African Population Prevalent Genetic Variations of Dihydropyrimidine Dehydrogenase as the 5-flourouracil Cancer Drug Metabolizing Enzyme: Computational Approaches towards Pharmacogenomics Studies

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

DOCTOR OF PHILOSOPHY

IN BIOINFORMATICS

RHODES UNIVERSITY,SOUTH AFRICA

Research Unit of Bioinformatics (RUBi)

DEPARTMENT OF BIOCHEMISTRY AND MICROBIOLOGY

Faculty of Science

by

MAUREEN BILINGA TENDWA

ORCID: https://www.orcid.org/0000-0002-2279-3326

FEBRUARY 2023

RHODES UNIVERSITY

Where leaders learn

Abstract

In an era of newly emerging cases of non-communicable diseases such as cancer, research is vital for both the medical and economic well-being of humanity. Pharmacogenomics has laid the groundwork for the identification of potential genes in cancer progression and treatment outcome investigations. Researchers are increasingly discovering heterogeneity in the efficacy and toxicity responses of drug metabolizing enzymes (DMEs) in diverse patient populations receiving anti-cancer therapy. DMEs comprise of Phase I (Cytochrome P450s) and Phase II (glutathione-S-transferases (GSTs), UDP-glucuronosyltransferases (UGTs), and dihydropyrimidine dehydrogenases (DPD) enzymes. The main cause of disparity in DME treatment outcomes is genetic variation, which causes missense mutations leading to structural and kinetic properties of the enzyme. These modifications have a deleterious impact on the pharmacodynamics and pharmacokinetics of drugs through multiple mechanisms. Presently, most cancer medicines are manufactured in developed countries based on the genetic background of non-African subpopulations. Thus, these drugs may not be optimally effective or can cause adverse side effects. Even though heterogeneity in toxicity and efficacy of these drugs has been observed in African descent, the basis of this population variance remains partially understood. For instance, a deficiency of DPD, the first-rate limiting metabolizing enzyme in the pyrimidine pathway, causes severe toxicity when exposed to 5-fluorouracil (5-FU) chemotherapy. However, minimum studies have been conducted to unravel its molecular mechanism which may unravel the observed drug treatment outcomes. The aim of this pharmacogenomics study was to determine the underlying mechanism by which DPD missense mutations, which are associated with an African ancestry subpopulation, provoke dysfunctional 5-FU metabolism, resulting in drug toxicity. This knowledge will be critical in designing drug modulators to aid in the restoration of DPD function, a hallmark of precision medicine. Therefore, in the first part of the research

we identified and reviewed the general role of Phase I and Phase II cancer drug metabolizing enzymes. We then used World Health Organization (WHO) essential medicine and drug.com to authenticate the usage of 5-FU as an anti-cancer treatment agent. The 3D structure and chemical structure of the agent was then downloaded from the Drug bank. Subsequently, Human Mutation Analysis - Variant Analysis PORtal (HUMA) and Mendelian Inheritance in Man (OMIM) were used to obtain data on DPD non-synonymous genetic variants. Additionally, the aggregate information of DPD missense mutations and their relation to human health were extracted from ClinVar and Pharmacogenomics Knowledge Base (PharmKGB). This information, along with additional data from single nucleotide polymorphisms (dbSNP), 1000 Genomes Project and Exome Sequencing Project (ESP MAF) considering variants classified based on their minor allele frequency (MAF) of 0.001, as well as research articles, consolidated information on missense mutations associated with African subpopulations. Finally, the wild type (WT) and detected mutation sequences were obtained from the Universal Protein Resources database (UniProt). However, because the 3D structure of human DPD was missing, the dimeric wild type (WT) human 3-dimensional (3D) structure was modeled via MODELLER using the pig's structure as a template. PRIMO, HHpred, and the Protein Data Bank (PDB) were all used to locate the suitable template. As a result, six clinical (C29R, M166V, Y186C, S534N, I543V, and D949V) and thirteen non-clinical (S201R, K259E, D342N, D432N, S492L, R592Q, A664S, G674D, A721T, V732G, T768K, R886C, and L993R) mutations were discovered. Using AMBER tools, we then determined accurate force field parameters for each monomer of DPD protein's Fe^{2+} centers. Following the creation of each mutation model structure in Discovery Studio, the resulting AMBER force field parameters were inferred. For each model structure, a drug free (inactive/open-conformation) and drug bound (active/closed-conformation) model structure was created (WT and mutations). The model structures were validated using the consensus of three

validation programs, namely ERRAT, PROCHECK, and ProSA. Similarly, the impact on structural functionalities was predicted by consensus from Variant Analysis Porta (VAPOR) web server, which include three support vector machines (SVM)-based tools; PhD-SNP, MUpro, and I-Mutation. After protonation in the H++ web server, the six clinical and thirteen non-clinical (six active site and seven non-active site) mutations identified were then exposed to 600 ns molecular dynamic (MD) simulation. The non-clinical data was divided into two categories to better understand the impact of the mutation based on its position in the protein: six catalytic-domain (R592Q, A664S, G674D, A721T, V732G, and T768K) and seven remote (S201R, K259E, D342N, D432N, S492L, R886C, and L993R) missense mutations.

The post-MD analysis was done using the typical existing computational global investigations [RMSD, all versus all RMSD, RMSF, RG, hydrogen bonds (H-bonds) and dynamic cross correlation (DCC)]. In addition, we used *in silico* tools newly developed within the Research Unit in Bioinformatics (RUBI) group, such as comparative essential dynamics (ED)-principal component analysis and dynamic residue network (DRN) multi-metric [*betweenness centrality (BC), closeness centrality (CC), degree of centrality (DC), eigen-centrality (EC) and Katz centrality (KC)*] analysis algorithms. From the analysis, it was observed that the loop regions of the mutation proteins had increased loop flexibility, particularly around the catalytic loop, which could account for the enhanced asymmetric behavior of the mutation's monomers compared to the WT. Notably, the A664S mutant showed relatively lower fluctuations, deviating from the observed heightened flexibility in other mutants. A general decrease in hydrogen bonds was observed in the 5-FU binding environment of the mutations compared to the WT. In particular, 5-FU contact analysis of the WT versus the mutation revealed a reduction in contact between core 5-FU binding residues and catalytic residues Cys671 and Ser670, which form hydrogen bonds that initiate DPD catalytic action. Additionally, *BC* was used to quantify the importance of a protein residue

based on how often it acted as a bridge along the shortest paths between other residues. It reflected the potential control or influence a residue may have over communication between different parts of a protein structure. *DC* assesses the number of connections or interactions a residue had with other residues in the protein, indicating its overall connectivity within the structure. In both drug free and drug bound state, DPD data from the active site hubs' *BC* and *DC* revealed a dimeric asymmetric communication pathway per monomer involving a cluster of newly introduced hubs ensemble along the oxidoreduction conduit from NADPH to 5-FU. The two *BC* communication pathways were located more on the interior of the oxidoreduction conduit, while the two *DC* communication pathways were located on the exterior. In both cases, one pathway dominated the other. Partially lost function reported in mutation systems could be credited to the compensation communication response to the catalytic site via the least compromised routes. Similar patterns were observed in allosteric communication pathways to the active site induced by remote mutations. Mutations may have destabilized the active-loop and 5-FU binding environment, resulting in a compensatory mechanism seen by the addition of new hubs to the communication network. Surprisingly, *EC* hubs in the WT were found within the catalytic site domain, indicating that the region is important in 5-FU metabolism. *EC* measured the importance of a residue by considering both its own degree of connectivity and the degrees of connectivity with its neighboring residues, highlighting its significance in information flow and communication. Herein, *EC* hubs in mutant systems were found to lose this importance, with active site domain mutations suffering the most. This could explain why non-clinical catalytic domain mutations R592Q, A664S, and G674D, as well as clinical catalytic domain mutations S534N and I543V, experienced drug exit in one of their monomers during simulation. In contrast, there was no 5-FU exit in the non-clinical remote domain. Additionally, aside from the active site, *KC* hubs were also found around the cofactors, indicating that these components were equally

important in DPD overall function. *KC* combines the concepts of both degree centrality and eigen-centrality, it incorporated both direct and indirect interactions to evaluate the importance of a residue, assigning higher centrality to residues that have connections to other highly central residues. Hence, providing a more comprehensive measure of influence within the protein network. More importantly, *CC* is known to measure how efficiently a residue can interact with other residues in the protein, considering the shortest path lengths. It indicates the proximity of a residue to others, suggesting its potential for information transfer or functional integration. *CC* revealed that the majority of persistent hubs were found within the protein-cores known as cold-spots.

Overall, this study highlighted the communication pathways triggered by active site domain mutations, as well as the allosteric communication pathways triggered by each remote mutation in both drug free and drug bound states of the DPD enzyme. Both clinical and non-clinical mutations revealed each protein's adaptive compensation mechanism, which results in partial function loss. In each case, the communication network of the different monomers changed from inactive to activated DPD protein. Cold-spot areas were discovered to contain key persistent residues involved in protein function and stability. These areas have been proposed as potential targets for new or repurposed pharmacological modulators that can restore enzyme function. In the pursuit of precision medicine, it also lays the groundwork for detecting and explaining the molecular mechanisms of other drug metabolizing enzymes related to the African-descent subpopulation.

Declaration

I, Maureen Bilinga Tendwa, declare that this is my own unaided work, except where duly acknowledged. It is being submitted for the degree of Doctor of Philosophy in Bioinformatics to the Faculty of Science at Rhodes University. It has not been submitted for any degree or examination at any other university and all the sources I have used or quoted have been indicated and acknowledged by complete references.

Signature: 

Date: 14th February 2023

Dedication

(Philippians 4:13)

(Thank you, Almighty God, in whom: I could do all things through Christ who strengthens me).

I dedicate this accomplishment to my mother, Mrs. Nellie Esther Tendwa, who has raised me in the ways of the Almighty God since my father's death as a child.

Table of Contents

Acknowledgements

I would like to express my gratitude and appreciation to:

My supervisor, Professor Özlem Tastan Bishop, for her unending support, mentorship, leadership, inspiration, dedication, guidance, and expertise. She gave me a whole new perspective on how to think and look at things through the eyes of a leader. Indeed, I stood on the shoulders of a giant.

Dr Thommas M. Musyoka (Kenyatta University, Kenya), my co-supervisor, who gave me guidance numerous times and helped with any troubleshooting. Thank you for the strong academic support.

Professor Kevin Lobb from the Chemistry Department, Rhodes University, for assisting me in the initial part of my research. I am forever grateful.

To my siblings, Mario, Gregory, Richard, Eugene, Innocencia and Christabel and in-laws Allan, Eric, Nosipho, Rehema and Moratuoa, thank you for believing and being there for me. Let this be proof to my daughter and all my nieces and nephews that hard work is the most important key to success.

To the staff and students at the Research Unit of Bioinformatics (RUBi), Rhodes University. Thank you for all your assistance, suggestions, and technical support. Indeed, each of you contributed to my intellectual growth. Dr. Olivier Sheik Amamuddy, for providing an innovative platform for our research group analysis. Dr. Rita Afriyie Boateng, who was very kind and open-hearted. Dr Lorna Chebon-Bore for mentorship in the initial part of the project. Chiratidzo Chamboko, Reine Bebey, Victor Barozi, and Mokgerwa Monama, thank you for the friendliness and encouragement. I am grateful to Mr and Mrs Glenister for guidelines, assistance, and technical support. I would also like to express my heartfelt appreciation to the National Institute of Health (NIH)'s Human Heredity and Health in Africa (H3Africa) partnership for partially sponsoring my research.

Research Outputs

Primary Publication

1. **Tendwa MB**, Chebon-Bore L, Lobb K, Musyoka TM, Tastan Bishop Ö. Force Field Parameters for $Fe^{2+}{}_4S^{2-}{}_4$ Clusters of Dihydropyrimidine Dehydrogenase, the 5-Fluorouracil Cancer Drug Deactivation Protein: A Step towards In Silico Pharmacogenomics Studies. Molecules. 2021 May 14;26(10):2929. doi: 10.3390/molecules26102929. PMID: 34069161; PMCID: PMC8156676.

Secondary Publications

2. Hamdi Y, Zass L, Othman H, Radouani F, Allali I, Hanachi M, Okeke CJ, Chaouch M, **Tendwa MB**, Samtal C, Sallam RM, Alsayed N, Turkson M, Ahmed S, Benkahla A, Romdhane L, Souiai O, Tastan Bishop Ö, Ghedire K, Fadlelmola FM, Mulder N, Kassim SK. "Human OMICs and Computational Biology research in Africa: Current challenges and prospects." PubMed 2021 April 1doi: 10.1089/omi.2021.0004.

3. Othman H, Zass L, da Rocha JEB, Radouani F, Samtal C, Benamri I, Kumuthini J, Fakim YJ, Hamdi Y, Mezzi N, Boujemaa M, Okeke CJ, **Tendwa MB**, Sanak K, Chaouch M, Panji S, Kefi R, Sallam RM, Ghoorah AW, Romdhane L, Kiran A, Meintjes AP, Maturure P, Jmel H, Ksouri A, Azzouzi M, Farahat MA, Ahmed S, Sibira R, Turkson MEE, Ssekagiri A, Parker Z, Fadlelmola FM, Ghedira K, Mulder N, Kamal Kassim S. African Genomic Medicine Portal: A Web Portal for Biomedical Applications. J Pers Med. 2022 Feb 11;12(2):265. doi: 10.3390/jpm12020265.

Conference Presentation

Micro-talk presentation

4. **Maureen Bilinga Tendwa**, Lorna Chebon-Bore, Kevin Lobb, Thommas Mutemi Musyoka, and Özlem Tastan Bishop. Laying the Foundation for In Silico Pharmocogenomics Studies through Parametrization of the $Fe^{2+}{}_4S^{2-}{}_4$ Clusters in Dihydropyrimidine Dehydrogenase: The 5-Fluorouracil Cancer Drug Metabolizer. 15th Centre for High Performance Computing (CHPC) National Conference, Johannesburg, South Africa. 1st – 3rd December 2021.

List of Figures

List of Tables

List of Supplementary Figures and Tables

Figures

Tables

Key Words

5-fluorouracil missense

AMBER force field parameters

Dihydropyrimidine dehydrogenase (DPD)

Dynamic residue network

Iron [$Fe^{2+}_4S^{2-}_4$] clusters

Molecular dynamics (MD) simulation

Mutation-induced communication network

Seminario approach

Quantum mechanics (QM)

Pharmacogenomics

Precision medicine

List of Equations

List of Abbreviations

3D	Three-dimensional
5-FU	Five-fluorouracil
6-MP	Six-mercaptopurine
ACPYPE	Ante-Chamber Python Parser interface
ALL	Acute-lymphoblastic leukemia
BC	*Betweenness centrality*
CC	*Closeness centrality*
CHPC	Center for High Performance Computing
CPU	Central processing unit
CYP	Cytochrome P450
DC	*Degree centrality*
DHFU	5,6-dihydrofluorouracil
DME	Drug metabolizing enzymes
DNA	Deoxyribonucleic acid
DPD	Dihydropyrimidine dehydrogenase
DRN	Dynamic residue network
dUMP	Deoxyuridine monophosphate
dTMP	Deoxy-thymidine-monophosphate
EC	*Eigen centrality*
ED	Essential dynamics
FAD	Flavin adenine dinucleotide
FβAL	Fluoro-alanine
FdUDP	Fluorodeoxyuridine diphosphate
FdUMP	5-Fluoro-2'-deoxyuridine-5'monophosphate

FMN	Flavin mononucleotide
FUDP	Fluorouridine diphosphate
FUMP	5-Fluorouridine monophosphate
FUPA	Fluoro-ureidopropionic acid
FUTP	Fluorodeoxyuridine diphosphate
GC/MS	Gas chromatography-mass spectrometry
GSTs	Glutathione-S-transferases
GWAS	genome-wide association studies
HIV	Human Immunodeficiency Virus
IARC	International Agency for Research on Cancer
KC	*Katz centrality*
LC/MC	Liquid chromatography-mass spectroscopy
MCBP	Metal center parameter builder
MD	Molecular dynamics
MM	Molecular mechanics
NADP	Nicotinamide adenine dinucleotide phosphate
NMR	Non-destructive nuclear magnetic resonance
OMICS	Genomics, Proteomics', Transcriptomics and Metabolomics
OPRT	Orotate phosphoribosyltransferase
PCA	Principal component analysis
PBC	Periodic boundary conditions
PDB	Protein Data Bank
PME	Particle Mesh Ewald
PRPP	Phosphoribosyl diphosphate
QM	Quantum mechanics

RESP	Restricted electrostatic potential
RNA	Ribonucleic acid
mRNA	messenger Ribonucleic acid
RNR	Ribonucleotide reductase
SSA	Sub-Saharan Africa
SN-38	7-ethyl-10-hydroxycamptothecin
SN-38G	SN-38-glucoronide
TB	Tuberculosis
TGN	Toxic thioguanine nucleotide
TK	Thymidine kinase
TP	Thymidine Phosphate
TMPT	Thiopurine methyltransferase
TS	Thymidine synthase
UDP	Glucoronosyltransferase
UGTs	UDP-glucuronosyltransferases
UK	Uridine kinase
UP	Uridine phosphorylase
URF	Five fluorouracil
UTP/DTTP	Uridine-5'-triphosphate/2'-deoxythymidine-5'-triphosphate
VFFDT	Visual force field derivation toolkit
WT	Wild type

List of amino acids

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

List of tools and web servers

Web server	URL
ClinVar	http://www.ncbi.nlm.nih.gov/clinvar/
DrugBank	https://go.drugbank.com/
Ensemble	http://www.ensembl.org/info/docs/tools/vep/index.html/
ESP-MAF	http://evs.gs.washington.edu/EVS/
H++	http://biophysics.cs.vt.edu/
HUMA-VAPOR	http://www.ncbi.nlm.nih.gov/omim/
MAFFT	https://www.ebi.ac.uk/Tools/msa/mafft/
MD-TASK	https://md-task.readthedocs.io/en/latest/home.html
MDM-TASK-web	https://mdmtaskweb.rubi.ru.ac.za/
OMIM	http://www.ncbi.nlm.nih.gov/omim/
PharmKGB	http://www.pharmgkb.org/variantAnnotations/
ProSA	https://prosa.services.came.sbg.ac.at/prosa.php
PROCHECK	https://servicesn.mbi.ucla.edu/PROCHECK/
RCSB PDB	https://www.rcsb.org/
UniProt	http://www.uniprot.org/
Verify3D	https://servicesn.mbi.ucla.edu/viewer/

THESIS SUMMARY

This thesis uses dihydropyrimidine dehydrogenase (DPD), a pyrimidine degrading enzyme, as a case study to demonstrate the impact of African population-related missense mutations in 5-fluorouracil 5-FU metabolism at the molecular level. 5-FU is a pyrimidine antimetabolite chemotherapeutic drug that is commonly used to treat cancer, specifically breast, colon, rectum, stomach, and pancreas cancers. However, the presence of missense mutations in DPD protein in the general population, including people of African descent, has been linked to increased toxicity due to altered 5-FU metabolism. Thus, understanding the effects of mutations on protein structure and function is critical, particularly for precision medicine. Using post-molecular dynamics studies and dynamic residue network (DRN) analysis, this thesis aims to identify African-related DPD protein missense mutations and decipher their effect on the protein structure and function.

It is divided into six chapters, beginning with an introduction (Chapter 1), followed by parametrization chapter (Chapter 2 Chapters 3, 4, and 5 were split into two major sections: Chapter 3 focused on clinically significant mutations, while Chapters 4 and 5 (Part II) addressed mutations of unknown significance (those that are yet to be validated). Using the mutations of unknown significance, the impact of allosteric/remote (Chapter 4) and catalytic domain (Chapter 5) mutations in dysregulated metabolism of 5-FU was ascertained. These findings shed light on how African population specific DPD mutations in various parts of the protein cause a variety of drug metabolism concerns. It also lays the groundwork for future research to improve treatment outcomes by discovering novel hits for personalised tailored precision oncology drugs. Guidelines for implementing precision medicine based on drug resistance mutations can thus be developed. Chapter 6 is a conclusion chapter.

The introduction chapter provides a general overview of pharmacogenomics and pharmacokinetics of drug metabolizing enzymes in precision medicine, with a particular emphasis on DPD and its implication in cancer therapy. In addition, the computational approaches used in the study were highlighted. Finally, primary research motivation, goals, and specific objectives were addressed.

The second chapter describes the generation of AMBER force field parameters for Fe^{2+} centers in DPD proteins using both the original Seminario (Model 1) and automated VFFDT Seminario (Model 2) approaches. This was done to pave the way for the subsequent wild type

(WT) and mutant protein analyses via molecular dynamics (MD) simulations and post-MD calculations. It is worth mentioning that this chapter was published in a peer-reviewed journal.

The third chapter elaborates on the effects of clinically significant DPD enzyme missense mutations, associated with people of African descent, on the drug metabolism.

Chapter 4 and 5, both, analyse the effect of mutations that are previously studied by *in vitro* and *ex vivo* approaches rather than direct clinical studies. Chapter 4 focuses on the mutations located away from active site area while Chapter 5 investigates the effect of mutations located in the vicinity of active site region.

Chapter 3, 4, 5 use combined *in silico* methods such as MD and post-MD analysis coupled with an approach that was established in our research group called dynamic residue network (DRN) while aiming to link the mutation effect into different level of drug metabolism.

The same analysis approaches as in Chapter 4 were used to depict the change in communication pattern within the active site residues caused by the presence of these mutations.

Chapter 6 concludes with a summary of all significant findings obtained in this work, their applications, and future work proposals.

CHAPTER 1
INTRODUCTION

1.1 General insight on pharmacogenomics of drug metabolizing enzymes in precision oncology

According to the 2019 United Nations World Population Prospects report, the global population is expected to grow by two billion people between 2019 and 2050, with 1.05 billion (52 percent) of them in Sub-Saharan African countries (1). Without a doubt, one of the reasons for this rise is the adequate supply and use of pharmaceutical drugs, which has resulted in prolonged life expectancy (2, 3). Technological advancements over the past decade have transformed the face of drug discovery beyond recognition (4). The incorporation of computational and experimental methods continues to play a pivotal role in drug design, opening up new opportunities for the drug discovery and development process (5). Computational tools (*in silico*) for drug metabolism prediction are at the scientific community's disposal. These tools have been used to generate improved drug target signatures pertinent to metabolic routes (6). Additionally, computational frameworks which evaluate drug metabolism and toxicological as well as pharmacokinetic effects, including the use of the adverse outcome pathway approach for risk assessment, have the potential to enhance the efficiency and speed of preclinical drug development (7). Generally, the pharmacokinetics and pharmacogenomics of a drug have a significant impact on its efficacy, dosing, and toxicity. Thus, influencing clinically relevant outcomes in drug dosing based on individuals' genetic predisposition, a concept known as precision medicine, is gaining increasing momentum (8). The rising paradigm in precision medicine holds great potential for providing efficacy and safety in drug design and development (9). It is also essential in oncology, where most anti-cancer medications have a narrow therapeutic index and diverse inter-individual pharmacokinetic variability (10, 11). Missense mutations caused by genetic variation in the gene encoding drug targets, transporters, and metabolizing enzymes can help to explain some of this variability (11). Thus, understanding the molecular mechanism by which missense mutations affect drug disposition and action in their metabolizing enzymes may hold the key to developing personalized (precision) drugs with improved efficacy and safety, including cancer drugs.

1.2 Cancer

The increasing disease burden in both infectious diseases such as tuberculosis (TB) and human immunodeficiency virus (HIV) and non-communicable diseases like cancer, cardiovascular diseases, and diabetes, demands the understanding of the basic molecular mechanism of susceptibility and response to drugs in the struggle towards improving healthcare solutions in Sub-Saharan Africa (SSA). For instance, cancer accounts for 7-10 million global deaths per annum (12). The impact is particularly pronounced within middle-income and low-income nations, characterized by a swift upsurge in cancer incidences, with a noteworthy 9.3 % of the populace facing a mortality risk from cancer before reaching the age of 75 years (13-15). Presently, 60% more Africans succumb to cancer than malaria (16). According to the International Agency for Research on Cancer (IARC), 693 487 Africans died from cancer in 2018, with a projected increase to 1 429 812 deaths by 2040 (15, 17). This figure represents a 106 % increase in deaths predicted between 2018 and 2040. Figure 1.1 illustrates the incidence and mortality rate of cancer in the year 2020. Cancer incidence in SSA has more than doubled in the last 30 years (18, 19). Without intervention, analysts predict that these figures will surge in the next 20 years, resulting in a doubling of cancer mortalities in Africa (20).

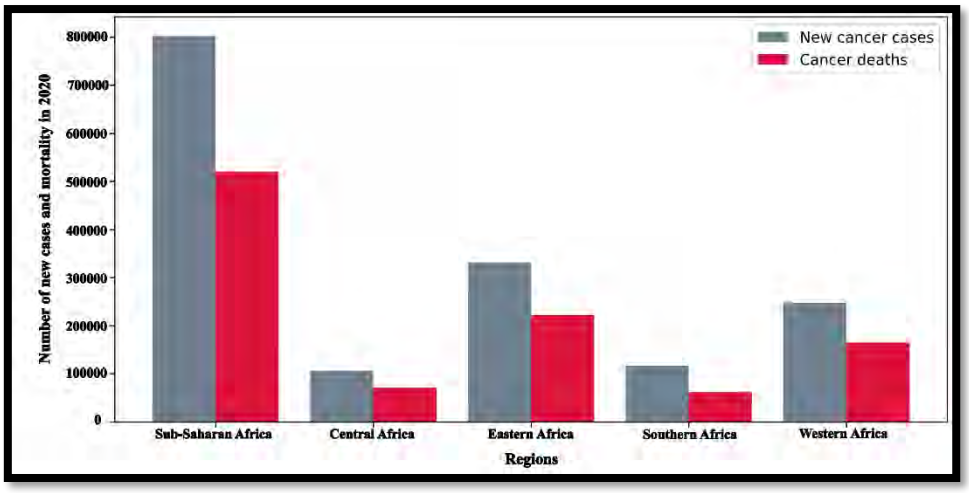


Figure 1.1: Cancer incidence and mortality rates in Sub-Saharan Africa sub-regions in 2020. Grey represents the number of new cases, while crimson represents the number of deaths. Furthermore, the Y-axis represents the number of new cases and mortality in 2020, while the X-axis represents the sub-regions of Sub-Saharan Africa. The statistical data was extracted from Ngwa and colleagues (20).

The observed increase in cancer incidence rates can, in part, be attributed to improved reporting practices (20). Over the years, advancements in diagnostic techniques, increased

awareness, and enhanced cancer registries have led to more accurate identification and reporting of cancer cases (21). This improved reporting not only captures previously undetected or unreported cases but also ensures comprehensive data collection (14). Studies have shown that the implementation of more rigorous reporting systems has resulted in a rise in cancer incidence rates, particularly for certain cancer types (22, 23). For instance, the introduction of widespread mammography screening has led to increased detection of breast cancer cases (24-26). Additionally, the implementation of population-based cancer registries has contributed to better data collection, enabling a more accurate assessment of cancer trends over time (14, 17, 21). Therefore, when evaluating the apparent growth in cancer rates, it is important to acknowledge the influence of improved reporting practices alongside other factors impacting cancer incidence (20). External and internal factors such as infection, changing population demographics, behavioural changes, environmental exposures, and genetics are all contributing to Africa's cancer epidemic (27).

1.3 Population genetic variation

Past external drivers of oncogenesis, genetics and tumour biology are also held accountable for the disproportionate impact of several cancers on the general population including people of African descent (28). Africa's population history reveals a high degree of genetic and phenotypic diversity due to the presence of numerous ethnic groups and tribes (29-31). Consequently, the genetic differences within these groups likely play a significant role in the observed disparities in cancer prognosis (32, 33). Migration patterns and adaptive pressures in ancient times led to genomic heterogeneity within African populations, resulting in the formation of five subgroups: Central African, Northern Khoe-San, Southern Khoe-San, East African, and West African (34, 35). Some individuals from these subgroups migrated out of Africa and formed what is now known as the "out-of-Africa" population (34, 36). Thus, Africa can be seen as a repository of human genomic diversity, offering an opportunity to explore the role of genomics in promoting human health equity (31). This repository is augmented by modern North African populations enriched with the genetic pool of “out-of-African” Euro-Asian populations. Nonetheless, Africa's contribution to world genetic and genomics data is grossly disproportionate to the diversity and size of its population. For instance, the HapMap and 1000 Genome projects included very few African populations (31, 34). This is a significant shortcoming for a group of people who account for more than 90% of human genomic diversity. According to a recent review of genome-wide association

studies (GWAS), African descents account for only 2.4% of individuals included in all GWAS studies (37).

In addition, research into human genetics has uncovered that the complexities of demographic history led to genetic diversity in different populations. Therefore, different ethnic groups should exhibit genetic variation in cancer-related genes (38, 39). Besides the prognosis, this may influence different ethnic groups' susceptibility to certain types of cancer and sensitivity to treatment options such as chemotherapy and radiotherapy (40, 41). Recognition of population genetic variation for anti-cancer agents through pharmacogenomics has significant potential impact on chemotherapeutic approach given the narrow therapeutics index of these drugs (9, 10).

1.4 Pharmacogenomics of drug metabolizing enzymes

Pharmacogenomics is defined as the study of individual differences in drug response, as well as the mechanism underlying the variable drug response, using an in-depth understanding of genomics, proteomics, transcriptomics, and metabolomics (9). Overall, genomics is the study of an organism's entire genetic complement (the genome). It uses DNA sequencing methods, recombinant deoxyribonucleic acid (DNA), and bioinformatics to sequence, assemble, and analyse the structure and function of genomes (42). This considers the genetic diversity that defines allele and allele frequency variation within and across human populations (43). The role of genomic variation in health and disease has been a major focus of human genomics research over the last three decades (31, 44). According to recent research, more than 90% of the human genome is transcribed to ribonucleic acid (RNA) (45). A cell's transcriptome is a collection of all the RNA molecules, or transcripts, that it contains. To create the transcriptome, an organism's DNA is first transcribed by RNA polymerase to produce complementary RNA strands, which are then spliced to remove introns, yielding mature transcripts containing only exons (46). The protein-coding messenger RNA (mRNA) transcripts are then used as templates for protein translation, whereas non-coding RNAs (ncRNAs) play a variety of structural and regulatory roles in the molecular cell biology (47).

Proteomics is the study of the interactions, function, composition, and structures of proteins, as well as their cellular activities (48). Compared to genomics, it provides a better understanding of the structure and function of the organism, although it is much more complex than genomics because protein expression varies over time and under different environmental conditions (49). Current proteomics research is primarily concentrated in three

areas: functional proteomics, structural proteomics, and expression proteomics. (50). Functional proteomics approaches are aimed at two major goals: revealing unknown proteins' biological functions and defining cellular mechanisms at the molecular level. Many proteins in cells exhibit their biological functions through rapid and transient association within large protein complexes (51). Understanding protein functions and unravelling molecular mechanisms within the cell are then dependent on identifying interacting protein partners. In fact, the association of an unknown protein with partners from a specific protein complex involved in a specific mechanism would be strongly suggestive of its biological function (51, 52). Expression proteomics studies, on the other hand, are concerned with the investigation of protein expression patterns in abnormal cells (i.e., malignant, drug-stimulated, etc.) relative to healthy cells. This comparative approach is commonly used in biomedical applications to identify proteins that are downregulated or upregulated in a disease-specific way for use as diagnostic markers or therapeutic targets (53-55). In structural proteomics, X-ray crystallography and nuclear magnetic resonance spectroscopy are frequently used to determine the three-dimensional (3D) structure and structural complexities of functional proteins (56, 57). A protein's amino acid sequence determines its 3D arrangement, and this correct structure is uniquely positioned to perform its intended biological function. Any changes in the amino acid sequence caused by diseases can result in changes in proteins dynamics, affecting their biological function (58, 59). The investigation of these changes has led to the discovery of the mechanism of drug binding in a protein and how it may affect its metabolism, paving the way for drug research and development (60).

Biological evaluations of drug metabolism frequently necessitate the use of high-precision analytical methods to identify and quantify metabolites produced by a drug's biotransformation (61). Furthermore, metabolomics which involves identification, quantification, and documentation of metabolites in various biological systems using modern analytical platforms has also become an important tool in pharmacogenomics (62, 63). Liquid chromatography-mass spectrometry (LC/MS) (64, 65), non-destructive nuclear magnetic resonance (NMR) spectroscopy (66), and gas chromatography-mass spectrometry (GC/MS) have all helped in the profiling of metabolites in complex biological matrices (67). Metabolomics, like transcriptomics, has been instrumental in establishing links between genetic regulation, metabolite phenotyping, and biomarker identification (68). In clinical practice, metabolomic analysis of human biofluids is used to distinguish between physiological and pathological states in humans, aid in early disease biomarker discovery, and predict individual response to drug therapy (69-71).

The omics (genomics, proteomics, transcriptomics, and metabolomics) subclass of biological studies (mentioned above) is gaining popularity in clinical and academic health arenas (72). All these “omics” technologies represent promising avenues for accurately characterizing diseases (73). Precision medicine is based on omics pathways to provide the most effective treatment for individuals (31, 74).

1.5 Precision medicine

Precision medicine is a concept that provides personalized care by considering both population characteristics and individual variability; this broadens biological knowledge and investigates the vast diversity of individuals (75). Understanding the underlying heterogeneity of numerous complex disease processes necessitates new strategies presented in precision medicine for prediction, prevention, and individualized treatment strategies (76). This framework must be tailored to each person’s unique omics, resulting in individualized disease management (77). Furthermore, it empowers clinicians to predict the most appropriate course of action for patients with complex diseases like cancer, diabetes, cardiomyopathy, and COVID-19 (78). For instance, numerous drugs' efficacy and toxicity are influenced by interindividual genetic variations in drug metabolizing enzymes and transporters (9). Recently, the precision oncology paradigm has reinforced pharmacogenomic studies. Better understanding of persons' responses to medication based on their genomic information has enabled the evaluation of their uniqueness to pharmacodynamics and pharmacokinetics of drugs (9). Not only does this allow for the effective and safe administration of existing medications, but it also holds the promise of increased safety and efficacy in drug design and development based on subpopulation genetic constitution (9, 79). The potentials of various drug metabolizing enzymes are influenced by interethnic genetic makeup. In addition, advances in pharmacogenomics have revealed that missense mutations caused by altered DNA may account for some of the variability in phase I and phase II drug metabolizing enzymes (DMEs) (9, 80).

Traditional drug metabolism is divided into three phases: modification (phase I), conjugation (phase II), and elimination (phase III) (through bile or urine) (11). The Cytochrome P450 (CYP) enzyme family is critical for drugs' phase I modifications (hydrolysis, oxidation, and reduction) (81). Enzymes such as uridine diphosphate glucuronosyltransferases (UDGs) and glutathione S-transferases (GSTs) carry out phase II conjugation reactions that result in drug inactivation or activation (UGTs) (82). The enzymes

that catalyse purine and pyrimidine analogue metabolism, thiopurine S-methyltransferase (TPMT) and dihydropyrimidine dehydrogenase (DPD), have several known polymorphisms in their encoding genes that cause deleterious amino acids substitution (missense mutations). These mutations may have an impact on the stability, structure, and dynamics of such DMEs (11).

1.6 Impact of missense mutation on phase I drug metabolizing enzymes related to Cytochrome P450

I. Cytochrome P450

The Cytochrome P450 (CYP) has been described as a broad and diversified heme-comprising enzyme superfamily, which takes part in the oxidative metabolism of structurally diverse molecules such as chemicals, fatty acids, and drugs (83, 84). Approximately a dozen enzymes from CYP1, 2 and 3 families of enzymes are responsible for 70-80% of all phase I-dependent drug metabolism, out of 57 putatively functional human CYPs (85). Additionally, CYPs 1A2, 2C8, 2C9, 2E1, and 3A4 are the most abundant in the liver, whereas CYPs 2A6, 2B6, 2D6, 2C19, and 3A5 are less abundant, and CYPs 1A1, 1B1, and 2J2 are mostly expressed extrahepatically (86). Mutations in cytochrome P450 enzymes (CYPs) can be classified either genetically or pharmacologically. Genetic classification primarily focuses on the influence of underlying genetic variations on enzyme structure and function. In contrast, pharmacological classification emphasizes the practical implications of these genetic variations on drug metabolism and interactions (87, 88). Depending on the impact on enzymatic function, these mutations can lead to either loss of function or gain of function, resulting in subcategories such as ultrarapid, intermediate, normal, and poor metabolizers (89-91). Gain-of-function in these mutations may increase medication clearance, resulting in a decrease in concentration, whereas loss-of-function mutations decrease clearance, resulting in an increase in concentration. For instance, mutation-induced loss of function in CYP2D6 has rendered pharmacological lack of production in active analgesic codeine (86). Although monogenic polymorphisms account for a significant portion of the differences in only a few enzymes (most notably CYP2D6), most enzymes are multi-factorially controlled, including additional polymorphisms in regulatory trans-genes and nongenetic host factors such as age, disease, sex, hormonal and diurnal influences, and other factors (Figure 1.2) (86).

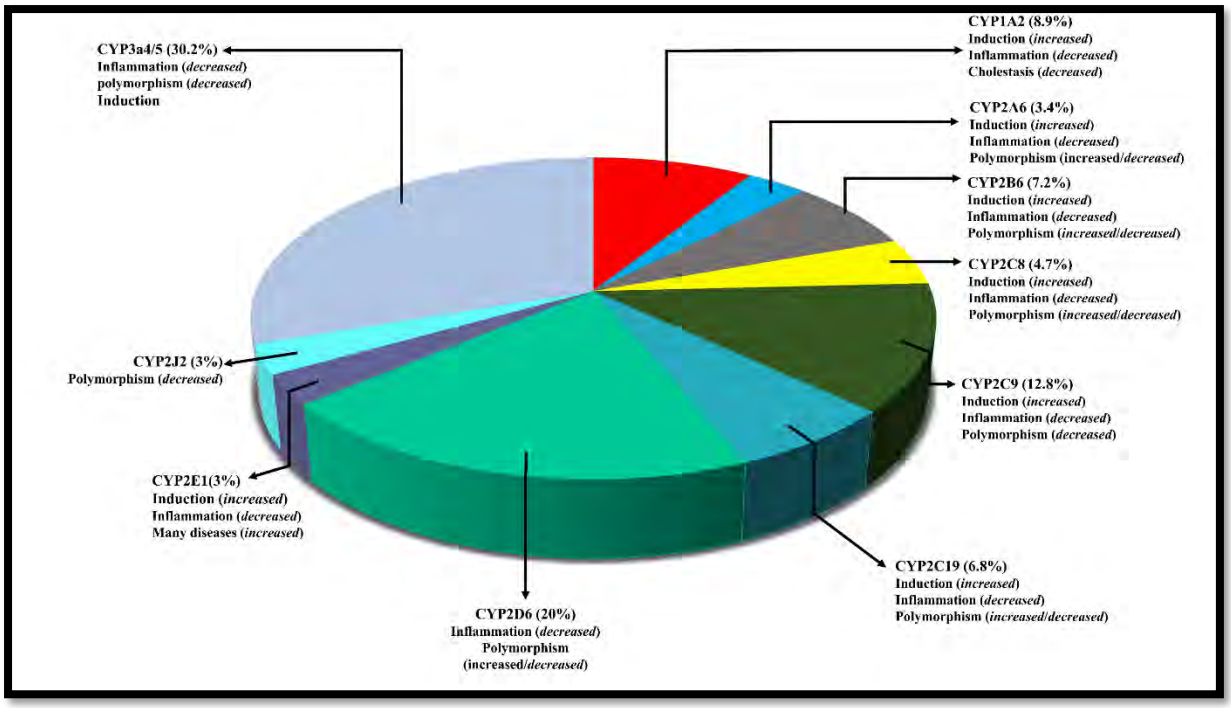


Figure 1.2: The percentage representation of clinically used metabolized isoforms (protein mutations) and the factors influencing variability. Significant variations are highlighted, with potential influence as increased and decreased activity or both. Statistical data obtained from Zanger and Swab (86).

1.7 Impact of missense mutation on phase II drug metabolizing enzymes

II. Thiopurine methyltransferase (TMPT) and 6-Mercaptopurine

The cytosolic enzyme Thiopurine methyltransferase (TMPT, EC 2.1.1.67) catalyses the 5-methylation of aromatic and heterocyclic sulfhydryl compounds (92). Included among its substrates are thiopurine drugs 6-mercaptopurine (6-MP) and the prodrug thioguanine (93). These chemotherapeutic agents have been essential components of treatment protocol for acute lymphoblastic leukaemia (ALL), autoimmune disease, and transplant organ recipients (94, 95). Many modern treatment protocols for paediatric ALL employ successive cycles of thioguanine and 6-MP beginning as early as induction consolidation treatment and continuing up to 6 months after diagnosis (96, 97). The anti-tumour effect of 6-MP is achieved by inhibiting the formation of nucleotides required for RNA and DNA synthesis (98-100). As such, the catalysis of S-methylation of 6-MP by TPMT results in the formation of inactive metabolites. Toxic thioguanine nucleotide (TGN) is incorporated into DNA chains, resulting in replication inhibition and consecutive cellular degradation (98, 99). Presence of missense mutation due to genetic variation in the TMPT gene hampers the inactivation process (101-103). These factors have a significant impact on 6-MP

bioavailability and toxicity. The accumulation of TGN, which is inversely proportional to TPMT activity, is associated with a higher risk of adverse reactions to 6-MP (104-106). Patients with TPMT missense mutations are at increased risk of adverse haematological toxicities when treated with 6-MP because the mutations result in a slower rate of 6-MP metabolism (106-109).

Polymorphic TPMT activity has a well-defined molecular basis. Three TPMT alleles, designated as TPMT*2, TMPT*3A, and TPMT*3C, were seen to account for nearly 95% of TPMT deficiency cases (110). Each of these mutation alleles encodes mutated TPMT proteins that degrade rapidly, resulting in enzyme deficiency (111). It has been reported that the types and frequencies of TPMT alleles differ between ethnic groups (112, 113). Individuals who are homozygous for variant TPMT alleles accumulate toxic metabolites 10- to 15-fold more (114). Though, in those individuals, a 5-10% reduction in the standard dose can result in an antineoplastic effect, with partial or total loss of myelosuppression (115). Before commencing 6-MP therapy, TPMT testing is now used to optimize doses in children with ALL (116).

III. Uridine diphosphate [UDP]-glucuronosyltransferase 1A1 and Irinotecan

UDP-glucuronosyltransferases are a superfamily of enzymes that metabolize a variety of lipophilic molecules such as bilirubin, steroids, toxins, and drugs, including SN-38, the active metabolite of irinotecan (117). Due to its potent anti-tumour activity against a wide range of carcinomas, irinotecan is one of the most commonly prescribed chemotherapeutic agents (116). After intravenous administration, this drug is converted by carboxylesterase enzymes to the active metabolite 7-ethyl-10-hydroxycamptothecin (SN-38), which in turn inhibits topoisomerase-I (118). The SN-38 metabolite causes cytotoxicity by forming a ternary drug-topoisomerase-DNA complex with the cleavable complex (119). This complex is supposed to prevent the re-ligation of single strand breaks, which would otherwise disrupt the moving DNA replication fork. The halting of replication and the interaction of replication enzymes and the ternary complex causes lethal double-stranded breaks in DNA, resulting in irreversible DNA damage and cell apoptosis (120, 121). Nonetheless, because SN-38 is lipophilic, it is then inactivated to SN-38 glucuronide (SN38G) by UDP-glucuronosyltransferase during phase II metabolism (glucuronidation). The derived conjugated SN-38 glucuronide is water-soluble and is predominantly excreted via bile, with

approximately 30% excreted via the kidneys (122). This process is impacted by genetic variation in the UDP-glucuronosyltransferase 1A1 (UGT1A1) enzyme, which results in significant interpatient variability in SN38G formation (123). Polymorphisms in the UGT1A1 promoter region, which contains several repeating TA elements, are the most common cause of variations in UGT1A1 activity. The presence of seven TA repeats (referred to as UGT1A1*28) rather than the normal number of six results in decreased UGT1A1 expression and activity (124). Thus, UGT1A1*28 has been linked to decreased glucuronidation of SN-38, increased exposure to SN-38, and increased clinical toxicity in irinotecan patients (125, 126). The frequency of UGT1A1*28 alleles varies greatly across ethnic groups: UGT1A1*28 alleles are found in approximately 35% of African Americans and Caucasians, but much less frequently in Asians (127, 128).

1.8 Impact of missense mutation on phase I drug metabolizing enzymes not related to CYP-450

The genesis of missense mutations is genetic variation in genes encoding drug metabolizing enzymes. As stated earlier, this influences the pharmacokinetic and pharmacodynamic properties of these enzymes leading to drug toxicity and resistance (9). This poses a challenge in the field of drug research and development. Well-designed criteria in the early stages of drug development can lead to greater success (59). This is the primary focus of this research which is illustrated by a case study of the Dihydropyrimidine dehydrogenase (phase I) 5-FU drug metabolizing enzyme. The phenomena of allostery and missense mutations are discussed, including how these biological occurrences can be integrated into *in silico* novel drug discovery or repurposing.

IV. Case Study: Dihydropyrimidine dehydrogenase and 5-fluorouracil (5-FU)

The metalloenzyme dimeric dihydropyrimidine dehydrogenase (DPD; EC 1.3.1.2) is the initial rate-limiting phase I drug metabolizing enzyme in the triple-step pyrimidine-based catabolic pathway (129, 130). This enzyme promotes the NADPH-dependent degradation of (thymine and uracil) to 5,6-dihydrouracil and 5,6-dihydrothymine, respectively (130). Each monomer contains highly specialized 4 × $Fe^{2+}{}_4S^{2-}{}_4$ clusters, which allow for a well-orchestrated electron transfer cascade that initiates a redox catabolic reaction that leads to

pyrimidine degradation (129-131). In addition to its biological nucleotide metabolism function, the enzyme is an antineoplastic drug target (132). It is fully liable for the phase I metabolism of 5-fluorouracil (5-FU), a pyrimidine-like antineoplastic medication that is commonly prescribed (133). However, the presence of missense mutations in the general population, including the subpopulation of African descent, causes toxicity due to altered 5-FU degradation.

5-FU and its oral prodrugs capecitabine and tegafur are three of the most used antineoplastic drugs in the adjuvant and palliative treatment of cancer patients (134, 135). These antimetabolite chemotherapeutic agents belong to the fluoropyrimidines (FP) family first developed by Heidelberger in 1957 (136). This medication is widely used in many neoplasms and is a standard treatment for gastrointestinal (pancreatic, colorectal, and gastric), breast, and head and neck cancers (137, 138). 5-FU is a prodrug that must be converted intracellularly into cytotoxic metabolites with anti-tumour effects. Its metabolism involves two routes (catabolic and anabolic pathway) that are in competition with each other (138). Initially, while 5-FU is administered intravenously (parenteral) as a bolus infusion, or continuous infusion over many days, capecitabine and tegafur are taken orally (133). After oral administration, capecitabine and tegafur are then converted to their only active metabolite, fluorouracil, by CYP2A6 and thymidine phosphorylase enzymes, respectively (139, 140).

a. Catabolic pathway

Thereafter, 80 - 85% of the administered dose is degraded in the liver leading to the formation of α-fluoro-β-ureido propionic acid (FUPA) and α-fluoro-β- alanine (FβAL) which is excreted through the kidneys (Figure 1.3) (135). The kinetic investigation of intravenous bolus administration 5-FU disclosed that 60 - 90% of this drug is excreted within 24 hours either as CO_2 exhalation or FβAL in urine. Oral administration of 5-FU prodrugs (capecitabine and tegafur) results in poor bioavailability associated with variability in DPD activity (genetic variation/circadian). Due to inter- and intra-patient variability in 5-FU adsorption/elimination, unpredicted 5-FU levels in the plasma are stated (136). Several 5-FU prodrugs have been modulated to overcome severe toxicities after 5-FU administration in DPD deficient cancer patients (135, 136).

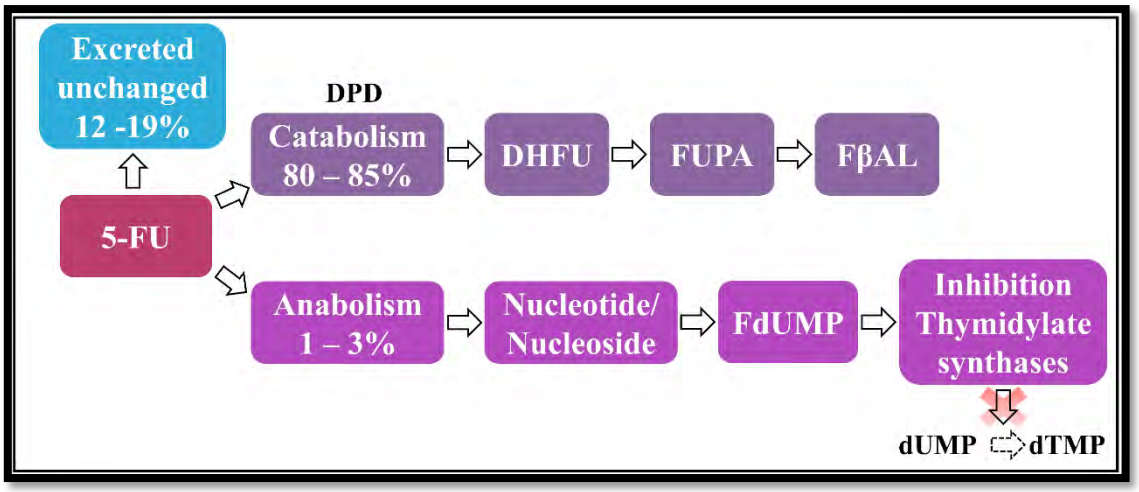


Figure 1.3: **An illustration of 5-FU multi-step activation pattern that leads to the production of cytotoxic 5-Fluoro-2′-deoxyuridine-5′-monophosphate (FdUMP).** FdUMP inhibits thymidylate synthase (TS) enzyme, causing intracellular accumulation of deoxy-uridine-monophospate (dUMP) and reduction of deoxy-thymidine-monophosphate (dTMP). This halts DNA synthesis. In addition, through catabolism dihydropyrimidine dehydrogenase (DPD) enzyme catalyses the reduction of 5-FU into 5,6-dihydrofluorouracil (DHFU). DHFU is then degraded into fluoro-ureidopropionic acid (FUPA) and fluoro-alanine (FBAL).

b. Anabolic pathway

In this pathway, the transformation of 1 - 3% of 5-FU to active metabolites competitively hinders the enzyme thymidylate synthase (TS) activities, hence thymine deficiency and inhibition of deoxyribonucleic acid (DNA) synthesis as well as cytotoxicity (141). Similarly, it inhibits ribonucleic acid (RNA) production, though to a lesser extent (Figure 3). These effects are most noticeable in rapidly growing cells and can result in cell death (141). Upon entry into the cells, 5-FU is converted into 5-fluorouridine monophosphate (FUMP) either indirectly through uridine phosphorylase (UP) and uridine kinase (UK) (FUR) or directly using orotate phosphorylase (OPRT) with phosphoribosyl transferase (PRPP) as a cofactor. FUMP is subsequently phosphorylated to fluorouridine diphosphate (FUDP), which is converted by ribonucleotide reductase (RNR) to the active metabolites fluorodeoxyuridine diphosphate (FdUDP) or fluorodeoxyuridine triphosphate (FUTP) (142). Ultimately, these active metabolites are mis-integrated into RNA and DNA instead of uridine-5'-triphosphate / 2′-deoxythymidine-5'-triphosphate (UTP/dTTP). Equally, 5-FU is converted to fluorodeoxyuridine monophosphate (FdUMP) either directly through the reduction of FUDP by RNR (a prevailing process) or indirectly through the sequential action of thymidine phosphorylase (TP) and thymidine kinase (TK) (29, 103, 143).

Thymidine synthase (TS) catalyses the methylation of deoxyuridine monophosphate (dUMP) to dTMP in the presence of 5-10-methylene-hydro folate, which aids in the biosynthesis of thymidylate (CH2THF). When FdUMP binds to TS, it forms a stable ternary complex that prevents dUMP access by constraining dTMP synthesis (Figure 4). A decline in dTMP levels causes a drop in deoxythymidine triphosphate (dTTP), which disrupts the level of other deoxynucleotides, resulting in severe restriction of DNA synthesis and repair. Furthermore, inhibiting TS causes an increase in dUMP, which raises the level of deoxyuridine triphosphate (dUTP). As a result, increased dUTP/dTTP and 5-FdUTP/dTTP cause dUTP and FdUTP misincorporation in DNA, resulting in DNA damage. The action of dUTPase (DUT), which generates dUMP and aids in preventing mis-integration into DNA, reduces dUTP accumulation, limiting the efficacy of 5-FU therapy (144-146). It should be noted that 12 - 19% of the administered parent drug is excreted unchanged in the urine (133, 147-149).

Herein, the human DPD protein is of particular interest. However, because we build the 3D human model using the crystal structure of pig homolog (PDB ID: 1H7X), the structural features will be defined based on the template structure.

The Dihydropyrimidine dehydrogenase (DYPD) gene, which is located on chromosome 1p21 and has 23 exons and a 3078-bp open reading frame (150), encodes the homodimeric DPD protein (130, 151). Within its five domains, each monomer of molecular mass ~110kDa contains four cofactors and a substrate. Domain I contains residues 27-173 and two sets of $Fe^{2+}{}_4S^{2-}{}_4$ clusters. Domain II (174-286 and 442-524) and domain III (residues 287-44) are $NADPH^-$ and FAD-binding domains, respectively, while domain IV (residues 535-847 and active site loop residues 675-679) is the FMN and URF-containing domain. The final domain V contains residues 1-26 and 848-1017 and the remaining two sets of $Fe^{2+}{}_4S^{2-}{}_4$ clusters (129, 130, 152, 153) (Figure 1.4).

The FMN and FAD are at opposite sides of the protein, separated by four $Fe^{2+}{}_4S^{2-}{}_4$ clusters: the first set of $Fe^{2+}{}_4S^{2-}{}_4$ (hetero atoms 1026 and 1027) clusters from each monomer are situated closer to the last set of $Fe^{2+}{}_4S^{2-}{}_4$ (hetero atoms 1028 and 1029) clusters from the opposing monomer (129, 130). Additionally, each monomer's catalytic domain (FMN and URF) is proximal to the C-terminal domain (set of $Fe^{2+}{}_4S^{2-}{}_4$ clusters) (47, 130, 154). This architecture is critical for electron transfer from the NADPH donor molecule to the catalytic sites (129, 130, 154). The precise mechanism by which these redox cofactors participate in the reaction, though, is largely unknown (155). Previous research has found that the $Fe^{2+}{}_4S^{2-}{}_4$ clusters act as a bridge between the FMN and FAD cofactors, allowing electrons to be transported to the active site (131, 152, 156).

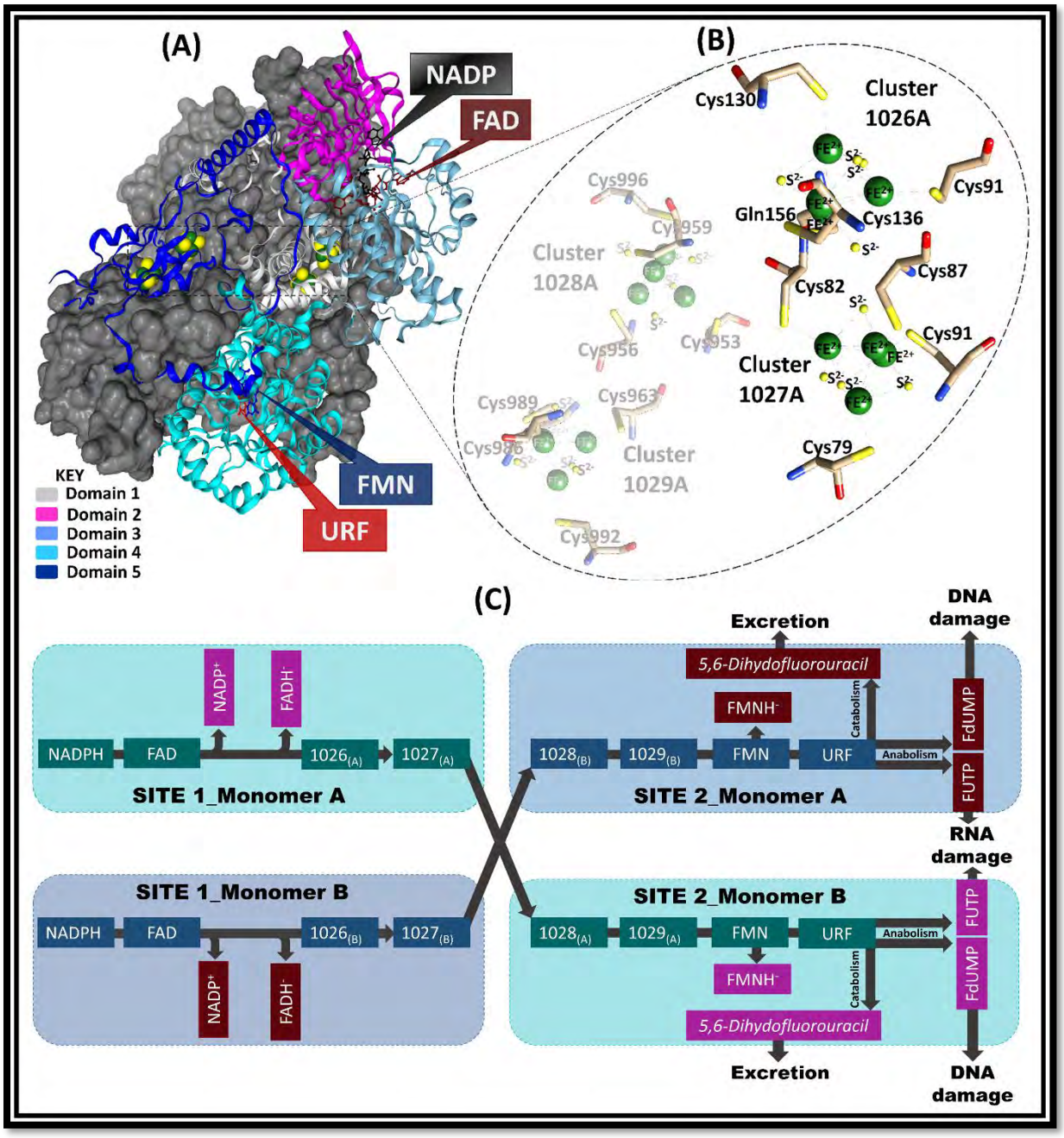


Figure 1.4: A detailed 3D representation of DPD pigs structure template (PDB ID: 1H7X). (A) Monomer A domains I – V cartoon representations are coloured in teal, magenta, light grey, light and dark blue, respectively. NADPH (black), FAD (maroon), FMN (royal blue) and red (URF) are Monomer A's stick representation of cofactors. Monomer B, the mirror image of Monomer A, is shown as a grey surface. (B) An illustration of Monomer A's $Fe^{2+}{}_4S^{2-}{}_4$ clusters coordinating environment. (C) The electron transport chains in which 2 electrons are lost from nicotinamide-adenine-dinucleotide phosphate (NADPH) in site 1 of each chain via flavin adenine dinucleotide (FAD) and $Fe^{2+}{}_4S^{2-}{}_4$ (1026 and 1027) clusters, for the reduction of URF (5-FU) in site 2 of the opposing chain through $Fe^{2+}{}_4S^{2-}{}_4$ (1028 and 1029) clusters and flavin mononucleotide (FMN). Adopted from our previous manuscript Tendwa et al., (147).

The geometry of $Fe^{2+}_4S^{2-}_4$ clusters is distorted tetrahedron cubane-like (129, 130, 157). The first (1026: Figure 4.B) cluster in each monomer has a unique coordination in which four Fe^{2+} atoms are interconnected by disulfide bridges, three of which are bonded to the protein through a cysteine residue side chain and the fourth by a glutamine residue ($[Fe^{2+}_4S^{2-}_4$ (S-Cys)$_3$(S-Gln)]) side chain. Conversely, each of the Fe^{2+} atoms in the subsequent three (1027, 1028, and 1029: Figure 4.B) clusters is coordinated by cysteine residues linked by disulfide brides ($[Fe^{2+}_4S^{2-}_4$ (S-Cys)$_4$]) (129-131, 147, 157).

1.9 Missense mutation in Dihydropyrimidine dehydrogenase (DPD)

Genetic variations in the DPYD gene can result in enzymes with low or no activity. Cancer patients with at least one copy of a non-functional DPYD variant will be at risk of metabolizing fluorouracil normally (158). Thus, these people are at risk of potentially fatal fluorouracil toxicity, such as bone marrow suppression, hand-foot syndrome, cardiotoxicities, alopecia, dermatitis, gastrointestinal toxicity, and, in rare cases, neurotoxicity (159-161). To curb the risk of toxicity, the clinical relevance of identifying patients with pathogenic DPD missense mutation is based on dose adjustment of fluoropyrimidine (5-FU, capecitabine, and tegafur) chemotherapy agents-based mutation status. Dose adjustment has been proposed to improve 5-FU tolerance and safety in patients with DPD missense mutations who are anticipated to develop severe toxicity (162, 163). This led to the genotype-guided dose adjustment (DPWG) recommendation by two main pharmacogenetics expert groups, namely: the Dutch Pharmacogenetics Working Group (DPWG) and the Clinical Pharmacogenetics Implementation Consortium (CPIC). Both consortia proposed classifying variant carriers according to their metabolizer capability (normal metabolizer, partial/intermediate metabolizer, and poor metabolizer) and agree that dose reductions of 25-50% ought to be introduced for patients who carry single variants of the four predominant Caucasian variants. For compound heterozygotes or homozygotes, extreme caution (dose reductions of 50-75% or over) and even 5-FU exclusion are advised (164, 165). Based on the DPWG's conclusion, four pathological clinically relevant variants, namely rs3918290 (DPYD *2A, IVS14+1G>A, c.1905+1G>A), rs55886062 (c.1679T>G), rs67376798 (c.2846A>T, D949V), and rs56038477/rs75017182 (c.1236G>A/HapB3), have gathered sufficient evidence to warrant their implementation in clinical care. The existing guideline exclusively presents recommendations concerning these four variants, while no recommendations are provided for other variants in DPYD as summarized in the Table 1.1 (162, 166, 167).

Table 1.1: A summary of clinically relevant DYPD variants

rs number	DPYD possible Phenotype	Activity score (AS)	DPYD variant	Consequences	Recommended dosage
rs3918290	Poor metabolizers	0: strong	DPYD*2A (DPYD*13)	No activity	Alternative therapy
rs55886062		0.5: moderate		Minimal activity	75% dose reduction
rs67376798	Intermediate metabolizers	1.0: strong	D949V HapB3 (	Reduced activity	50% dose reduction
rs56038477 or rs75018182		1.5:moderate		Reduced activity	25% dose reduction
	Normal metabolizers	2: strong	Wild type	Normal activity	Recommended dose administration

Of these variants only D949V (rs67376798) has been identified on the African descent population (168). The most prevalent missense mutation related to DPD deficiency is the IVS14 + 1 G > A, also known as DPYD*2A (rs3918290). This mutation specifically affects the splice recognition sequence located in intron 14, where a G > A alteration occurs. As a consequence, an exon is omitted during the mRNA processing, resulting in the deletion of 165 base pairs in the DPD mRNA. Consequently, a truncated protein is produced, which exhibits almost no enzyme activity. (169). Similarly, c.1679T>G (DPYD *13, rs55886062) mutation has been associated with absent DPD activity, whereas HapB3 has exhibited decreased protein function (170). Furthermore, cancer patients with M166V, Y186C, D342N, I543V, R592Q and G674D DPD missense mutations, including people of African descent, have a significantly increased risk of adverse drug reactions from 5-FU treatment (171-177). A prospective observational study examining the relationship between DPD variants and toxicity in advanced cancer patients receiving capecitabine (5-FU pro-drug), discovered that the D949V mutation was significantly associated with an increased risk of developing grade 3 toxicity (178). Lower DPD activities have been observed in D949V, M166V and Y186C variants with altered concentrations of uracil and/or its metabolite (164, 179).

Furthermore, a detailed examination of five DPD mutations found in cancer patients suffering from 5-FU-related side effects revealed a significant reduction in DPD activity in four missense mutations (C29R, M166V, S534N, I543V) (180). Mutations C29R, M166V, and I543V, as well as others (M406T, E412E, F632G, and V732I), have been isolated in the Tunisian population, but their proclivity to cause 5-FU-related toxicity was not determined (181). Nonetheless, mutation Y186C, which is notably specific to Sub-Saharan Africans, has been linked to severe toxicity cases (176). It has been shown to inhibit the activity of the DPD enzyme (176, 182). Indeed, it is proposed that this mutation be considered for inclusion

in clinical guidelines for the Sub-Saharan population after validation in clinical trials (182). Offer et al. found that variant M166V had higher enzymatic activity, whereas R592Q and D342N reported reduced DPD activity of up to 12.5% to 25% compared to the wild type. G674D, on the other hand, was the least active (<12.5%) (174).

An *in vitro* investigation of 296 volunteers from various ethnic subpopulations including Ghanaian, British-Caucasian, Kenyan, and South-West Asian displayed different DPD activities (183). Their mean DPD activities from peripheral blood mononuclear cells (PBMC) samples determined by HPLC analysis were compared within subgroups, and then against reference cohorts of Korean and French-Caucasian populations (184, 185). Similar DPD activities were observed in Kenyan, British-Caucasian, and South-West Asian subpopulations (0.194, 0.215, and 1.92 nmol/min/mg, respectively). Ghanaian subjects, on the other hand, showed significantly lower mean activity (0.119 nmol/min/mg). Considering reference cohorts, the Korean population had a significantly higher mean than the French-Caucasian population (0.227 nmol/min/mg) (185, 186). *In vitro* functional analysis using *E.coli* expression system identified two mutations (G366A and T768K) from 150 healthy Japanese volunteers. G366A mutation demonstrated reduced 5-FU intrinsic clearance (Clint), lowering DPD activities by 50%, but no change was observed in T768K. On the other hand, T768K-associated activity decreased at a faster rate than the wild type (WT) DPD, indicating protein instability (165). Using 5-FU as a substrate, Offer et al., measured the enzymatic activities of 80 non-synonymous variants expressed from HEK293T/c17 cells. DPDs with the mutations P1023T, K861R, E828K, and M166V were found to be significantly more active than wild type DPD. D949V and 31 other mutations had significantly lower activity than wild type DPD (168).

1.9.1 General computational approaches and resources to study the effects of missense mutations

1.9.1.1. Missense mutation identification

Single nucleotide variants that result in amino acid substitutions at the protein level, known as missense mutations, are a major research focus because they are linked to the majority of known genetic diseases (182, 183). Most of this information is available from the Online Mendelian Inheritance in Man (OMIM) database, which is a freely accessible public database that contains bibliographic data on human genes and genetic disorders (184). It links to curated comprehensive mutation listings, such as Human Gene Mutation Database, as well

as many locus-specific mutation databases (LSDBs) including Cystic Fibrosis Mutation Database (CFRdb), Phenylalanine Hydroxylase Locus Knowledgebase (PAHdb) and Human Genome Variation Society (185). The Genome-wide single-nucleotide variant array (GWSAs) contains specific information about missense mutations, including those related to the African descent population (186, 187). This array contains the greatest amount of linkage data for the entire genome, comprising whole-exome sequencing (WES). By employing WES sequencing, novel connections between coding polymorphisms with minor allele frequencies (MAFs) greater than 0.01 and prevalent metabolic phenotypes can be unveiled (188). Similarly, the Genome Aggregation Database (gnomAD) is a publicly available database that provides information on genetic variants found in human populations (187). The primary objective of gnomAD is to aggregate and curate genetic variant data from large-scale sequencing projects, such as the Exome Aggregation Consortium (ExAC) (188) and the Genome Aggregation Database (gnomAD) itself (187). The exome sequencing data, focuses on the protein-coding regions of genes, and genome sequencing data, which encompasses the entire genome (188). It includes information on millions of genetic variants, including single nucleotide variants (SNVs), insertions, deletions, and structural variants. Each variant entry provides details such as its genomic location, allele frequencies in different populations, functional annotations, and predicted impact on protein function (188).

ClinVar, Pharmacogenomics Knowledge Base (PharmKGB) and DrugBank are drug-related resources providing information on drug and/or related missense mutation (189). PharmKGB is a knowledge resource base which captures the associations between phenotype/diseases, genes, and drugs implicated in pharmacodynamics (PD) and pharmacokinetics (PK) (190, 191). The clinical information is incorporated from potentially clinically actionable gene-drug associations, clinical guidelines, genotype-phenotype relationships, and drug labels (192, 193). Literature reviews of gene-drug-disease interactions and genetic variant drug dosing guidelines in a pharmacogenomic approach as well as FDA label curations are a good source of this information (194). On the other hand, ClinVar combines information about the genomic variation and its association to human health (195). Some of these variants may lack 3D crystal structure, which is critical in determining the effects of mutation on protein structure and function using an *in silico* approach. As a result, they need to be modelled.

1.9.1.2. Homology modeling

Homology modeling is the most reliable computational tool for making structural models in the absence of experimentally determined crystal structure. This approach predicts the 3D structure of a query protein based on template protein sequence alignment. Target identification, sequence alignment, model construction/model refinement and model validation are the four steps involved in homology modeling as summarized in the Figure 1.5 (189, 190).

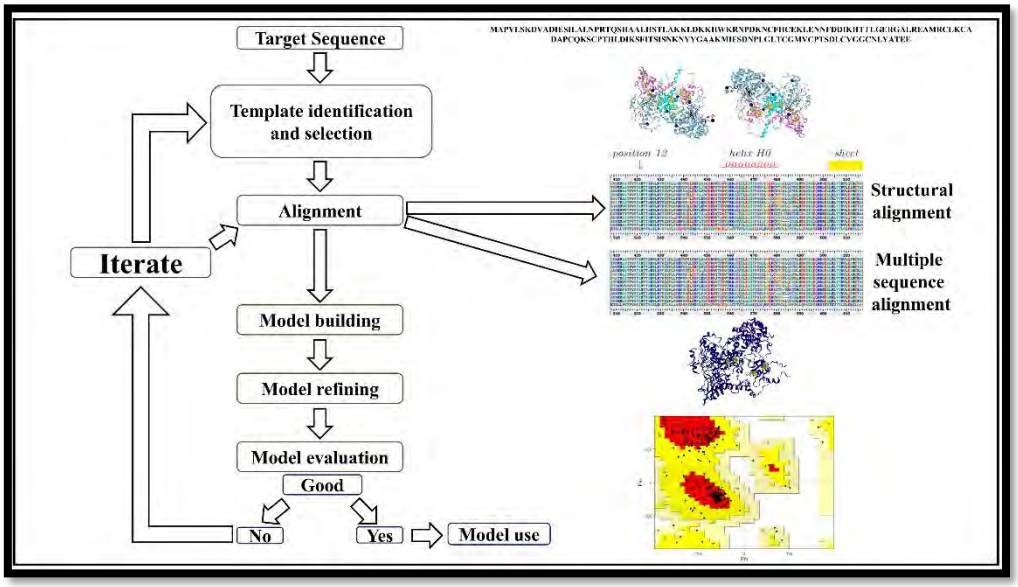


Figure 1.5: A simplified representation of the modeling process. The first step is to retrieve homologous sequences in order to create an alignment. This alignment will serve as the foundation for the model. The alignment is then fine-tuned using external data such as secondary structure information, known motifs, and conserved features. This aids in the creation of a precise alignment. The model is then built using software. This step is usually very quick (a few minutes). After the model has been built, it must be inspected and refined. In some cases, an iterative process is used in which the model is built, then inspected, and then the alignment is adjusted before rebuilding the model to include nicotinamide-adenine-dinucleotide phosphate (NADPH), flavin adenine dinucleotide (FAD) and $Fe^{2+}{}_4S^{2-}{}_4$ (model.1026, 1027, 1028 and 1029) clusters and flavin mononucleotide (FMN) and 5-fluorouracil (5-FU) Adapted from Tastan Bishop et al., (190).

Identification and selection of templates: protein databases, in addition to drug-related resources, have evolved into an essential component of identifying correct protein sequences and their related templates in modern biology. Large amounts of information are being generated for protein structures, functions, and certain sequences. Database searches are frequently the first step in the investigation of a new protein (196). Comparison of proteins or protein families provides knowledge about the relationship between proteins within a genome or across different species, and thus provides more information than studying an isolated

protein (196, 197). Protein databases such as PIBASE (198), Proteins, Interfaces, Structures and Assemblies (PDBe PISA) (199), UniProtKB/Swiss-Prot (200), Reference sequence (RefSeq) (201), Protein Information Resource (PIR) (202), and the Protein Data Bank (PDB) (203) provide diverse information regarding protein sequences, crystal structures, domains, protein-protein interactions, and post-translation modification (204). Aggregate information of the majority of protein databases can be accessed through the National Center for Biotechnology Information (NCBI) (205). Among other things, the NCBI promotes the use of software and databases to conduct research on advanced computer-based information processing approaches for analysing the structure and function of biologically important molecules such as proteins (196, 205). The sequences for proteins and how they are related to other proteins can be obtained with blast local alignment search tool (BLAST). Furthermore, databases such as HHPred (191) and PRotein Interactive MOdeling (PRIMO) (192) are popular simplified tools for finding eligible templates for target sequences. In this study, PRIMO was used; it searches for templates in a PDB (193) containing NCBI databases via BLASTp algorithms. The HHSuite packages, which include HHSearch (194), explore databases such as the PDB for suitable templates. It is critical to obtain a template with high sequence similarities to the target sequence in order to generate highly accurate 3D structures. A minimum similarity of 25% between the template and target sequences has been determined to be suitable for homology modeling (195). Aside from high sequence similarity, environmental factors like solvent type, pH, and the presence of bound substrates, as well as the experimental structure resolution, all play a role in selecting an eligible template (196).

Multiple sequence alignment (MSA): following the identification of the eligible template, the target and template sequences are aligned, a critical step in homology modeling. MSA has become a cornerstone of *in silico* studies in modern molecular biology, ranging from evolutionary research to the prediction of 2D/3D structures, intermolecular interaction, and molecular functions (197, 198). In placing the sequence in the context of the entire family, MSA can be used in identifying conserved features and highlighting their specificities and differences (198). Essentially, three or more related sequences are matched as closely as possible. MSA's goal is to arrange a set of sequences so that as many characters from each sequence as possible are matched using the same scoring function. Aligning a set of sequences can be challenging and time consuming. Thus, computational algorithms such as heuristic or dynamic programming methods including: ClustalW (199), T-Coffee (Tree based Consistency Objective Function For alignment valuation) (200, 201), MUSCLE (MUltiple Sequence Alignment by Log-Expectation) (202), MAFFT (Multiple Alignment using Fast

Fourier Transform) (203), and PROMALS3D (PROfile Multiple Alignment with predicted Local Structures) as well as 3D constraints are used to generate and analyze this alignment (204). In this study, the sequence identity of the query to the template protein was 93%, therefore most of the aforementioned processes were not necessary.

Model creation: some of the computational programs invented for model creation include SWISS MODEL (205), MODELLER (206), PRIMO (192), and i-TASSER (207). Herein, MODELLER was used to set up a protein model by performing comparative modeling on the target and template sequences provided. Besides being used for protein optimization and loop modeling, the program calculates the nonhydrogen atoms to create a model (206, 208). It satisfies spatial constraints by either employing distances or optimizing techniques. In addition, this program includes the CHARMM energy terms that ensure valid stereochemistry when combined with spatial constraints (190). Afterwards, the model's characteristics are verified by reaching a consensus among various metric programs. The initial evaluation of the protein structure is then conducted using the normalized Discrete Optimized Protein Energy (zDOPE) score (190). Structures with zDOPE scores lower than -0.5 are indicative of wild type-like structures.

Model validation: the accuracy of the built models will determine their future applicability in various fields. As a result, further validation of the built models was done via PROCHECK (209), ProSA (210), and ERRAT (211). The PROCHECK web server (209) defines stereochemical properties of previously crystallized protein structures to determine the geometric quality of protein structures. Hence, a Ramachandran plot describes: (i) most preferred, (ii) generally allowed, (iii) additionally allowed, and (iv) disallowed regions. Individual residues are assigned to each region depending on the torsion angles' (φ – ψ) favorability (209, 212). Another validation program, ProSA (210), generates an energy plot of all residues in order to identify potentially problematic plots of a structure. A calculated statistical score (z-score) comparing a range of z-scores of the same sized proteins to those determined experimentally is also provided (210). Furthermore, ERRAT (211) is a non-bonded atomic interaction "overall quality factor" that categorizes atoms as carbon, hydrogen, or oxygen. It then compares the distribution of non-bonded interactions between atoms in an individual structure to other reliable, high-resolution structures from a database (211). Where the generally accepted range is greater than 50, higher quality models are scored higher. The higher the value, the better the quality (213).

Structure prediction with AlphaFolds: The recently released neural network-based model, AlphaFold (214), represents a computational method capable of consistently

predicting protein structures with atomic precision, even in instances where there is no existing similar structure. It integrates cutting-edge neural network architectures and training methods that leverage evolutionary, physical, and geometric principles inherent in protein structures. Advancements, such as a novel architecture that simultaneously incorporates multiple sequence alignments (MSAs) (215) and pairwise features, a refined output representation and associated loss function that ensures precise end-to-end structure prediction, an innovative equivariant attention architecture, iterative refinement of predictions through the utilization of intermediate losses, joint training of masked MSA loss with the structure, and the ability to learn from unlabeled protein sequences via self-distillation and self-estimated accuracy. Employing the primary amino acid sequence and aligned sequences of homologues as inputs, the AlphaFold network directly forecasts the 3D coordinates of all heavy atoms within a given protein (214).

1.9.1.3. The structural and functional effect of mutations

MUpro (216), I-Mutation (217), and PhD-SNP (218) are web-tools used in the prediction of the structural and functional effects of mutations on proteins. I-Mutation 2.0 (68) predicts protein stability by calculating the direction of changes in protein stability as well as the energy values associated (G). If the variations have no effect on stability, they are classified as neutral. PhD-SNP, on the other hand, predicts the deleterious effect of SNPs in both coding and non-coding regions of a protein structure, categorizing each SNP as either neutral or deleterious (218). MUpro equally engages SVM and neural network approaches. With a confidence score ranging from -1 to 1, the output of this tool describes the changes in energy (G) caused by residues substitution. A variation with a score less than 0 reduces the stability of the protein (216). The three approaches mentioned have been incorporated into the web servers for Variant Analysis Porta (VAPOR) (219). This is a workflow tool for single nucleotide polymorphism (SNP) analysis which has been integrated into the HUMA platform (219). Other machine learning computational tools like Rhapsod*y* (220) have also, been utilized in predicting the pathogenicity of human missense mutations. This approach integrates dynamics-dependent qualities such as local residue flexibility and other allosteric and mechanical parameters derived from the protein's structural dynamics using the Elastic Network Model (ENM) (221).

1.9.1.4. Metal ion Parametrization

Metalloenzymes play critical roles in cellular processes, constituting nearly 30-40% of the proteome (222). Iron, copper, zinc, and magnesium are among the most abundant metal ions found in enzymes. Iron, for example, forms clusters within the DPD enzyme. These clusters coordinate a redox reaction across the protein, activating 5-FU degradation in the active site (147, 223-225). This process is hampered by DPD missense mutations, resulting in 5-FU drug toxicity (189). MD simulation studies are an integral tool in deducing the molecular mechanism of action of proteins. To conduct this investigation, force field parameters for the Fe^{2+} centers coordinating environment has to be determined. Computational studies, particularly molecular dynamics (MD) simulation, can provide insights into metal coordinating environments. MD calculations, on the other hand, are heavily reliant on force fields derived via quantum mechanics (QM) and molecular mechanics (MM) approaches (226, 227). Classical-type models are used in the MM method to predict the amount of energy in a molecule based on its conformation (228). This method is computationally less expensive and sufficient for describing atomic interactions and dynamics in a purely organic system relative to QM approaches (229). Furthermore, *de novo* QM/MM calculations have been used to describe the precise electron structure of atoms around a metal center to cater for the electronic effects of metals (224, 230, 231). Since metals are needed in protein function, novel force field parameters for defining key transition metal architectures are being developed (232). This has resulted in several modified force fields which have been integrated into a number of force field families, including Gronigen molecular simulation (GROMOS), optimized potentials for liquid simulations (OPLS-AA) (233, 234), chemistry at Harvard molecular mechanics (CHARMM) (235, 236), and assisted model building with energy refinement (AMBER) (234). The latter two have been extensively used. They offer a large palette of atom types, permitting numerous organic molecules to be described by assigning atom types based on chemical similarity (237).

OPLS-AA (238, 239) optimizations concentrate on the small molecule condensed phase properties and have since been extended to include a diverse set of small molecule model compounds. Atom type assignment, on the other hand, must be done manually. It is interesting to note that a commercial OPLS-AA implementation with atom typing functionality is available (240).

CHARMM, instead, has been enhanced with the CHARMM general force field (CGenFF), that incorporates not only a wide range of chemical groups present in biomolecules and drug-like molecules, but also many heterocyclic scaffolds (241-243). A

web interface for automatic atom typing as well as analogy-based parameter and charge assignment are also now available (244).

The GROMOS force field atom type palette encompasses a broad pool of options for building small molecule models with force fields sourced from biopolymer parameters (233). AMBER (233, 245, 246) now comprises of the general AMBER force field (GAFF) (247) and the antechamber toolkit, which assists users to generate an AMBER (234, 248) force field model for any input molecule. Barring the related simulation speeds and exportable parameters, the advancement of a Python-based metal parameter builder (MCPB.py) (249) that also supports various AMBER force fields and >80 metal ions has made it easier to parametrize inorganic constituents in proteins. With these benefits, AMBER is the most popular platform for developing metal parameters for use in metalloprotein simulations (250). To characterize such proteins, various methods including the polarization model and non-bonded (246), semi-bonded (251), and bonded models (252) have been used. Metal coordinating environments are defined by their van de Waals interactions in the non-bonded approach (21, 246). It is limited by the long-range nature of electrostatic forces which can cause the metal to escape from its coordinating centers or ignore its correct connectivity (253, 254). The semi-bonded approach entails surrounding the metal with dummy atoms that are covalently connected to it based on the metal's predefined coordination geometry (255). This method has the same limitations as the non-bonded approach (256). Consequently, the bonded model has been used to overcome the limitations of other approaches because it can preserve the metal in its correct coordination geometry by considering the metal's connections covalently (257). Parameters in the bonded approach can be obtained experimentally or calculated quantum mechanically at a high level of theory (258). It employs defined harmonic energy terms that have been introduced into possible energy functions to account for the formation of bonds between atoms and metal centers (252, 259, 260), as proposed by Seminario et al., (261). The method is a popular metal parametrization technique that has been automated in various parametrization programs such as Visual Force Field Derivation Toolkit (VFFDT) Seminario (250).

VFFDT is a user-friendly automated software based on the Seminario method. It permits the user to generate bond angles bending and bond stretching force field parameters for metalloproteins. In addition, the software assists the user in customizing, viewing, copying, and pasting the force fields in an alternative manner. MCPB.py, MCPB, and ANTECHAMBER in AmberTools are function complements to VFFDT (250). The

developed human model for the Fe^{2+} center in DPD protein included a new set of AMBER force field parameters defined by equation 1. The molecular system describes the AMBER total potential function (V_{total}) as a sum of non-bonded (Coulomb and Lennard-Jones) and bonded (bond, angles, and dihedrals) terms.

$$V_{total} = V_{bond} + V_{angle} + V_{dihedral} + V_{Lennard\text{-}Jones}\ V_{total}$$

$$V(r^N) = \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} \sum_n \frac{1}{2} V_n[1 + cos(n\omega - \gamma)] + \sum_{j=1}^{N-1} \sum_{i=J+1}^{N} f_{ij} \left\{ \in_{ij} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0ij}}{r_{ij}} \right)^{6} \right] + \frac{qi\, qj}{4\pi\epsilon_0 r_{ij}} \right\}$$

Equation 1: AMBER Force Field Potential Energy Function

Where k_b and k_a are the force constants used to restore their respective bond angles and lengths to their equilibrium values. The force constant k_Θ defines the height of the dihedral rotation barrier; the equilibrium values b_0 and Θ_0; n defines the number of minimums, V_n defines the dihedral angle; γ is the phase angle; rij defines the distance between atoms i and j; ε is the potential well depth; qi and qj are the atoms' point charges; and Є0 is the permittivity in vacuu

1.9.1.5. Molecular Dynamic Simulations

The primary goal of simulation is to examine the molecular system in terms of the structure and interactions of the molecules involved. There are numerous simulation techniques, but the two most basic are simulation Monte Carlo (MC) and molecular dynamics (MD) simulation; all other techniques are hybrids of these two (262). MD simulation technique uses classical Newton's equation of motion to analyze the trajectories, movements, and interactions in an atomically modeled molecular system (263, 264). These atoms and molecules are allowed to run for a set amount of time, and their potential energies and forces within the molecules are analyzed using molecular mechanics force field interpretations (59).

1.9.1.6. Post-Molecular Dynamic Analysis

Mutations and ligand binding generate dynamics that affect global effects, or consequences that extend far beyond the site of the perturbation. As a result, local dynamics

(perturbations) of a protein can produce global changes that cause the protein to: i) lose function (disease), ii) acquire new or recovered function (adaptability), or iii) maintain its function (robustness) (59, 265, 266). Therefore, the combination of global and local analysis gives a better understanding of the effect of mutations on protein function.

I. Global MD analysis

Different global analysis approaches such as root mean square deviation (RMSD), framewise RMSD, root mean square fluctuation (RMSF), radius of gyration (Rg), dynamic cross-correlations (DCC), hydrogen bonds (H-bonds), and comparative essential dynamics have been used to depict dynamics of a protein.

Root mean square deviation (RMSD): the RMSD between pairs of equivalent Cα atoms computed after optimal superposition of the two structures is commonly used to assess the similarity of two protein three-dimensional structures. This is done when two conformations of the same protein (native, mutation, bound/unbound, monomeric/oligomeric, etc.) are compared, or when two different proteins with different amino acid sequences are compared, though the equivalencies between pairs of Cα atoms may be defined or discovered differently depending on the degree of similarity between the two proteins that are compared. For instance, the smaller the RMSD, the more similar the two structures and vice versa. The use of the RMSD may appear unusual, given that many other similarity measures have been proposed and used in various fields of structural and molecular biology (267-270). The calculation is dependent on the simulation's final and original reference point, as well as all successive frames. RMSD is defined as follows:

$$\boldsymbol{RMSD} = \sqrt{\frac{\mathbf{1}}{\boldsymbol{N}}\sum_{\boldsymbol{i-1}}^{\boldsymbol{N}}\left(\boldsymbol{r}_{\boldsymbol{final\,(i)}} - \boldsymbol{r}_{\boldsymbol{initial\,(i)}}\right).\left(\boldsymbol{r}_{\boldsymbol{final\,(i)}} - \boldsymbol{r}_{\boldsymbol{initial\,(i)}}\right)}$$

Equation 2: RMSD Function

In which N is the total number of atoms and $r_{initial(i)}$ and $r_{final(i)}$ are the initial and final structures of their respective present atom coordinate i.

Root mean square fluctuation (RMSF): the RMSF is a measure of the average movement of atomic (e.g. protein residue) positions over time. It can be measured by capturing the total positional displacement of an atom for each conformation generated and then averaging it by the number of conformations generated (271). Frequently, RMSF

calculates whether a structure is stable in simulation time or diverging from the initial coordinates. Divergence from the initial coordinates is usually interpreted as a sign that the simulation is not equilibrated. However, at equilibration the structure fluctuates around a stable average condition. At this point the fluctuation of each subset of the structure (e.g. each atom) relative to the simulation's average structure is calculated (272-274).

Radius of gyration (Rg): the protein Rg is the root mean square distance between each atom in a protein and its center of mass. The distance between the rotating point and the point at which the energy transfer has the greatest effect gives Rg. If a protein is folded stably, it will most likely maintain a relatively constant value of Rg, whereas if a protein unfolds, Rg will change over time (275, 276). Rg is determined as follows:

$$\boldsymbol{Rg} = \sqrt{\frac{\mathbf{1}}{\boldsymbol{N}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}}(\boldsymbol{r}\,(\boldsymbol{i}) - \boldsymbol{r_{center}}).(\boldsymbol{r}\,(\boldsymbol{i}) - \boldsymbol{r_{center}})}$$

Equation 3: Radius of gyration Function

For which *N* denotes the total number of protein atom coordinates and *r (i)* is the atom *i*. r_{center} represents the center of mass (COM) of the whole protein structure *i*.

Dynamic cross-correlation (DCC): DCC is an approach for analyzing residue motions during simulation. It evaluates backbone motions using C^{β} (C^{α} for Gly) atoms (277). This technique has the drawback of ignoring side chain motions, which contribute to the multimodal motions of the residues. However, the introduction of multimodal DCC (mDCC) has addressed this issue (277, 278). Extraction of dynamical pairwise correlations and identification of key residues from large molecular dynamics trajectories is critical in explaining processes such as ligand binding, mutational effects, and long-distance interactions (278). Despite this progress, most researchers continue to use the unimodal DCC in trajectory analysis, with the value for each cell in the correlation matrix (C_{ij}) for the i^{th} and j^{th} residue calculated using the formular below:

$$Cij = \frac{(\Delta ri \Delta rj)}{\left((\Delta r_i^2)(\Delta r_i^2)\right)^{1/2}} \ldots$$

Equation 4: Calculation for DCC

Residues *i* and *j*'s spatial positions are represented by r_i and r_j, respectively, and their displacement from their original position by vectors Δr_i and Δr_j (278).

Essential dynamics (ED) based on PCA: ED is a process that includes principal component analysis (PCA) to extract biologically plausible motions from atomic trajectories of proteins (279). MD trajectories generate collective motions that are segmented into components (principal components), which together represent the functional dynamics of the protein structure (280). Once all the ensemble configurations are superposed, the internal motion from overall rotation and translation is filtered during this analysis. This is accomplished by comparing each configuration to a reference structure and determining the configuration with the least squares fit (281). The 'fitted' trajectories are then used to build the variance-covariance matrix, which is then diagonalized. The diagonalization of the covariance matrix yields a collection of eigenvectors and eigenvalues, with the eigenvalues ordered decreasingly (282). As a result, the principal mode represents the atom's collective motion, with the corresponding eigenvalue representing its magnitude and the eigenvector its direction (281). Amamuddy and colleagues (283, 284) have introduced a unique comparative essential dynamics approach to evaluate the distribution of conformational sampling across MD simulations for the WT and mutations. This approach, which employs the "compare essential dynamics.py" tools, has been incorporated into the MDM-TASK-web (283, 284). It involves data filtering and alignment of all MD frames to a common reference (using C-α atoms), prior proceeding to a single decomposition. Thereafter, the regions of interest in the WT and mutation frames are aligned to a single WT simulation frame, which is used to compute and decompose the covariance matrix (285-287)

II Local MD analysis

In the field of protein analysis, several approaches such as dynamic residue network (DRN), contact maps and contact heat maps have been developed to investigate local properties and interactions within protein structures.

Dynamic residue network (DRN): DRN analysis algorithms evaluates the patterns of change and conservation of key nodes according to network centrality (*BC, CC, DC, EC,* and *KC*) independent criteria (283). Each residue Cβ (Cα for Gly) is treated as a node in this approach, and the inter-node proximity is used to define interactions between residues.

NetworkX (288) internally uses a N x N adjacency matrix (where N is the number of residues) from the resulting DRN graph to compute the following metrics at each time frame: *betweenness centrality (BC), closeness centrality (CC), degree centrality (DC), eigenvector centrality (EC), and Katz centrality (KC). BC* stipulates the importance of a residue based on its usage for protein communication. It is proportional to the number of shortest paths which pass through that node across all vertices to the rest.(283). Conversely, by measuring how close a node is to all other nodes in a network, *CC* determines how quickly information flows through it (287). Moreover, the total number of unique connections to a node is defined as *DC* (287, 289). Hence, the more edges it has, the more central it is locally. *EC* calculates a node's level of influence within a network by highlighting the significance of node neighbors recursively (283, 287). *KC*, like *EC*, calculates a node's relative impact within a network by successively gauging node centrality based on the centrality of neighboring nodes (287). *BC, CC, DC, EC* and *KC* are determined as follows (Table 1.2).

Table 1.2: Dynamic residue network centrality metrics and their formulas (283, 284)

Centrality metric	**Formula**	**Implication**
BC **Equation 5**	$\frac{1}{BC(V)} = \frac{1}{M}\sum_{i=1}\sum_{s,t\in v}\frac{\delta(siti/vi)}{\delta(si,ti)}$	In which V defines the entire set of nodes, m denotes the number of frames, and d (s, t|v) shows the number of shortest paths among nodes *s* and *t* that pass through another node *v*. *d(v, u)* gives the shortest distance between u and *v*, where *n* is the number of nodes in a given graph.
CC **Equation 6**	$CC\ (x) = \frac{1}{\sum_y d(v,u)}$	In which $d(v,u)$ defines the total distance between all other residues *u* and residue *v*
DC **Equation 7**	$DC(K) = \frac{1}{m(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ijk}$	Where n signifies the number of residues, m denotes the number of frames, and A_{ijk} represents adjacency in time frame *i*, with 1 specifying adjacency if residues having indices *j* and *k* are adjacent, and 0 alternatively.
EC **Equation 8**	$EC_{ik} = \in_k^{-1}\sum_t^n =_1 A_{ijk}\,.\,EC_{jk}$ (a) $EC_i = \frac{1}{m}\sum_{k=1}^{m} EC_{ik}$ (b)	(a) The i[th] factor (residue) of the eigenvector (defined by power iteration) comparable to the k[th] frame is denoted by ECik. is the eigenvalue

		of the k^{th} frame's eigenvector. (b) Divide the number of frames by the average *EC* for i^{th} residue.
KC **Equation 9**	$KC = \alpha \sum_j A_{ij} X_j + \beta$	For the k^{th} frame, A_{ij} is the adjacency between nodes *i* and *j*. The parameters α and β correspond to the attenuation factor and weighting parameter for the centrality, whereas x_j represents to the *jth* component of the *eigencentrality* vector at frame k.

m = Frame numbers

Weighted residues contact map and contact heatmaps: These are a representation residues inter-relationship in proteins. In their 3D structure, residue contact maps are described as a pair of spatially proximal residues. A PDB file represents these residue coordinates in the form of x, y, and z (3D structure) and thus contacts can be defined using a distance threshold (290). A pair of residues are in contact if the distance between their specific atoms (usually carbon-alpha (Cα) or carbon-beta (Cβ)) is less than a distance threshold (typically 7 Å) (291). Contact maps can be used to visualize the interaction frequency between the residue of interest and other proximal residues throughout the MD simulation. This information can then be aggregated using a contact heatmap (284).

1.10 Problem statement

The International Agency for Research on Cancer predicts that the number of new cancer cases in Africa will double over the next decade (20, 21). However, Africa lacks adequate infrastructure to support research in the unprecedented cancer landscape that will allow for a better understanding of the disease and the development of informed treatment plans (20, 21). According to the GLOBOCAN 2021 (21), an estimate of 110,000 new cases of cancer was diagnosed in South Africa in 2020, with more than 56,000 cancer-related deaths, accounting for one-quarter of all premature noncommunicable disease-related mortality (292). Patients undergoing chemotherapy have varying treatment outcomes in terms of tumour response, as well as incidence and severity of adverse effects. Fundamentally, the causes of these differences are multifactorial, but genetic factors play an important role (293). Currently, most cancer medications are manufactured in developed countries and are used in

Sub-Saharan African populations without further testing. As a result, these drugs may be ineffective or cause negative side effects (186). Even though variability in the toxicity and efficacy of these drugs has been observed in the African ancestry population, the cause of this population variation is only partially understood (186, 294).

1.11 Research Aim and Objectives

Aim

The primary aim of this study was to investigate the pharmacogenomic impact of DPD missense mutations in African descent subpopulations that contribute to dysfunctional 5-FU metabolism. The study is arranged into two major sections focusing on:

1. Generating force field parameters for the Fe^{2+} center environment of DPD proteins, since existing force field parameters could not accurately describe the Fe^{2+} center coordination displayed by this enzyme. This would have posed a challenge for the subsequent *in silico* studies.
2. Deciphering the mechanism of action within DPD missense mutations and wild type. We used a variety of *in silico* approaches and tools developed recently by our research group along with existing computational tools.

The missense mutations were divided into two categories: (a) clinically significant mutations; and (b) non-clinical mutations (mutations that are not clinically validated yet). Based on the non-clinical mutations, the effect of (i) allosteric (remote) and (ii) catalytic domain mutations in dysfunctional metabolism of DPD was determined. The obtained information will be explored further for the purpose of improving treatment outcomes by identifying novel hits for personalised tailored precision cancer drugs. Furthermore, these studies may help to elucidate the molecular mechanism of action of missense mutations in other drug metabolizing enzymes in relation to the WT. Hence, guidelines for implementing precision medicine based on drug resistance mutations can be developed. The aims were divided into the following objectives:

Objectives

1. To obtain DYPD variation data on African descent population genetics from various databases or literature, and to identify mutations related to the DPD protein.
2. To retrieve DPD sequences for each mutation and filter them through various predicting tools before identifying mutations based on predicting tool consensus.

3. To model human WT through homology modeling using pig's crystal structure as a template and DPD Fe^{2+} centers parametrization, as well as parameters validation to maintain the metals within their coordinating environment. This was done for both drug (5-FU)-free and drug bound models.
4. To model the missense mutations based on the calculated WT template and transfer the previously generated force field parameters (WT) to each of them. This was done for both drug free and drug bound models.
5. To conduct molecular dynamic simulations to assess the behaviour of the WT in relation to the mutations.
6. To separate the mutations based on clinical and non-clinical data, as well as further sub-divide non-clinical data to allosteric and catalytic-domain mutations.
7. To perform global analysis of the simulation (RMSD, RMSF, Rg, DCC, comparative essential dynamics, and hydrogen-bonds) to highlight the global conformational changes of both the drug free and drug bound systems of the mutations in respect to the WT.
8. To perform local analysis of the simulation (DRN multi-metric analysis: *BC, CC, DC, EC,* and *KC*) to pinpoint the behaviour of the mutations in their drug free and drug bound states at residual level.

Contribution: 3D structures of DPD WT and 19 missense mutations were generated by **Maureen Bilinga Tendwa.** Generation of AMBER Fe^{2+} Force Fields Parameters for the WT using original Seminario method was done by **Maureen Bilinga Tendwa**, Lorna Chebon-Bore and Thommas Mutemi Musyoka. On the other hand, the generation of AMBER Fe^{2+} Force Fields Parameters using collation features Visual Force Field Derivation Toolkit was performed by **Maureen Bilinga Tendwa** and Kevin Lobb. Inference of the generated force field parameters and MD simulations for the WT and the 19 mutations of drug free and drug bound ensembles were done by **Maureen Bilinga Tendwa.** All global analyses were done by **Maureen Bilinga Tendwa**, as well as local analyses. However, Victor Barozi performed the DRN calculation for the identification of *BC, CC, DC, EC,* and *CC* hubs via global top 4% calculations using the script written by Olivier Sheik Amamuddy. Finally, all the paper drafts were written by Maureen Bilinga Tendwa.

CHAPTER 2
DETERMINATION OF AMBER FORCE FIELD PARAMETERS FOR THE IRON CENTER PARAMETRIZATION OF DIHYDROPYRIMIDINE DEHYROGENASE

The primary focus of this chapter is on modeling DPD WT and generating AMBER force field parameters for the protein's Fe^{2+} ion centers. The obtained parameters are integral part of the subsequent studies using MD in order to unravel the molecular mechanism of DPD protein, which metabolizes the anti-cancer drug 5-FU.

After obtaining proper authorization, Chapter 2 was founded on the published information listed below.

Tendwa MB, *Chebon-Bore L, Lobb K, Musyoka TM, Tastan Bishop Ö. Force Field Parameters for* $Fe^{2+}{}_4S^{2-}{}_4$ *Clusters of Dihydropyrimidine Dehydrogenase, the 5-Fluorouracil Cancer Drug Deactivation Protein: A Step towards In Silico Pharmacogenomics Studies. Molecules. 2021 May 14;26(10):2929.* ***https://doi.org/10.3390/molecules26102929***

Authorship contribution: The conceptualization, study design and experimental procedure were supervised by Özlem Tastan Bishop and Thommas Mutemi Musyoka. Data retrieval was performed by **Maureen Bilinga Tendwa**, AMBER Force Field Parameter calculation and analysis of results by original Seminario method was performed by **Maureen Bilinga Tendwa** under the mentorship of Lorna Chebon-Bore, AMBER Force Field Parameter calculations and analysis by collation features Visual Force Field Derivation Toolkit was done by **Maureen Bilinga Tendwa** under the supervision of Kevin Lobb. **Maureen Bilinga Tendwa** wrote the final draft.

2.1 Introduction

The dimeric dihyropyrimidine dehydrogenase (DPD) metalloenzyme that metabolizes 5-fluorouracil (5-FU) has highly specialized 4 × $Fe^{2+}_4S^{2-}_4$ clusters per monomer (129, 130). These clusters catalyze the rate-limiting step in the pyrimidine degradation pathway via a harmonized electron transfer cascade that initiates a redox metabolic reaction (131). The therapeutic potential of 5-FU is largely dependent on the catabolism of a small percentage (1-3%) of the administered dose to active fluorodeoxyuridine monophosphate (FdUMP) and fluorouridine triphosphate (FUTP) metabolites in the liver. Both RNA and DNA synthesis are inhibited by the process (262, 295). Conversely, the DPD enzymes deactivate 80-85% of the dose to dihydroflourouracil (DHFU) via anabolism, with the remaining 12-19% excreted in the urine unmetabolized (295, 296). The presence of DPD mutations in the general population, including the African descent subpopulation, has toxic consequences (21, 144). Deciphering the effect of these mutations on the catalytic mechanism via *in silico* studies such as molecular dynamic simulation can thus greatly improve oncology treatment protocols (297). It is challenging to conduct such computational studies due to the presence of $Fe^{2+}_4S^{2-}_4$ clusters.

The homodimeric DPD is a 222kDA protein with 1015 residues per monomer. Each monomer has five distinct domains: domain I (residues 27-172 and $Fe^{2+}_4S^{2-}_4$ clusters), domain II (residues 173-286, 442-524, FAD binding), domain III (residues 287-441, NADPH binding), domain IV (residues 525-847 and pyrimidine binding), and domain V (residues 1-26, 848-1025 and $Fe^{2+}_4S^{2-}_4$ clusters) (168). Pyrimidine binding domain IV functions as the 5-FU drug recognition and catalytic core domain. The remaining domains are implicated in redox reduction, with $Fe^{2+}_4S^{2-}_4$ cluster domains acting as a bridge between FMN and FAD cofactors for electron transport to the active site (131, 152, 153, 156). This is due to the proximity of the domain IV 2 × $Fe^{2+}_4S^{2-}_4$ (hetero atoms 1028 and 1029) clusters of Monomer A to the domain V 2 × $Fe^{2+}_4S^{2-}_4$ (hetero atoms 1026 and 1027) clusters of Monomer B (152, 156). Clusters 1027, 1028, and 1029 share a coordination of cysteine residues linked by disulfide bridges ($[Fe^{2+}_4S^{2-}_4(S\text{-}Cys)_4]$). A unique coordination, however, is observed in cluster 1026, where the Fe^{2+} atoms are linked by disulfide bridges, three of which are bonded to the protein backbone via a cysteine residue side chain and the fourth through a glutamine residue ($[Fe^{2+}_4S^{2-}_4(S\text{-}Cys)_3(S\text{-}Gln)]$) side chain (129, 131).

To accurately predict the behavior of a protein over time using MD simulation, potential energy functions known as force fields are required (226). The majority of available

force fields can be applied in the structural and mechanistic features of proteins in the absence of non-protein components such as ligands or cofactors. However, since they do not account for polarization and ligand metal charge transfer effects, these pairwise-additive force fields cannot be used for metalloproteins (130, 224, 298). To factor in the metals' electronic effects, de novo QM/MM calculations were used to define the precise electron structure of atoms surrounding a metal center (224, 230, 231). New force field parameters are being developed to describe various transition metal architectures for metalloproteins (258). This has resulted in a slew of modified force fields being incorporated into various force field families, such as assisted model building with energy refinement (AMBER) (234). AMBER is widely used because it offers a diverse set of atom types. This enables the representation of a number of organic molecules by assigning atom types based on chemical similarity (237, 299). It includes the antechamber toolkit and the general AMBER force fields (GAFF) (247). Hence, a user is able to generate an AMBER force field for any input molecule (234, 248). Apart from the corresponding simulation speeds and exportable parameters, the emergence of a Python-based metal parameter builder (MCPB.py) (249), which supports multiple AMBER force fields and more than 80 metal ions, has simplified the parametrization of inorganic constituents in proteins. These precedents make AMBER the most popular platform for developing metal parameters for use in metalloprotein simulations (245). Studies have widely applied this approach in characterizing Fe^{2+} centers in a number of different metalloproteins (252, 259, 300, 301). For instance, Carvalho and colleagues (259) successfully generated AMBER force field parameters for $Fe^{2+}{}_4S^{2-}{}_4$ clusters coordinated by cysteine residues. There were no parameters associated with the coordination of glutamine residues to the Fe^{2+} center, nor did they generate parameters for the structures of composite multiple clusters in any of these studies.

In this study, force field parameters for Fe^{2+} centers in DPD protein were computed via the bonded method of QM and Seminario techniques (258). The density functional theory (DFT) of the QM technique, in particular, was used to generate Fe^{2+} center AMBER parameters for two models utilizing different Seminario methods. The first method (viz. Model 1) used the original Seminario (260) method (249), whereas the second method (viz. Model 2) used collation features Seminario (250) Visual Force Field Derivation Toolkit (VFFDT). The parameters from the two approaches were compared, and their reliability was assessed using all atom MD simulations.

2.2 Materials and Methods

Figure 2.1 depicts a visual summary of the procedures employed in this study.

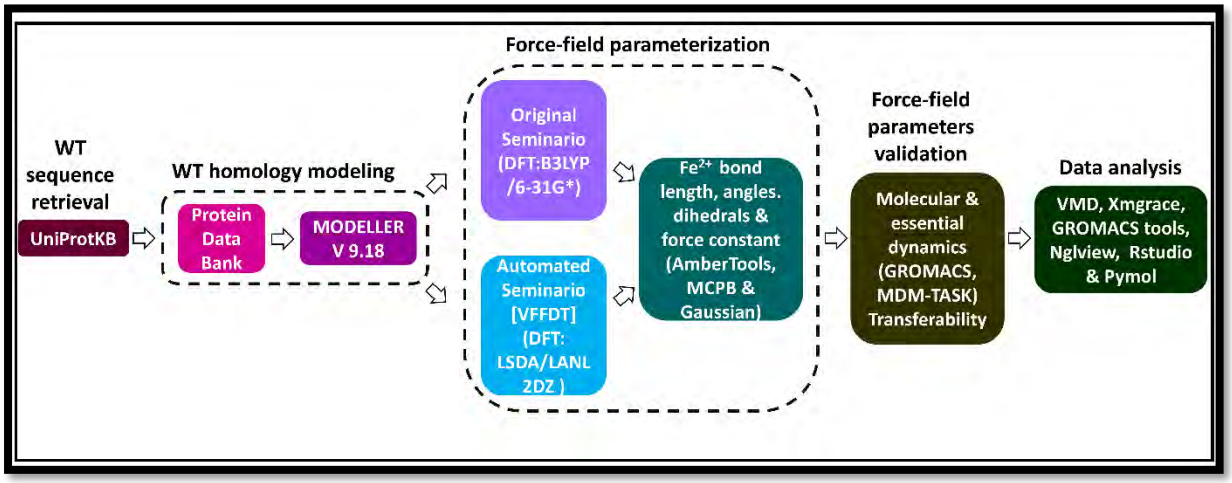


Figure 2.1. Shows a flow diagram that summarizes the methods and tools used to generate and validate $Fe^{2+}{}_4S^{2-}{}_4$ force field parameters for a human DPD protein model.

2.2.1 Software

AMBER and AmberTools17, University of California, San Francisco, CA, USA; Discovery Studio v4.5, Dassault Systems BIOVIA, San Diego, CA, USA; GROMACS v2018.2, University of Groningen, Uppsala Sweden; R v3.6.1, RStudio v1.1.456, R Core Team, Boston, MA, USA, GaussView 5.0.9, Carnegie Mellon University Gaussian, Wallingford, Connecticut, USA; PyMOL Molecular Graphics System, v1.8.2.3 Schrödinger, New York, NY, USA; MODELLER, University of California, San Francisco, CA 94143, USA; MD-TASK v1.0.1, Research Unit in Bioinformatics (RUBi), Rhodes University, Makhanda, South Africa; and Maestro v12.5, Schrödinger, New York, USA.

2.2.2 Human DPD protein homology modeling

The pig crystal structure of Dihyropyrimidine dehydrogenase (PDB 1H7X) was used to obtain the complete geometry of the initial structure (129, 130). However, due to the lack of human DPD X-ray crystallographic data, a homolog 3D structure was built. The human DPD enzyme residue sequence was first obtained from UniProt (UniProt accession: Q12882) (302) database. Subsequently, PRIMO (192) and HHpred (191) were both used to find a good template for modeling the human DPD protein. The selected template guided the homology modeling of 100 human DPD holo models at the "very-slow" refinement level using

MODELLERv9.15 (303). The resulting models were ranked based on their lowest normalized discrete optimized protein energy (z-DOPE) score, with the top three models being chosen for further modeling. To incorporate the non-protein structural information, each of the chosen models was superimposed on the template separately in Discovery Studio 4.5 (304), and all non-protein information was copied. The drug and cofactor coordinates were then directly transferred to the modeled structures. VERIFY3D (305), PROCHECK (306), QMEAN (307), and ProSA (210) were used to evaluate the quality of the resulting complexes further.

2.2.3 DPD models titratable residues protonation

To represent the system's correct protonation states, all DPD titratable residues were protonated in the H++ web server at pH 7.5 (308), 0.5 M system salinity, and internal and external default dielectric constants of 80 and 10, respectively. Pronated protein structure files were created using system topology (*top*) and coordinates (*crd)* files. A visual inspection of all titratable residues was performed, and incorrect protonation was corrected using Schrödinger Maestro version 11.8 (309).

2.2.4 DPD parametrization

Two distinct coordination subset arrangements, 1026A (4 × Fe^{2+}, 4 × S^{2-}, 3 × Cys, and 1× Gln) and 1027B (4 × Fe^{2+}, 4 × S^{2-}, and 4 × Cys) were identified prior to the parameter generation procedure. Following that, two approaches were used to determine the force field parameters describing the coordinating interactions in these unique centers. Initially, the original Seminario method (Model 1) was executed in AmberTools16 (250) incorporated with the Python-based metal center parameter builder (MCPB) (249) using the bonded model approach. The protonated protein's Gaussian 09 (310, 311) input files (*com*) integrating the subset structures (1026A and 1027B) were prepared. Their geometries were then optimized using the hybrid DFT method at the B3LYP correlation function level of theory. This method made use of double split-valence and a polarization [6-32G(d)] basis set (311, 312) (Table S1). The metal geometry force field parameters were calculated using sub-matrices of the Cartesian Hessian matrix (260). Harmonic potential fitting yielded bond and angles and force constants. AMBER force field parameter determined the potential energy of the relative position of each atom in the system (Equation 1). We then visualized the

minimized/optimized structures in GaussView 5.0.9 (313) to ensure that the bonds in the centers were still intact. The electrostatic potential was then used to calculate the atomic charges of the optimized subset structures (ESP). However, ESP gives the buried atoms unreasonably charged values, impairing their conformational transferability. To resolve these concerns, the restrained electrostatic potential (RESP) fitting technique, which uses the Coulomb potential to calculate electrostatic interaction, was used. This method has been revered and extensively used for ascribing partial charges to various molecules utilizing B3LYP/6-31G(d) gas phase (314). Penalty functions are used to constrain the buried atoms, resulting in multiple possible charged values. As a result, the quality of fit to the QM ESP is unaffected (315). The metal centers were given a default Merz-Kollman restrained electrostatic potential (RESP) radius of 2.8. An additional approach (hereafter referred to as Model 2) was used, which made use of the collation features of the Seminario: VFFDT program (250). Following optimization of subset Fe^{2+}-S^{2-}, Fe^{2+}-Cys, and Fe^{2+}-Gln coordination, analysis data were obtained; calculations were performed using density functional theory (DFT) with the LSDA/LANL2DZ (Table S2) (316). This considered the internal covalent bonds; however, the calculation failed at the B3LYP level of theory (317). Therefore, GFN1-xTB (318, 319) was used to calculate the external non-covalent bonds. The force field parameters for the entire molecule were retrieved using the Protocol menu item "FF" for the entire "General Small Molecule." Due to the symmetry of the system in this study, the atom types remained Fe or S. The AMBER force field parameters for the Fe^{2+} metal center bond and angles were then automatically generated. Individual detailed statistics were computed, but only the final values were used in subsequent calculations. Using the LEaP (314) program, the obtained parameters were then inferred to the other clusters in the modeled structures as well as the template crystal structure (PDB ID: 1H7X). This was based on the clusters' coordinating geometry being similar. Consequently, cluster 1026A was inferred to be 1029B, and cluster 1027A was inferred to be 1027B, 1028A, 1028B, 1029A, and 1029B because they have the same coordination geometry. For each model, two ($[Fe^{2+}{}_4S^{2-}{}_4(S\text{-}Cys)_3(S\text{-}Gln)]$) and six ($[Fe^{2+}4S^{2-}{}_4(S\text{-}Cys)_4]$) cluster parameters were calculated. In the PDB, there was no other 3D structure with metal centers, such as the human DPD coordinating environment. As such, the pig crystal structure was used to validate the dependability and accuracy of the force field parameters that were generated.

2.2.5 Validity and analysis of force field parameters

Duplicate all-atom MD simulations were performed using the GROMACS 5.14 MD package (320) to assess the reliability of the generated parameters derived from the original and automated Seminario approaches. For simulation studies, the holo (protein with only cofactors) and holo-drug (5-FU) complexes were considered for each model system (Model 1, Model 2, 1H7X crystal structure). Initially, AMBER topologies were generated for each system using Leap modeling with the AMBER ff14SB force field to incorporate all of the generated parameters (189). Using the AnteChamber Python Parser interface (ACPYPE) tool, the resulting system topologies were converted to GROMACS-compatible input files for the structure (*gro*) and topology (*top*), with the correct atom types and charges (321). The infinite systems were then solvated in an octahedron box system with a padding distance of 10 set between the protein surface and the box face using the simple point charge (SPCE216) water model (322). After that, the net charge for all systems was neutralized by adding 0.15 M NaCl counter-ions (323). The neutralized systems were then subjected to an energy minimization phase (without constraints) with the steepest descent integrator 0.01 nm, resulting in a maximum force tolerance of 1000 kJmolTnmm. This was required to eliminate any steric clashes that may have occurred during the incorporation of the parameters and water molecules. Following that, the systems were equilibrated to ensure they reached the proper temperature and pressure using a two-step conical ensemble (each 100 ps). Using a modified Berendsen thermostat, the temperature was first set at 300 K (NVT-number of particles, volume, and temperature). The Parrinello-Rahman barostat algorithm (324) was used to equilibrate pressure at 1 atm (NPT-number of particles, volume, and temperature). The particle mesh Ewald (PME) (325) method was used to describe long-range electrostatic interactions with a gap cut of 8.0, and the LINCS algorithm was used to constrain bonds between all atoms (326). Finally, at the Centre for High Performance Computing (CHPC) in Cape Town, South Africa, production MD simulations of 150 ns were performed for all systems using 72 Linux CPU cores and a time integration step of 2 fs. Every 10 ps, coordinates were written to file. All periodic boundary conditions (PBC) were removed from the obtained MD trajectories, which were then fitted to the reference starting structure.

2.2.6 The Root Mean Square, the Root Mean Square fluctuation, and the Radius of Gyration analysis

Various GROMACS modules were used to determine the global and local conformational behaviors of the replicate ensembles. RStudio (327) was used to calculate gmx rms, gmx rmsf, gmx gyrate, and gmx distance. These packages were used to calculate the root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), and inter-center of mass (ICM) between groups of interest. To ensure that the derived parameters correctly maintained the geometry of the various $Fe^{2+}{}_4S^{2-}{}_4$ clusters, the overall conformational changes per system were observed using visual molecular dynamics (VMD) (328). MD simulations and post-MD analysis such as RMSD, RMSF, Rg, essential dynamics (PCA), VMD visualization, and center of mass (COM) distances calculations allowed for the validation and evaluations of force field parameters. The procedures of these investigations are discussed in detail under their corresponding titles in the section that follows.

2.2.7 Principal Component Analysis

Principal component analysis (PCA) was used in MDM-TASK-web to investigate the time evolution of the protein's conformational changes in MD trajectories (284, 329). PCA is a linear transformation technique that extracts the most important element from a data set by using a covariance matrix constructed from atomic coordinates defining the protein's accessible degree of freedom. After applying RMS best-fit to an average structure, the coordinate covariance matrix for the C and C atoms was calculated (284, 329). A diagonalized matrix was then used to obtain the corresponding eigenvectors and eigenvalues. The protein coordinates were projected using eigenvectors, after which normalized primary and secondary projections were used to generate PC1 versus PC2 plots.

2.2.8 Additional Analytical Methods

PyMOL v1.8 (330), Anaconda 4.3.1 Jupyter Notebooks (331), and various open-source Python libraries, including matplotlib (332), Seaborn, Pandas (333), NumPy (334), and NGLview (335), were used to create the molecular graphics. The average bond lengths and force constants from the derived parameters were compared to those of the X-ray structure to determine how accurate the generated force field parameters were. All statistical

calculations were carried out in RStudio v1.1 using the Welch t-test 456 (327), where p-values less than 0.05 were considered significant.

2.3 Results and discussion

2.3.1 Homology modeling approaches determined human DPD 3D wild type (WT) complete structure

MD simulation experiments rely on the availability of a suitable protein crystal structure to gain a better understanding of the protein's atomic-level molecular function (224). As such, since the human DPD X-ray structure was lacking in protein data bank (PDB) (151), MODELLER v9.15 (303) and DiscoveryStudio 4.5 (304) were used to derive an accurate human 3D model. The pig's X-ray structure (PDB ID: 1H7X, 2.01 Å) was used as a template. The high sequence identity (93%) to the target human DPD enzyme influenced the template selection. It was also in complex with the drug of concern (5-FU) and had an entire query coverage of 100%. MODELLER v9.15 was used to generate 100 apo (without cofactor and 5-FU drug) proteins at a very slow refinement level. The three models with the lowest z-DOPE score values (1.36, 1.36, and 0.88) were then selected for further validation. It is noteworthy that the z-DOPE score is rooted in atomic distance-displacement statistical potential (336). This value considers a model's closeness to the native structure, with a score of ≤ -1.0 indicating a near-native structure. Thus, holo (apo and cofactors) and holo-drug (5-FU) complex structures were generated in Discovery Studio 4.5 by incorporating the template's non-protein coordinates. VERIFY3D (337), protein structure analysis (ProSA) (210), qualitative model energy analysis (QMEAN) (307), and program to check the stereochemical quality of protein structures (PROCHECK) (306) web servers were used for model quality assessment (Table S1).

VERIFY3D estimates the local quality of a model using pairwise interaction derived energy potentials based on each residue structure environment (337). More than 80% of residues in high-quality structures are predicted to have a 3D-1D score of 0.2 or higher (116). In 85.01% of the modeled structures, the 3D-ID scores were 0.2 or higher (Table S1).

ProSA appraises the input model's quality by calculating its potential energy and comparing the resulting score to that of the experimental structures in PDB (210). Each monomer of the holo and holo-drug complexes had a z-score between -13.41 and -13.56, which is comparable to NMR structures of the same size.

QMEAN calculates a value corresponding to the overall quality of the structure and compares it to the determined QMEAN scores of 9766 high-resolution experimental structures (307). The QMEAN scores of the modeled DPD holo and holo-drug complexes were 0.90 and 0.89, respectively, which are comparable to high-resolution experimental structures.

PROCHECK evaluates the stereochemical reliability of the input protein models based on their phi/psi angle arrangement and then generates Ramachandran plots that demonstrate the positions of protein residues on the most favored, allowed, and disallowed regions (209). More than 83.8%, 16.0%, and 0.2% of the residues in each generated model were in the most favored, allowed, and disallowed regions, suggesting a good distribution of torsion angles (Table S2.1). Overall, holo and holo-drug complexes were constructed with consistently high-quality scores.

Notably, the steric clashes in the generated models (holo and holo-drug) were eliminated through 100 minimization steps with the steepest descent algorithm conducted using GROMACS 5.14 MD simulation package, and the models were found to be appropriate for successive calculations (320).

2.3.2 Subset determination and generation of AMBER force field parameters derived via bonded approaches

Metal coordination geometries in proteins are strongly influenced by the protonation states of the implicated residues. Therefore, the protonation states of all titratable residues were ascertained at 7.5 pH using the H++ web server (http://biophysics.cs.vt.edu/H++, accessed on 12 December 2019) (Table S2.2) (308). This ensured that the desired geometry arrangement in the human DPD protein was attained as verified by visual inspection of all titratable residue with Schrödinger Maestro version 11.8 (309). Nevertheless, the protonation states of the residues forming bonds with the metal ions in the $Fe^{2+}{}_4S^{2-}{}_4$ clusters are shown in Table 2.1. Cys was protonated to form CYM, which interacted with the Fe^{2+} center via a sulfur (SG) bond. In contrast, Gln was protonated as GLH and formed a bond with the Fe^{2+} ion via the oxygen (OE) atom.

Table 2.1 Metal coordinating residues' protonation states and pKa values in human DPD protein model. Table adapted from Tendwa et al., (2021).

Residue name	AMBER protonated residue name	Residue number	pKa value
Cysteine (Cys)	[1]CYM	79	8.37
		82	>12.00
		87	>12.00
		91	8.92
		130	>12.00
		136	>12.00
		140	>5.55
		953	>12.00
		956	1.93
		959	<0.00
		963	11.69
		986	>12.00
		989	8.92
		992	<0.00
		996	10.50
Glutamine (Gln)	[2]GLH	156	0.00

[1] CYM denotes protonation state of cysteine (Cys) as specified by AMBER.
[2] GLH denotes protonation state of glutamine (Gln) as specified by AMBER.

The original Seminario (Model 1) and the collation features Seminario approach in Visual Force Field Derivation Tool (VFFDT) (Model 2), were used to calculate AMBER force field parameters of the human DPD protein's $Fe^{2+}{}_4S^{2-}{}_4$ clusters. Two unique residue coordinating environments were observed in each monomer (chain). Clusters 1027, 1028, and1029 (4 × Fe^{2+}, 4 × S^{2-}, and 4 × Cys) coordination differed from cluster 1026 (4 × Fe^{2+}, 4 × S^{2-}, 3 × Cys, and 1 × Gln). Internal coordinates were formed by the four Fe^{2+} (Fe1, Fe2, Fe3, Fe4) ions bonding to the four S^{2-} (S1, S2, S3, S4) ions. Four cysteine, in contrast, bound the four Fe^{2+} (Fe1, Fe2, Fe3, Fe4) via a sulfide link (Cys [SG]) to form the external coordinates of clusters 1027, 1028, and 1029. Additionally, cluster 1026 coordinates externally to the four Fe^{2+} (Fe1, Fe2, Fe3, Fe4) via three Cys [SG] and the oxygen atom of Glutamine (Gln [OE]). Given the spitting image nature of the two chains, the $Fe^{2+}{}_4S^{2-}{}_4$ clusters of the same geometry orientation were assigned a similar number, with each corresponding monomer represented by a different letter: Monomer A (1026-A, 1027-A, 1028-A, and 1029-A) and Monomer B (1029-A) (1026-B, 1027-B, 1028-B, and 1029-B). For QM calculations, the subset structures describing all possible coordination environments for Fe^{2+} centers in DPD protein were considered. When compared to including all the clusters,

this approach significantly reduced computational time and resource utilization. Therefore, QM calculations for $Fe^{2+}_4S^{2-}_4$ subset clusters (1026-A and 1027-A) (Figure 2.2.A) were derived for Model 1 using Becke three-parameter hybrid exchange and Lee Yang Parr (B3LYP) correlation function level of theory (312, 317, 338). However, Model 2 calculations failed at the B3LYP level of theory, so the parameters for single internal coordinates (S3 and Fe3) were achieved using a Los Alamos double-zeta basis (LSDA/LANL2DZ) approach (312). Those for the external coordinates ((Gln and Fe^{2+}) and (Cys and Fe^{2+})) on the other hand, were calculated utilizing a geometry, frequency, noncovalent, extended TB (GFN1-xTB) method (Figure S2.1) (318, 319).

2.3.3. Geometry optimization of subset structures

At step number 238, the subset structures for Model 1 reached the local minima, kicking off the optimization process (Figure 2.2.C, D). A significant energy variation was observed between steps 120 and 230 of the optimization process. The formation of a repulsive bond between Fe^{2+} and Fe^{2+} ions rather than Fe^{2+} and S^{2-} ions in cluster 1026 was the primary cause of the energy variation. Nonetheless, as shown in Figure 2.2.B, the subset structures achieved correct optimization while retaining their geometry.

Individual point value parameters for the subsets in Model 1 were calculated using the original Seminario method (Table S2.3). The VFFDT (Model 2) approach/method, on the other hand, derived average related parameters for internal bond angles and length, while the external parameters were manually averaged (Table S2.4). The equilibrium bond angles and length calculated values from QM (Models 1 and 2) indicated that the crystal structure was slightly distorted (Table 2.2, Table 2.3, and Table 2.4). These discrepancies could be attributed to a lack of phase information on the X-ray structure, as it provides a static snapshot of the dynamic structure, which could lead to spurious values (223). Besides, the disparity could well be based on the lack of solvent effects and intermolecular interactions during the QM gas-phase optimization step (223, 339). The average bond length and angle for Model 2 were, as expected, within the range of those obtained for Model 1. Moreover, in both models, the bond distances between Gln(OE)-Fe^{2+} appeared to be lower (Model 1: 1.92 and Model 2: 1.93) (Table 3.1) than the bond distances between Cys(S)-Fe^{2+}, with force constants of 60.40 and 24.97 kcal·mol^{-}·Å^{-}, respectively. The shorter bond length is possibly due to the smaller atom radius of oxygen in Gln compared to sulfur in Cys (129, 130).

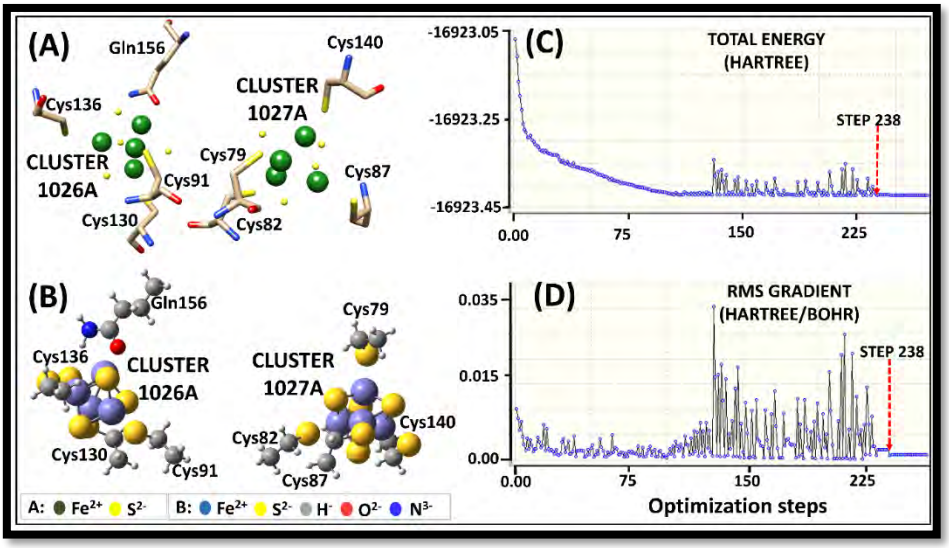


Figure 2.2: Parametrization of human DPD $Fe^{2+}{}_4S^{2-}{}_4$ clusters utilizing original Seminario approach. A) 3D illustration of $Fe^{2+}{}_4S^{2-}{}_4$ clusters coordinating geometry, B) The optimized geometry of Model 1 subset structures (1026A and 1027A) at B3LYP/6-31G* level of theory. C, D) A GaussView visualization of the energy potential depicting the starting point of optimization at the lowest energy level (step 238). Figure adapted from Tendwa et al., (2021).

These values agreed with those obtained in previous studies on Fe^{2+} and Cys (259, 340, 341). Even so, there is a scarcity of literature on Fe^{2+} and Gln force field interactions, which this study adequately addresses. Although there were minor differences, the values of the force constant from both systems (Models 1 and 2) were in the same range and agree with earlier studies (259, 341). Frequently, the force field parameter values of a model derived under different systems are not exact, but fall within an accepted range (250, 260, 342). Additionally, the state of the structural geometry optimization is thought to be a contributing factor to the various observations when generating new parameters (343). Previous findings (344) attributed the disparities to the various methods used to obtain the force constant and the reversed ways in which the connectivity's were defined. Most relevantly, the generated values demonstrated that both models sustained the structural geometry of the subsets after the optimization step.

Table 2.2: A summary comparison of average bond length (Å) values obtained using X-ray, [2]DFT, [4](B3LYP), and [5](LSDA) methods for the molecular cluster models ([$Fe^{2+}_4S^{2-}_4$ (S-Cys)3(S-Gln)]) and ([$Fe^{2+}_4S^{2-}_4$ (S-Cys)4]) of native DPD protein. Table adapted from Tendwa et al., (2021).

$Fe^{2+}_4S^{2-}_4$ Cluster Number	Geometry	Bond Length (Å)						
	Model System	$Fe^{2+}_4S^{2-}_4$(S-Cys)$_3$(O-Gln) and ([$Fe^{2+}_4S^{2-}_4$(S-Cys)$_4$]) Clusters						
	Bond	X-ray	[1] QM ([2] DFT)				AFTER [3] MD	
	Bond Description	1H7X	[4] B3LYP (Model 1)		[5] LSDA (Model 2)		Model 1	Model 2
		Average Bond Length (Å)	Average Equilibrium Bond Length [r_{eq}] (Å)	Force Constant [K_r] (kcal·mol^{-1}·Å^{-2})	Average Equilibrium Bond Length [r_{eq}] (Å)	Force Constant [Kr] (kcal·mol^{-1}·Å^{-2})	Bond Length (Å) Mean and[6] SD	Bond Length (Å) Mean and [6] SD
Cluster 1026A	FE-S	2.54	2.24	58.63	2.22	89.23	2.24 ± 0.21	2.23 ± 0.22
	FE-SG (Cys)	2.35	2.37	48.72	2.33	39.77	2.37 ± 0.01	2.33 ± 0.01
	FE-OE (Gln)	1.89	1.92	60.40	1.93	54.97	1.91 ± 0.01	1.93 ± 0.04
Cluster 1027A	FE-S	2.46	2.24	57.11	2.22	89.23	2.25 ± 0.15	2.23 ± 0.16
	FE-SG(Cys)	2.31	2.38	40.85	2.33	39.77	2.38 ± 0.05	2.33 ± 0.01
	FE-S	2.58	2.24	57.11	2.22	89.23	2.25 ± 0.23	2.23 ± 0.25
Cluster 1028B	FE-SG (Cys)	2.36	2.38	40.85	2.33	39.77	2.38 ± 0.01	2.33 ± 0.02
	FE-S	2.48	2.24	57.11	2.22	89.23	2.23 ± 0.18	2.23 ± 0.18
	FE-SG (Cys)	2.32	2.38	40.85	2.33	39.77	2.38 ± 0.04	2.33 ± 0.00
Cluster 1029B	FE-S	2.54	2.24	58.63	2.22	89.23	2.24 ± 0.21	2.23 ± 0.22
	FE-SG (Cys)	2.35	2.37	48.72	2.33	39.77	2.37 ± 0.01	2.33 ± 0.01
	FE-OE (Gln)	1.89	1.92	60.40	1.93	54.97	1.91 ± 0.01	1.93 ± 0.04

[1] QM: quantum mechanics, [2] DFT: density functional theory, [3] MD: molecular dynamics, [4] B3LYP: Becke, three-parameter, Lee Yang Parr, [5] LSDA: local spin density approximation, [6] SD: standard deviation.

Table 2.3: A summary comparison of average internal angle (°) values obtained from X-ray, [2] DFT [4] (B3LYP), and [5] (LSDA) methods for the molecular cluster model ([$Fe^{2+}_4S^{2-}_4$(S-Cys)$_3$(S-Gln)]) and ([$Fe^{2+}_4S^{2-}_4$(S-Cys)$_4$]) of native DPD protein. Table adapted from Tendwa et al., (2021).

$Fe^{2+}_4S^{2-}_4$ Cluster Number	Geometry	Angle (°)						
	Model System	$Fe^{2+}_4S^{2-}_4$(S-Cys)$_3$(O-Gln) and ([$Fe^{2+}_4S^{2-}_4$(S-Cys)$_4$]) Clusters						
	Angle	X-ray	[1] QM ([2] DFT)				AFTER [3] MD	
	Angle Description	1H7X	[4] B3LYP (Model 1)			[5] LSDA (Model 2)	Model 1	Model 2
		Average Angle (°)	Average Equilibrium Angle [Θ_{eq}] (°)	Force Constant [K_Θ] (kcal·mol^{-1}·rad^{-2})	Average Equilibrium Angle[θ_{eq}](°)	Force Constant [K_Θ] (kcal·mol^{-1}·rad^{-2})	Angle (°) Mean and [6] SD	Angle (°) Mean and [6] SD
Cluster 1026A	FE-S-FE	67.98	67.32	52.64	66.28	26.86	62.91 ± 3.59	68.10 ± 0.08
	S-FE-S	106.03	108.50	39.12	109.21	39.52	109.25 ± 2.28	106.99 ± 0.68
Cluster 1027A	FE-S-FE	68.39	67.61	49.30	66.28	26.86	64.55 ± 2.72	68.24 ± 0.11
	S-FE-S	107.21	108.14	40.39	109.21	39.52	110.0 ± 1.98	108.07 ± 0.61
Cluster 1028B	FE-S-FE	68.22	67.61	49.30	66.28	26.86	66.13 ± 1.48	68.30 ± 0.06
	S-FE-S	106.51	108.14	40.39	109.21	39.52	107.02 ± 0.36	106.97 ± 0.33
Cluster 1029B	FE-S-FE	67.97	67.61	49.30	66.28	26.86	65.15 ± 1.99	67.48 ± 0.35
	S-FE-S	107.62	108.14	40.39	109.21	39.52	106.74 ± 0.62	107.30 ± 0.23

[1] QM: quantum mechanics, [2] DFT: density functional theory, [3] MD: molecular dynamics, [4] B3LYP: Becke, three-parameter, Lee Yang Parr, [5] LSDA: local spin density approximation, [6] SD: standard deviation.

Table 2.4: A summary comparison of average external angle (°) values obtained from X-ray, [2] DFT [4] (B3LYP), and [5] (GFN1-xTB) methods for the molecular cluster model ([$Fe^{2+}_4S^{2-}_4$(S-Cys)$_3$(S-Gln)]) and ([$Fe^{2+}_4S^{2-}_4$(S-Cys)$_4$]) of native DPD protein. Table adapted from Tendwa et al., (2021).

$Fe^{2+}_4S^{2-}_4$ Cluster Number	Geometry	Angle (°)						
	Model System	$Fe^{2+}_4S^{2-}_4$(S-Cys)$_3$(O-Gln) and ([$Fe^{2+}_4S^{2-}_4$(S-Cys)$_4$]) Clusters						
	Bond	X-ray	[1] QM ([2] DFT)				AFTER [3] MD	
	Bond Description	1H7X	[4] B3LYP (Model 1)		[5] GFN1-xTB (Model 2)		Model 1	Model 2
		Average Angle (°)	Average Equilibrium Angle (°)	Force constant (kcal·mol^{-1}·rad^{-2})	Average Equilibrium Angle (°)	Force Constant (kcal·mol^{-1}·rad^{-2})	Angle (°) Mean and [6] SD	Angle (°) Mean and [6] SD
Cluster 1026A	C-Gln(OE)-FE	117.29	130.30	75.86	115.32	41.23	115.29 ± 1.41	114.42 ± 2.02
	C-Gln(OE)-H	104.50	122.90	80.00	118.02	44.55	113.34 ± 6.25	116.93 ± 8.78
	Gln(OE)-FE-S	107.18	109.53	48.56	113.08	40.55	111.10 ± 2.77	112.09 ± 3.47
Cluster 1027A	CT-Cys(SG)-FE	106.87	106.27	102.22	107.39	100.90	107.56 ± 0.49	109.52 ± 1.87
	Cys(SG)-CT-H	108.92	109.50	50.80	104.33	23.56	101.39 ± 5.32	106.06 ± 2.02
	Cys(SG)-FE-S	110.17	110.68	53.74	113.28	36.14	108.84 ± 0.94	112.60 ± 1.72
Cluster 1028B	CT-Cys(SG)-FE	106.72	106.27	102.22	107.39	100.90	111.35 ± 3.27	115.99 ± 6.55
	Cys(SG)-CT-H	107.42	109.50	50.80	104.33	23.56	106.53 ± 0.63	104.94 ± 1.75
	Cys(SG)-FE-S	110.37	110.68	53.74	113.28	36.14	110.89 ± 0.37	112.35 ± 1.40
Cluster 1029B	CT-Cys(SG)-FE	110.70	106.27	102.22	107.39	100.90	105.99 ± 3.33	116.21 ± 3.90
	Cys(SG)-CT-H	110.45	109.50	50.80	104.33	36.14	105.07 ± 3.80	103.38 ± 5.00
	Cys(SG)-FE-S	110.02	110.68	53.74	113.28	36.14	110.58 ± 0.40	111.20 ± 0.83

1 QM: quantum mechanics, [2] DFT: density functional theory, [3] MD: molecular dynamics, [4] B3LYP: Becke, three-parameter, Lee Yang Parr, [5] LSDA: local spin density approximation, [6] SD: standard deviation.

2.3.4 RESP charges

The restrained electrostatic potential (RESP) has been used previously to derive accurate atomic charges for organic molecule (223, 345). This approach involves fitting the electrostatic potential (ESP) generated by a molecule to a set of restrained parameters, such as bond lengths and angles (315, 346). The ESP is computed at various points around the molecule, and these data are used to generate a set of target values for the charges. The fitting process involves minimizing the difference between the computed ESP and the target values by adjusting the charges under specific restraints. By applying these restraints, the RESP approach calculates the atomic charges that best reproduce the experimental or theoretical electrostatic potential (346). These derived charges can then be used in molecular simulations and other computational studies to accurately represent the electrostatic interactions between atoms in the system (223). For the optimized subset structures, partial atomic charge computations were done for each atom interacting with the Fe^{2+} center. Figure S2.2 and Table S2.5 reveal variation in the WT DPD atomic charge distribution in the oxidized subsets. Based on the atom-centered point charge model, the RESP approach calculated these charges by fitting the molecular electrostatic potential derived from the QM calculation. Moreover, due to the large electrostatic environment surrounding the protein's metal sphere, atoms around the DPD Fe^{2+} center (S^{2-}, Gln, and Cys) had varying atomic charges in their oxidized state. Such differences are found to affect charge transfer at the redox center, resulting in metalloprotein stability around the coordinating sphere (342). Consequently, they are critical components in achieving accurate inter- and intramolecular potential electrostatic interaction.

2.3.5 Mirroring the derived QM force field parameters to the identical clusters

The newly derived AMBER Fe^{2+} force field parameters for subsets 1026-A and 1027-A (Tables S2.6 and S2.7) were transferred to the remaining Model 1 DPD clusters relating to the previously mentioned geometries. Additionally, the generated external and internal parameters (Table S4) for Model 2 were assigned to the appropriate clusters. Ultimately, for the DPD ($Fe^{2+}{}_4S^{2-}{}_4$) clusters, each model included a holo (drug free) and a holo-drug (drug bound) (5-FU cancer drug) protein complex, totaling 32 external (30 Cys-Fe; 2 Gln-Fe) and 64 internal (Fe-S) parameter calculations. There were no significant differences in the energy profiles and range of force constants for Models 1 and 2 in terms of DPD Fe^{2+} ion coordination to Cys and Gln residues, and S2 ions. Tables 2, 3, and 4 provided a summary of

equilibrium bond lengths, angles, and related force constants, with more information available in the supporting material (Tables S2.6 and S2.7). Dihedral-related force constants for the corresponding structures were also manually calculated from the structures (Table S8).

2.3.6 Validating of generated force field parameters through MD simulation

Accurate parameters are needed to hold the coordinating geometry of a metal center in metalloproteins (301). As a result, all atom MD simulations (150 ns) for the holo system and holo-drug complexes were performed to assess the accuracy and reliability of the generated parameters (Models 1 and 2). The derived parameters were validated using the root mean square deviation (RMSD) (Figure 2.3.A), radius of gyration (Rg) (Figure 2.3.B), and root mean square fluctuation (RMSF) (Figure 2.3.C). Simulations of both models for holo (drug free) and holo-ligand (drug bound) complexes revealed very little deviation from their original structures, which remained constant throughout the simulation process (Figure 7A). Model 1 systems (holo and holo-drug) violin plots showed a multimodal RMSD distribution, indicating that they sampled multiple local minima, whereas Model 2 proteins only attained a single local minimum (unimodal distribution). The Rg (Figure 2.3.B) demonstrated that the compactness of the different protein models remained constant during dynamics. There were, however, differences between the drug free and drug bound proteins. In both model systems, the drug-free protein had a higher Rg than the drug free protein. This could be due to the presence of the drug. Proteins from both models had comparable RMSF profiles (Figure 2.3.C). The drug bound proteins, on the other hand, appeared to be slightly more flexible than the drug free proteins. As expected, the loop regions, which account for 43% of the total protein structure, including the active site loop (residues 675-679), were the most flexible, while metal site residues fluctuated the least (Figure S2.3). Visual molecular dynamics (VMD) projection of the different trajectories (328) confirmed a high conformational change of the loop areas, while the protein central core containing Fe^{2+} clusters had vibration-like movements.

The violin representation of the RMSDs (Figure 2.3.A) showed greater variation in conformational changes within all systems. These differences were more visible in the proteins of the Model 1 system than in the proteins of the Model 2 system. Given the similarity of protein behavior to drug binding, it is clear that both models demonstrated similar atomic tendencies in drug free and drug bound systems. The disparities caused by

conformational changes were due to minor differences in the approaches used to prepare the models. Fixed bond parameters were assigned between Fe-S, Fe-Fe, and the connecting residues (Fe-Cys or Fe-Gln) of Model 2 based on crystallographic structure averages (Table S2.8), whereas Model 1 parameters were obtained from crystallographic structure single point atom calculations. In all models, the RMSF values of both the drug free and drug bound complexes demonstrated a region of greater flexibility between residues (Figure 2.3.C).

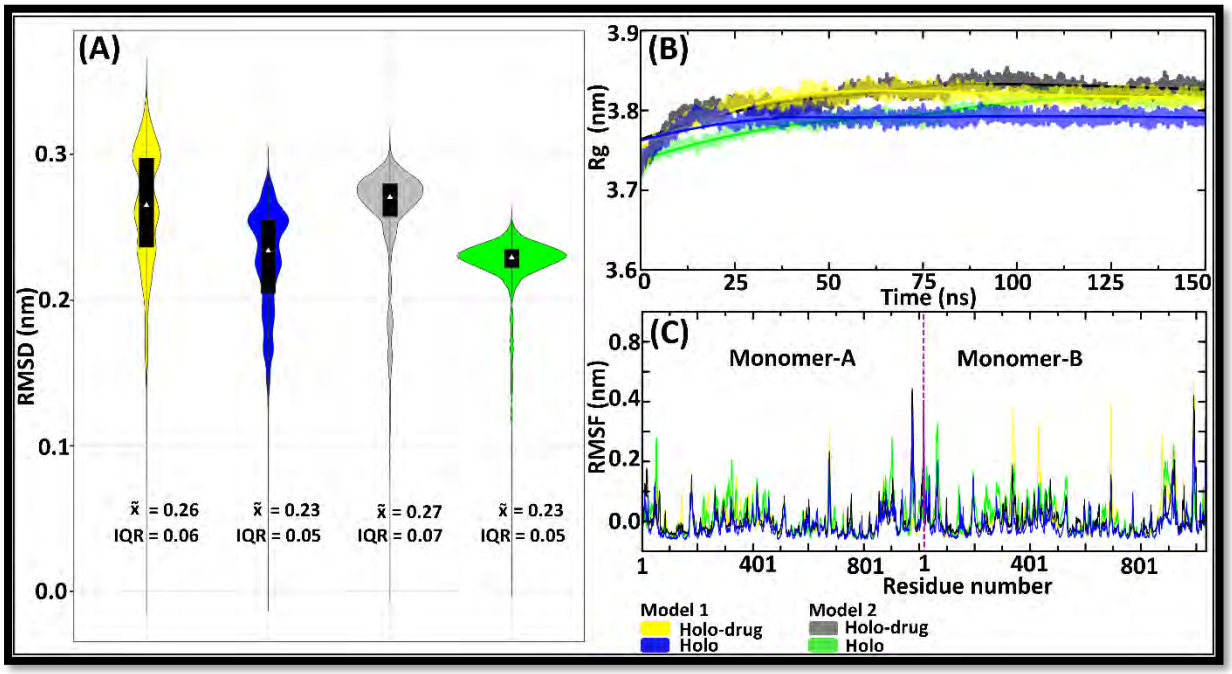
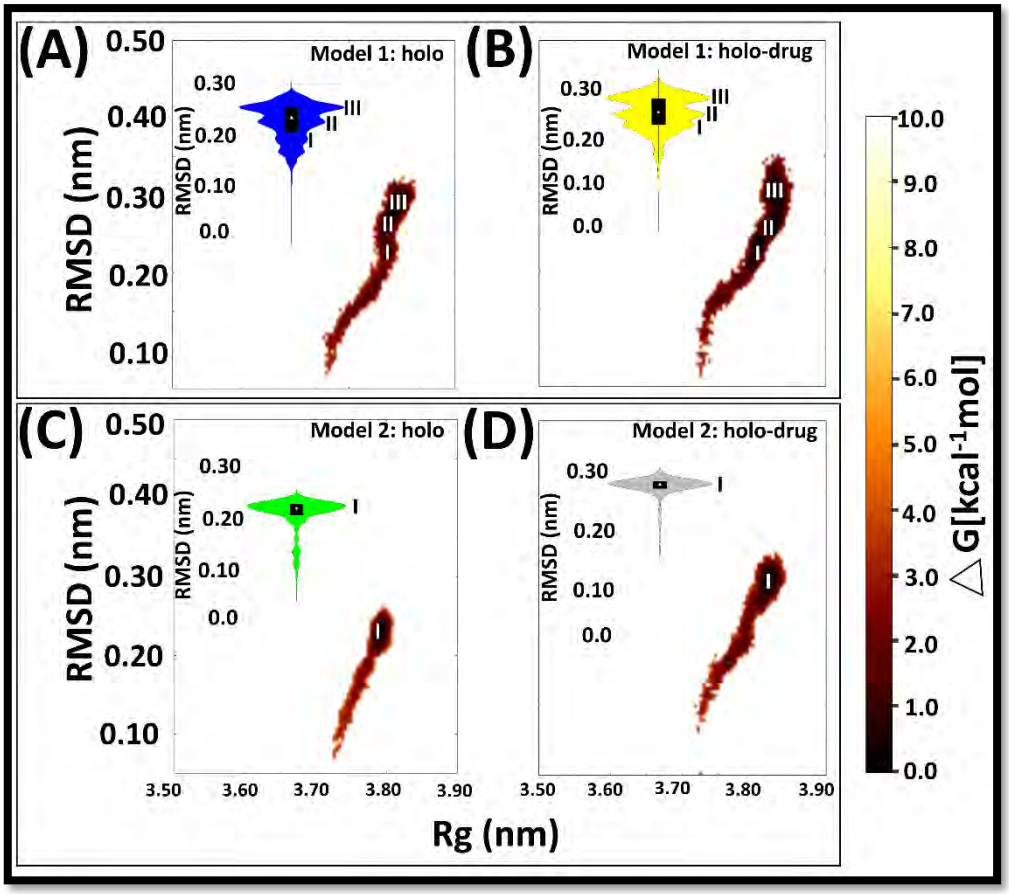


Figure 2.3: A representation of 150 ns MD simulation showing DPD protein stability, compatibility, and flexibility through root mean square of deviation (RMSD), radius of gyration (Rg), and root mean square of fluctuation (RMSF). Model 1 is displayed by yellow (drug) and blue (drug free) systems. Model 2: grey (drug) and green (drug bound) systems **A)** Backbone RMSDs with boxplot overlays on kernel density estimation graphs. Box plots depict the median, upper and lower quartiles. The median and interquartile ranges are used to represent non-parametric kernel density plots (IQR). **B)** Rg line graphs demonstrating the compactness of all systems. **C)** RMSF depicting residue fluctuation, with the purple line separating chains A and B. Figure adapted from Tendwa et al., (2021).

Proteins are dynamic entities that experience conformational changes as part of their feature and function. Understanding how their functionality is maintained requires a deciphering of these changes (347). Thus, we assessed the conformational difference sampled by each system during simulation by plotting the free energy of each system snapshot as a function of RMSD and Rg using the Boltzmann constant (Figure 2.4). Free energy investigations in both models revealed similar tendencies to the violin plots in all systems. In

Model 1, both drug free and drug bound proteins occupied three major conformations. The holo (drug free) protein had three energy minima at 0.18, 0.20, and 0.25 nm, whereas the drug bound protein energy minima were achieved later, at 0.22, 0.25, and 0.35 nm. Model 2, however, equilibrated at single energy minima for both the drug bound (0.28 nm) and holo (drug free) (0.22 nm) complexes. Model 1 proteins tried repeatedly to find a high probability region that guaranteed greater thermodynamic stability for their conformational states than



Model 2. Nevertheless, when the drug was bound, the conformation entropy increased in both models, destabilizing the transition state and slowing protein equilibration. The trajectories were visualized in VMD to determine the cause of the trimodal ensemble, which revealed alternating movements in the loop regions, which included the C-terminal, N-terminal, and active site loop areas. More importantly, the geometry of the $Fe^{2+}{}_4S^{2-}{}_4$ cluster was preserved during simulation (Figure S2.4).

Figure 2.4: A depiction of the free energy landscape of the four system snapshots as RMSD and Rg values derived from the Boltzmann constant in relation to the violin plots. A) Drug bound system (Model 1) displaying three (I, II, and III) major conformational changes sustained by the protein during MD simulation. **B)** A holo (drug free) system (Model 1) with three major conformational changes (I, II, and III). **C)** Drug bound system (Model 2)

with one (I) major conformational change. **D)** Holo bound system (Model 1) protein exhibiting one major conformational change during 150 ns simulation. Model 1: 5-FU drug bound protein (yellow); holo bound protein (blue). Model 2: drug bound protein (grey) and holo bound protein (green). Landscape of free energy (maroon). Figure adapted from Tendwa et al., (2021).

2.3.7 $Fe^{2+}{}_4S^{2-}{}_4$ clusters demonstrated stability during MD simulation

The evaluation of inter- or intramolecular distances between groups of interest can be used to investigate changes in stability during MD simulations (348). Distances between the centers of mass (COM) of: 1) the entire DPD protein and each of the eight $Fe^{2+}{}_4S^{2-}{}_4$ clusters (Figure 2.5.A); 2) each monomer and its four $Fe^{2+}{}_4S^{2-}{}_4$ clusters (Figure 2.5.B); and 3) the active site of each monomer and its $Fe^{2+}{}_4S^{2-}{}_4$ clusters (Figure 2.5.C) were measured for each model (Models 1 and 2: holo and holo-drug). These calculations were used to assess the overall stability of the key components involved in the electron transfer process. In both models, the inter-COM distances between the various groups were nearly identical (Figure 2.5.A-C).

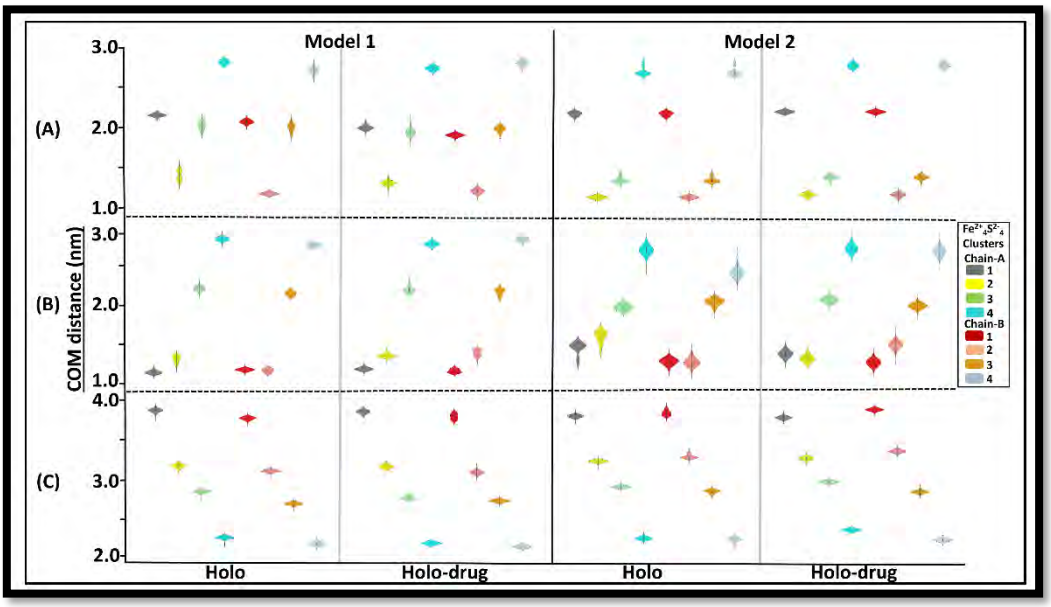


Figure 2.5: The distribution of COM distances across different $Fe^{2+}{}_4S^{2-}{}_4$ clusters and the protein, the monomers (Monomer A and Monomer B), and the active sites are depicted by violin plots. Grey represents cluster 1 (1026A), yellow represents cluster 2 (1027A), green represents cluster 3 (1028A), cyan represents cluster 4 (1029A), red represents cluster 5 (1026B), salmon represents cluster 6 (1027B), orange represents cluster 7 (1028B), and light blue represents cluster 8 (1027B). The interquartile range and median are displayed within the violin plots. **A)** COM distances distribution between clusters and proteins, **B)** Monomer A (clusters 1, 2, 3, and 4), Monomer B (clusters 5, 6, 7, and 8), and **C)** active site. In both models, a unimodal distribution was observed across all clusters. The distance between the Fe^{2+} cluster and the protein's backbone remained constant. During dynamics, the distance

between the Fe^{2+} cluster and the protein backbone remained constant. Cluster compactness is an indicator of system stability. The corresponding clusters are colored accordingly. Figure adapted from Tendwa et al., (2021).

Furthermore, as seen in most violin plots, data were distributed with a lower standard deviation (unimodal distribution), indicating that distances within metal clusters remained in the same range throughout the 150 ns simulation and maintained stability within the metal clusters. As a result, the two methods can be used to obtain accurate parameters for other metalloproteins. The distances between the Fe^{2+} centers and the coordinating residues for the holo-drug complexes were calculated in additament to the group inter-COM distances in both models (Figure 2.6). Using this approach, the integrity of the coordinating geometry could be assessed during simulations. A highly consistent bond length was observed across all $Fe^{2+}{}_4S^{2-}{}_4$ centers, indicating that the derived parameters accurately described the cluster geometries. Moreover, the obtained bond lengths agreed with those previously reported (259, 301). The preservation of bond distances indicated that the desired functionality and stability were not jeopardized, as this is dependent on the protein environment (259).

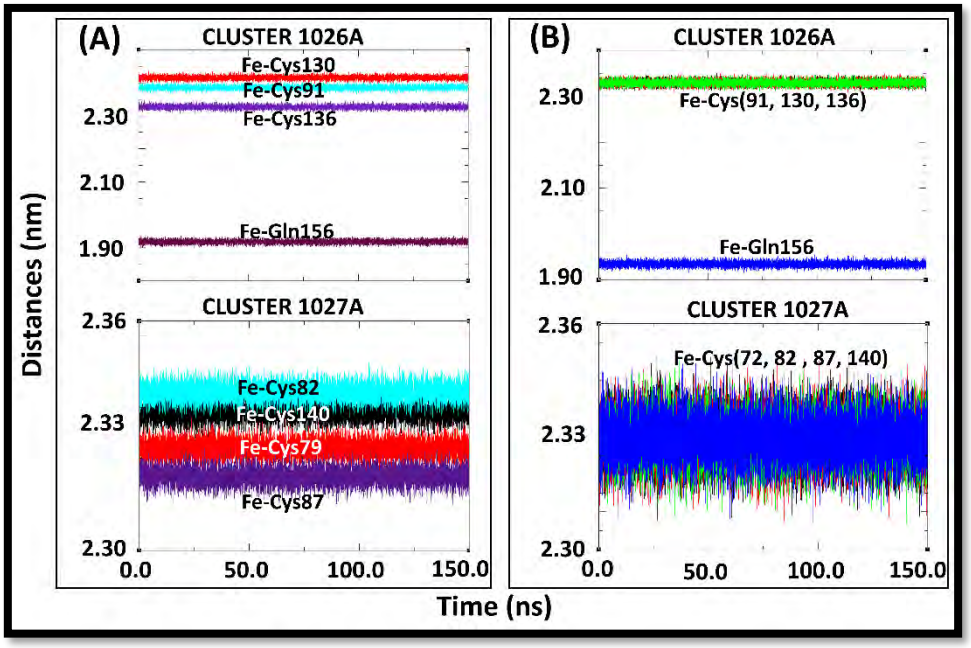


Figure 2.6: Subset structures representing residues coordinating to the Fe^{2+} centers of the drug bound systems during 150 ns MD simulations for Model 1 and Model 2. A) Model 1 holo-ligand clusters 1026A and 1027A, and **B)** clusters 1026A and 1027A of Model 2, the drug bound system. Both models show that the coordinating distances between the Fe^{2+}

and the connecting residues are maintained throughout the simulation. Adapted from Tendwa et al., (2021).

Remarkably, Zheng et al., (250) protocol for evaluating metal-binding structure confirmed that the coordinating tetrahedral geometry of $Fe^{2+}{}_4S^{2-}{}_4$ clusters was preserved throughout the simulation run. While our calculations agreed with previous findings (259, 260, 340, 341), it is important to note that, to the best of the authors' knowledge, none of the studies included the force field parameters for glutamine interaction with a single or multiple $Fe^{2+}{}_4S^{2-}{}_4$ clusters in a single protein.

2.3.8 1H7X crystal structure was used to validate generated parameters

For additional validation, the derived $Fe^{2+}{}_4S^{2-}{}_4$ parameters which coordinated uniquely to Cys and Glu residues were inferred to the template structure (PDB ID: 1H7X). The four $Fe^{2+}{}_4S^{2-}{}_4$ clusters in each monomer of the template, like the modeled human structures, maintained the correct geometry, as shown in Figure 2.6.

2.3.9 Comparative essential dynamics revealed similar motion of protein in phase space

Proteins are dynamic entities for which molecular motions are linked to a variety of biological activities, including redox reactions. Collective coordinates obtained from atomic fluctuation principal component analysis (PCA) are applied to predict a low-dimensional subspace wherein essential protein motion is bound to occur (284). These molecular motions are essential for biological processes. As a result, PCA was calculated to probe the 2D conformational investigation and internal dynamics of the holo and holo-drug complexes of both models (Model 1 and Model 2). The first (PC1) and second (PC2) principal components identified the dominant protein motions of all atoms in the 150 ns MD simulation (Figure 2.7). During this period, both holo structures (Model 1 and Model 2) demonstrated a U-shaped time evolution from an unfolded state (yellow) emerging from simple Brownian motion to a native state (dark blue). Interestingly, the projection of holo-drug complexes from both models (1 and 2) followed a V-shaped time evolution space, beginning in an unfolded state (yellow) and concluding in a native state (dark blue). Model 1 and Model 2 holo

structures constituted 44.95% of total global structural variances. Model 1 and Model 2 holo-drug complexes accounted for 48.95% and 36.5% of global total variance, respectively. In comparison to their respective holo structure, the holo-drug complexes (Model 1 and Model 2) exhibited altered conformational evolution over time. This indicated that, while drug binding alters conformation stability, newly derived force field parameters in both models had no effect on protein function.

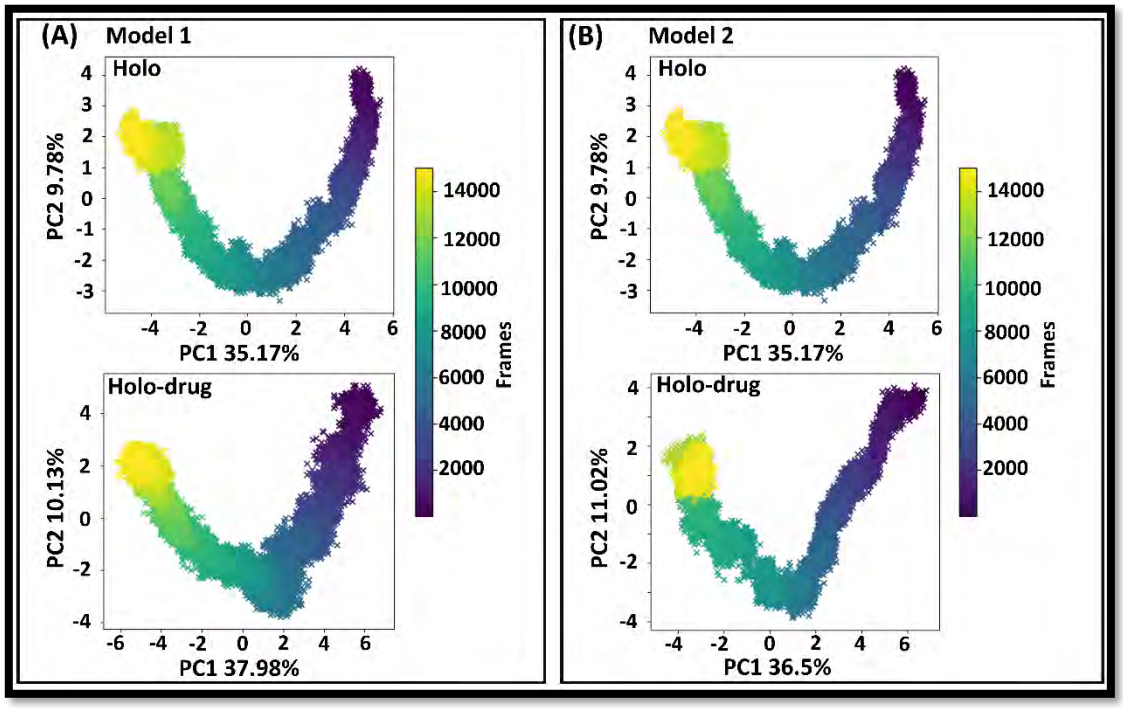


Figure 2.7: A representation of the first and second principal component analyses (PC1 and PC2) of human DPD wild type model obtained from essential dynamics. The time progression of the DPD protein's unfolded state (yellow) arising from simple Brownian motion and ending in the native state (dark blue) is over 150 ns. **A)** The first two PCs of Model 1, constituting 44.95% and 48.95% of the total structural variance of the holo and holo-drug complexes, respectively. **B)** The first two PCs of Model 2 account for 44.9% and 47.52% of the total structural variance of the holo (drug free) and holo-drug (drug-bound) complexes, respectively. Figure adapted from Tendwa et al., (2021).

2.4 Conclusion

The DPD metalloenzyme serves as an important anti-cancer drug target, particularly with respect to its nucleotide metabolizing function in humans (148, 149, 295). However, the presence of mutations causes DPD enzyme deficiency or dysfunction, resulting in increased exposure to active fluoropyrimidine metabolites and severe toxicity (149). Computational approaches, such as MD simulations, have become essential for understanding protein

function and the implications of mutations (295). Such simulations can reveal the conformational transformation of protein systems over time during a reaction (231, 234, 236). They require the use of proper mathematical functions as well as a set of parameters known as force fields, which define the protein energy as a function of its atomic coordinates. Further descriptors are required when adequate parameters, particularly those describing non-protein components in a system, are lacking. We reported new validated AMBER parameters that can be used to accurately describe the complex $Fe^{2+}_4S^{2-}_4$ clusters in the DPD protein and related systems in this work, which serves as a foundation for future research into anti-cancer personalized medicine. This was motivated by the lack of ready-to-use force field parameters that would allow for *in silico* studies on the DPD system. The evolution of combined QM/MM methods has resulted in the most effective, accurate, and theoretical description of the molecular system (311). They allow for a thorough examination of the structural, functional, and coordinating environments in metal-binding sites (231). As a result, we highlighted the capabilities of two similar methods, but with different approaches and aspects of the algorithms for determining authentic force field parameters for Fe^{2+} centers in DPD protein.

To begin, we reported on the generation of force field parameters using the original Seminario method (260). We went one step further and used the VFFDT Seminario method's collation features to obtain the force field parameters of the same Fe^{2+} ions as a supportive measure (250). This was accomplished by bearing in mind the human DPD protein's dimeric functionality, which is based on well-organized inter-chain electron transfer across an eight $Fe^{2+}_4S^{2-}_4$ cluster complex. A double displacement reaction between the two chains results in the activation and deactivation of the world's third most commonly prescribed anti-cancer drug (5-FU) (349). We were able to obtain the favored force constants and bond distances for the Fe^{2+} centers using both Seminario approaches, which was remarkable. Other studies' parameters (259) did not address the clusters' coordinating geometry. Furthermore, none of the studies focused on force field parameters for multiple protein clusters. As a result, obtaining averages of such force fields from the range of force field parameters generated by both approaches would be preferable for future use in other similar systems. These averaged values will have some transferability.

Overall, the derived parameters are easily transferable to consolidated MM packages. Furthermore, we discovered that the original Seminario approach is not inferior to the modified Seminario (collation features VFFDT) approach, regardless of the DFT (B3LYP HF/6-31G* and LSDA/LANL2DZ and GFN1-xTB) logarithm application. Despite the

importance of DFT calculations (such as B3LYP) in understanding the reactivity mechanisms of DPD systems, the method has a major limitation in that it ignores dispersion interactions (350). As a result, the calculations include additional correction methods such as DFT-D3 (351), DFT-D (352), and BJ-damping [115]. The DFT-D3 correction, which is part of Grimme's GFN1-xTB, was used in calculations where the dispersion interactions were most critical in Model 2. Although, for Model 1, future calculations will consider the most appropriate DFT correction method. Due to the possibility of paramagnetism effects attributed to the prevalence of unpaired electrons in the non-trivial DPD system $Fe^{2+}_4S^{2-}_4$ clusters, an attempt to implement unrestricted calculations in Model 2 resulted in a higher energy than under restricted conditions. The MD simulations used to validate the Fe^{2+} force field parameters obtained from this study produced satisfactory results. This will provide more information on the impact of site-specific interactions on the reaction path in the DPD protein and the deleterious mutants (230, 234, 236).

Most relevantly, the authors acknowledge no other compatible parameters for this unique system when it comes to the generation of AMBER force field parameters. Through higher-level MD simulation methods, the generated novel force field parameters laid the groundwork for further simulations and improved the mechanistic understanding of metal cluster function in the human DPD protein. Moreover, the derived parameters have been used in subsequent chapters to investigate the structural and stability effects of existing mutations in the human DPD protein. The findings of these studies have provided atomistic details of mutation effects on the DPD protein. This will provide a foundation for the implementation of *in silico* cancer pharmacogenomics and drug discovery research on the efficacy and toxicity effects of 5-FU drugs.

2.4.1. Final considerations

The use of the term "validation" in the context of simulations conducted with force field models field may differ from its general understanding. Typically, validation refers to the process of comparing the results of a model or simulation to a known, reliable reference or ground truth. It implies a rigorous evaluation against an established benchmark to determine the accuracy and reliability of the model's outputs. However, in the context being discussed, the concept of validation seems to deviate from this traditional understanding. The simulations being described do not have a specific ground truth or reference against which they are being validated. Instead, the simulations are seen as a demonstration that the outputs of different models appear reasonable and show a high degree of similarity to each other.

Therefore, this term is being used more to indicate a general agreement or consistency among the model outputs, rather than a strict comparison to an external reference.

CHAPTER 3

EXPLICATING THE ROLE OF CLINICALLY SIGNIFICANT DIHYDROPYRIMIDINE DEHYDROGENASE MISSENSE MUTATIONS WITHIN AFRICAN ANCESTRY POPULATIONS, IN THE PHARMACOGENOMICS OF 5-FLUOROPYRIMIDINE DYSFUNCTIONAL METABOLISM: A PENTA-METRIC DIRECT RESIDUE NETWORK APPROACH

This chapter presents the identification of DPYD specific missense mutations present in people of African ancestry through available clinical studies and analysis of their effect on the protein via computational approaches. The impact of these mutations on the protein communication pattern has been highlighted.

Contribution: Study design and *in silico* studies were conducted under the supervision of Özlem Tastan Bishop and Thommas Mutemi Musyoka. All the global and local calculations and their analyses were done by **Maureen Bilinga Tendwa.** However, the script for the identification of global top 4% high significant multi-metric (*BC, CC, DC, EC* and *KC*) DRN residues was generated by Olivier Sheikh Amamuddy (for another study); and DRN hub calculations were performed by Victor Barozi in this study. The paper draft was written by Maureen Bilinga Tendwa.

3.1 Introduction

Genetic variations in drug metabolizing enzymes have been revealed to impact drug dosage, pharmacokinetics, pharmacodynamics, and other factors that impact therapeutic outcomes (353). Missense mutations that result in amino acid changes, in particular, interrupt functional sites liable for protein activity such as catalysis and substrate/ligand binding, or disulfide bond sites (354). They can also change the structure, stability, and protein folding, inducing an altered response to medications (355). This can account for clinical manifestations of individuals' susceptibility to disease and drug response, as well as an increased risk of adverse drug reactions (ADRs) (356). Thus, pharmacogenetic studies have revealed the clinical significance of missense mutations in drug metabolizing enzymes such as human DPD. Similarly, this information might be a factor in determining disparities in kinetics and response to a drug, which is critical for drug trials (357).

The functional consequences of missense mutations in clinical settings have been investigated. Some of these variants demonstrated partial deficiency of enzyme activity, while others exhibited diminished or enhanced activity when compared to the native (168, 176). Presence of mutation in the initial and rate-limiting DPD enzyme in the hepatic metabolism of the anti-cancer drug 5-FU results in pharmacokinetics efficacy and toxicity (357, 358). 5-FU application is constrained by a narrow therapeutic index and may be complicated by severe toxicity of WHO grade III-IV symptoms such as granulocytopenia and mucositis (359). Even so, severe adverse occurrences have been reported with capecitabine, an orally bioavailable prodrug of 5-FU (360). The first 5-FU toxicity was reported in 1985 in a patient with familial pyrimidinemia who experienced severe toxicity following 5-FU treatment, indicating that a genetic deficiency in pyrimidine catabolism could be linked to fluoropyrimidine-associated toxicity (361, 362). Subsequent studies by Diasio and colleagues demonstrated that 5-FU metabolism after a 25 mg/m^2 test dose was practically absent in a second patient with familial pyrimidinemia, and that DPD enzyme activity was abnormally low (363). Moreover, the first mutation to be documented as functionally relevant was DPYD*2A, which was discovered in patients with complete DPD deficiency who were homozygous for DPYD*2A (359, 364). This variant is catalytically inactive (176). DPD activity is reduced by 50% in heterozygous patients, resulting in a significant increase in 5-FU exposure (141, 365, 366).

The ability of DYPD genotype-guided dosing to prevent severe toxicity was investigated by Deenen et al., (366). In their study, 2038 patients were prospectively screened for DPYD*2A. 1631 of these patients received fluoropyrimidine-based chemotherapy, with DPYD*2A heterozygous variant allele carriers (n = 18) receiving a 50% reduced fluoropyrimidine starting dose. As a randomized trial was deemed unethical, the primary endpoint - the incidence of severe toxicity - was compared to historical controls (n = 48) receiving full dose therapy. The incidence of severe toxicity was reduced from 73% (35/48) in historical controls to 28% after DPYD*2A genotype-guided dosing (5/18; p 0.001). The rate of severe toxicity after genotype-guided dosing was comparable to that observed in standard-dosed DPYD*2A wild type patients (n = 1613; 28% vs. 23%, respectively; p = 0.64). The occurrence of fatal treatment-related toxicity decreased from 10% to 0% (p = 0.19). The pharmacokinetic analysis revealed that a 50% lower dose in DPYD*2A variant allele carriers resulted in comparable systemic drug exposure as the standard dose in wild type patients. This study demonstrated that upfront DPYD screening was feasible and cost-effective in routine clinical practice (366). Specific starting doses can be recommended depending on the intensity of the variant's effect on DPD activity (Table 3.1) (163). The four tabulated variants [c.1905G >A (rs3918290, also IVS14 +1G >A), c.1679T >G (rs55886062), c.2846A>T (rs67376798), and c.1129-5923C>G. (rs75017182, HapB3)] are thought to be clinically significant due to their known effects on the protein, observed toxic effects, and population frequency (164, 294). In Europeans (164), this last variant is in linkage disequilibrium with the synonymous variant c.1236G> A (rs56038477) and is frequently used in mutational screening analyses. As indicated, most of the clinical studies on genetic variants that contribute to DPD deficiency are well discussed in the Caucasian population. However, there is limited data on the effect of clinically significant missense mutations in non-Caucasian populations (186).

Table 3.1: An illustration of clinically relevant DPYD variants and dosing recommendations

Allele characteristics	dbSNP Identifier	DPYD variants	Consequences	Dosing guidance for homozygous patients (%) (162)	Activity score (165)	Clinical vulnerability to severe toxicity
No function	rs3918290	c.1905 + 1G > A (DPYD*2A)	Exon 14 is spliced out, skipping a portion of the pyrimidine binding site ((367)	50	1	5.2% [95% CI: (3.0 - 8.9)] (368)
No function	rs55886062	c.1679T > G (DPYD *13; I560S)	Amino acid substitution that results to missense mutation. Thus, protein structure destabilization. (129, 171, 369-371)	50	1	0.3% [95%CI: (0.0 - 0.6%)] (166)
Decrease function	rs67376798	c.2846A > T (D949V)	Amino acid substitution that results to missense mutation. Thus, inhibits electron transport or cofactor binding (120, 171)	25	2	5.4% [95% CI: (1.7 - 16.1)] (368)
Decrease function	rs75017182	c.1129-5923C > G (haplotype B3, in linkage with c.1236G > A)	This results in slightly abnormal pre-mRNA splicing. (372).	25	2	6.4% [95%CI: (4.2–8.6%)] (373)

This chapter focused on identifying clinically significant missense mutations associated with the African descent population and uncovering the mutation induced dysfunctional drug metabolism mechanisms. The computational approaches recently developed by our research group were applied (59, 283, 284, 287, 374). The purpose of this study was to determine the structural and functional effects of six clinically proven mutations (C29R, M166V, Y186C, S534N, I543V, and D949V) found in the African population. This will serve as a guide in the development of DPD enzyme gene therapy approach as well as novel drug modulators that will restore the enzyme's proper function in specific carrier patients. Hence, curbing the toxicity associated with these mutations. The interrelationship between function, structure, and mutation has previously been demonstrated experimentally in other proteins. For example, in the study of tumour suppressor p53 missense mutations, it was discovered that mutations in the DNA-binding structural domain resulted in functional loss more frequently. Previous *in silico* research has demonstrated the interdependence of function, structure, and mutation (375). The study of KatG missense mutations, for example, revealed significant loop flexibility around the heme-binding pocket, which influenced asymmetric protomer behaviour in comparison to the native protein (374).

Herein, we generated 3D structures of mutations using our earlier modeled human native DPD protein with recently derived Fe^{2+} center AMBER force field parameters. Thereafter, their conformational dynamics over time were investigated via molecular dynamics (MD) trajectory analysis. Post-MD analysis revealed a wide range of mutational effects on the DPD protein structure, characterized by changes in protein flexibility and compactness of the mutant systems relative to the wild type (WT). One of our novel approaches, comparative essential dynamics (ED) analysis, revealed destabilization and increased conformational changes in the 5-FU (cancer drug) binding environment (active site) of the mutated proteins (284). For the first time in DPD, key residues in the communication network of WT and mutation proteins were discovered using combined dynamic residue network (DRN) multi-metric analysis (*betweenness centrality (BC), closeness centrality (CC), degree of centrality (DC), eigencentrality (EC) and katz centrality* (*KC*)), as previously done in other studies within our research group (283, 287, 374). Overall, this study employs several computational approaches to shed light on the structural and residue communication changes caused by clinically significant missense mutations found in people of African descent. As a result, dysfunctional degradation of 5-FU occurs, resulting in drug toxicity.

3.2 Materials and Methods

The flow chart below (Figure 3.1) illustrates the schematic flow of investigative approaches used in this chapter.

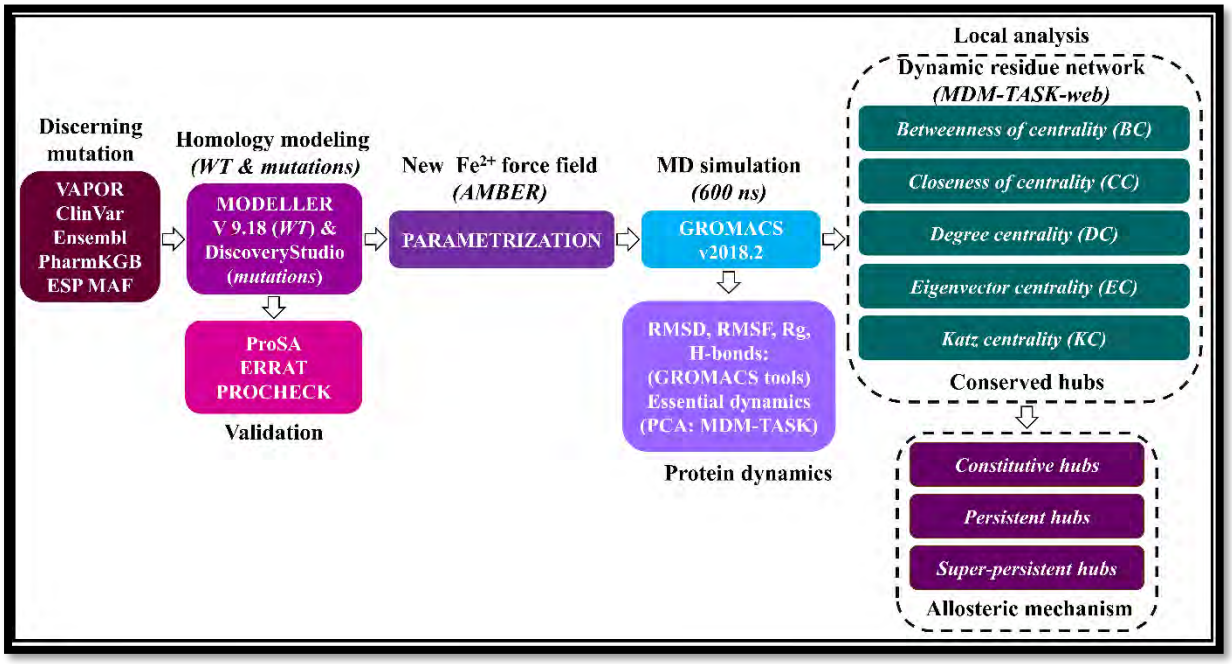


Figure 3.1: The comprehensive computational approaches and tools used in this chapter

3.2.1 Software

MODELLER, University of California, San Francisco, CA 94143, USA; AMBER and AmberTools17, University of California, San Francisco, USA; RStudio v1.1.456, R Core Team, Boston, MA, USA; GROMACS v2018.2, University of Groningen, Uppsala Sweden; R v3.6.1, R Core Team, Vienna, Austria; Discovery Studio 2019, Dassault Systems BIOVIA, San Diego, CA, USA; MD-TASK v1.0.1, Research Unit in Bioinformatics (RUBi), Rhodes University, Makhanda, South Africa; Maestro v12.5, Schrödinger, New York, USA; and PyMOL Molecular Graphics System, v1.8.2.3 Schrödinger, New York, NY, USA.

3.2.2 Dataset collection

Non-synonymous genetic variant information was obtained from the online Mendelian Inheritance in Man (OMIM) (376) and Human Mutation Analysis - Variant Analysis PORtal web servers (219), whereas ClinVar (377), Pharmacogenomics Knowledge Base (PharmKGB) (378), and DrugBank (379) databases were used to collect information on

these mutations and their relationship to human health. Data from the Exome Sequencing Project featuring variants with minor allele frequency (ESP MAF) (380) and ensemble (381) databases were used to obtain additional information about the identification of missense mutations associated with the African population. The minor allele frequency (MAF) of 0.001 of allele frequency related to African population was used to classify the variants. Following that, the clinically significant missense mutations (C29R, M166V, Y186C, S534N, I543V, and D949V) linked to 5-FU adverse drug reactions in people of African descent were chosen for further investigation. The human DPD protein sequence (UniProt accession: Q12882) and its associated mutations were obtained from Universal Protein Resources (UniProt) (382). However, because the 3D crystal structure of the same protein was unavailable, the pig (PDB: 1H7X) (129) structure with a resolution of 2.01 and a sequence identity of 93% to the human DPD protein was obtained from the RCSB Protein Data Bank (PDB) (383, 384).

3.2.3 Structure preparation: 3D structure generation and protonation in DPD wild type and mutations

The PDB did not contain the crystal structures of the human DPD wild type (WT) or the six identified mutations (C29R, M166V, Y186C, S534N, I543V, and D949V). As a result, as stated in our previous work, the human 3D WT model was created using MODELLER version 9.15 (385). The protein models for the mutation were created using the WT as a template. This was accomplished by using Discovery Studio 2019 (304), mutating the MET166, TYR186, ASP342, ARG592, GLY672, and ASP949 to the desired mutation, followed by structure energy minimization. The subsequent step was used to account for any steric clashes that could occur because of replacing a small residue with a larger one. There were seven drug free and seven drug bound models in total. The model quality of each mutation was then evaluated and validated in comparison to the WT protein using ProSA (210), ERRAT (386), and PROCHECK (209). Then, using the H++ web server (308, 387, 388), all titratable residue groups in the DPD WT and mutation protein models were protonated to the correct state (389, 390) at a pH of 7.5 (130), a system salinity of 0.15 M, and internal and external default dielectric constants of 80 and 10, respectively. To generate protonated protein coordinate (*crd*) and topology (*top*) files, the AmberTools19 software, which is integrated into H++, was used. All titratable residues were inspected visually, and incorrect protonation was corrected using Schrödinger Maestro version 12.5 (391).

3.2.4 *In silico* prediction of functional effects and stability of DPD mutations

The amino acid sequence of a protein defines its structure and, as a result, its function. A single amino acid substitution can cause changes in protein structure, stability, and function, potentially leading to disease (392). Predicting the effects of amino acid substitutions on protein stability is therefore essential for protein design and understanding the role of missense mutations in disease. The structural and functional effects of the six identified missense mutations of DPD protein were predicted using Variant Analysis Porta (VAPOR) (219) web servers. VAPOR is a single nucleotide polymorphism (SNP) analysis workflow tool integrated into the HUMA (219) platform that combines three SVM-based tools to predict the effect of mutation on protein stability and function (MUpro (216), I-Mutation (217) and PhD-SNP (218)). To make these predictions, MUpro employs both SVM and neural network approaches. With a confidence score ranging from -1 to 1, the output of this tool describes the changes in energy (G) caused by amino acid substitution. A variation with a score less than 0 reduces the stability of the protein. I-Mutation 2.0 computes the direction of protein stability changes as well as the energy values associated (G) with predicting stability. If the variations have no effect on stability, they are classified as neutral. PhD-SNP predicts the deleterious effect of SNPs in both coding and non-coding regions of a protein structure, categorizing each SNP as either a neutral or deleterious polymorphism.

3.2.5 Molecular dynamics simulation establishing global conformation variation in DPD WT and mutations

GROMACS version 5.1.5 (393) was used to run 600 ns MD simulations of the human DPD WT and the missense mutations to investigate the effect of mutations on structure and dynamics. To determine the point of system convergence for the WT, duplicate MD simulations were run. The topology files were created using the AMBER ff14SB (394) forcefield and a cubic box with a cut-off distance of 10 (the protein is placed 10 Å from the box's edge). As a solvent, water molecules conforming to the TIP3P (395) water model were introduced into the box, and the system was neutralized with 0.15 M NaCl. Prior to energy minimization, the derived AMBER topology files were translated to GROMACS (393) topology using ACPYPE (321), resulting in *gro* and *top* files. Converting topologies from AMBER to GROMACS with ACYPE keeps all previously set Leap parameters, such as cubic box cut-off distance, consistent across programs, allowing the generated topologies to be directly minimized. The energy of the solvated system was minimized for 5000 steps

using the steepest descent algorithm with an initial step of 0.01 nm and no constraints until a maximum force of 1000.0 kJ mol-1 nm-1 was achieved. After the system had converged, a bi-phase equilibration step (every 100 ps) was performed to ensure that the solvated system's temperature and pressure were correct. The Berendsen thermostat was initially set to 300 K (NVT - constant number of particles, volume, and temperature). Following that, a pressure equilibration at 1 atm was performed using the Parrinello-Rahman Barostat (NPT - constant number of particles, pressure, and temperature) (396). The equilibrated systems were then subjected to MD production for 600 ns with a 2 fs (femtoseconds) integration step. This was accomplished at the Center for High-Performance Computing (CHPC) in Cape Town, South Africa, with 148 cores and an average CPU time of 2,880 hours. All bonds were constrained by the LINCS algorithm during the equilibration and production phases (397). Long-range electrostatics were calculated using the particle-mesh Ewald (324) algorithm with 0.16 nm Fourier grid spacing. For the Coulomb and van der Waals interactions, a cut-off distance of 1.4 nm was used. Following the production phase, each system's trajectory was stripped of periodic boundary conditions and centered within the simulation box using *gmx* trjconv. *gmx* rms, *gmx* rmsf, and *gmx* gyrate were used to determine the global and local conformational changes for the entire protein, each monomer, and the active site. Furthermore, during the simulation, g H-bond was used to calculate the number of distinct hydrogen bonds formed by specific residues to other amino acids within the protein (NH bond). The distance between the donor and acceptor was less than 0.35 nm, and the donor-hydrogen-acceptor angle was greater than 150. Further research was carried out with the aid of visual molecular dynamics (VMD) (328), LigPlot+ (398), RStudio (327), and Python libraries such as Numpy (334), Pandas (333), NGLview (335), Seaborn (399), and Matplotlib (332).

3.2.6 Dynamic cross correlation

Dynamic cross correlation (DCC) analysis for the Cα (backbone) atoms in each protein were performed for the three groups of mutations (group 1: clinical, group 2: remote/allosteric non-clinical and group 3: orthosteric/catalytic-domain non-clinical) featured herein. MD-TASK (284) software generated matrices over the entire 600 ns trajectories using the calc correlation.py script. A Python script was used to generate correlation matrix heatmaps for each trajectory.

3.2.7 Analysis of essential dynamics

The comparative essential dynamics tool from the MDM-TASK web featuring compare essential dynamics.py was used to investigate the most prevalent and collective motions of the protein systems (284). Additionally, comparative essential dynamics, by which the native trajectories were paired with each mutation trajectory, were aligned to those of the reference structure (native) through the Cα atoms prior to the decomposition and calculations of the covariance matrix. Consequently, conformation sampling of the mutations was obtained in the same Eigen subspace as the native. Due to increased flexibility, the last 10 and 13 residues from the C and N terminals were excluded respectively. This method was used in several groups of investigations in this study: (group 1) the dynamics of the DPD dimers of Model 1 (built using the original Seminario approach) and Model 2 (built using the automated VFFDT Seminario approach); (group 2) the dynamics of the WT relative to six clinically significant missense mutations; (group 3) the dynamics of the WT relative to seven remote/allosteric non-clinically significant mutations; and (group 4) the dynamics of the WT relative to six orthosteric/catalytic-domain non-clinically significant missense mutations. Similarly, this approach was used to investigate the dynamics of the active site binding environment by specifying the residue of interest in each of the systems in groups 2, 3, and 4. PC1 and PC2 protein motions were plotted as scatter plots for each system, with the percentage variance explained by each PC shown on the axes. On the PC1 and PC2 scatter plots, the time stamps in picoseconds for the lowest energy conformations calculated from 2D kernel density estimates were indicated.

3.2.8 Hydrogen bond interactions and center of mass distances (COM)

Hydrogen bonds (H-bonds) are interactions formed between an electronegative atom (N, O, or F) and another electronegative atom. These interactions contribute significantly to protein structure and are critical to its function and stability (400, 401). Therefore, the frequency of H-bond interactions between the drug and the protein ensembles for the entire simulation period (600 ns) was analyzed using GROMACS *gmx* hbond (393) tool per system and presented as line graphs. On the other hand, CPPTRAJ (402) was used to identify the frequency of prominent hydrogen interactions within the active site. The maximum distance between a donor and an acceptor was set to 3.5 by using the hbond command in AmberTools17's CPPTRAJ (402).

COM distances: to fully understand the 5-FU drug positional stability in mutations ensembles, the 5-FU center of mass (COM) relative to the COM Cα of each monomer was calculated using the *gmx* distance tool (393).

3.2.9 Interface investigation

The human DPD modelled structure was subjected into interface analysis by submitting the structures individually to the PDBePISA (403), PPCheck (404), PyMOL InterfaceResidues (405)and Robetta (406)web servers. As validation of the interface residues, a consensus of at least three programs was required in both cases. A representative multimeric structure was used to depict the location of these residues on the CAs from each class. NGLview (335)was used for visualization.

3.2.10 Multi-metric dynamic residue network analysis and the identification of centrality hubs

For the DRN analysis, Cβ atoms (Cα for glycine) of each residue were treated as nodes in the equilibrated 30 ns part of the simulations, and any connection between the nodes within a cutoff distance of ≤ 6.7 Å was treated as an edge. Using the MDM-TASK-web server scripts, five DRN metrics, that is averaged *BC*, averaged *CC*, averaged *DC*, averaged *EC*, and averaged *KC* were calculated for each snapshot extracted from the same time frame of an equilibrated 30 ns section of each monomer's trajectories (Table S3.1 and S3.2). Utilizing a recently developed DRN analysis algorithm (283, 287, 374), metric-specific results per monomer for all ensembles were unified into a vector before being sorted in descending order while keeping track of the residue numbers (indices). Following that, a threshold for DRN values within the top 4% of the entire set was determined by multiplying the total number of monomer residues (1017) by the number of ensembles (six: clinical mutations; seven: remote non-clinical mutations; and six: catalytic-site mutations) and the percentage cutoff (0.04). The 4% cutoff was chosen to include all hub residues while regulating noise in the data due to the protein's size. Finally, the threshold was applied to generate a binary matrix with dimensions similar to the original set. Thereafter, the highly significant residues (hubs) were demonstrated as a heatmap for each metric.

3.2.11 Analysis of residue contact maps

For each system trajectory, residue contact maps were generated to show the degree of interaction at point of mutation, as well as the identified persistent hubs through multi-metric DRN analysis. This analysis was carried out to detect the differences in residue-residue interactions caused by mutations, in the equilibrated 30 ns of the corresponding time frame section of each monomer's system trajectories in the simulation (Table S3.1 and S3.2). The weighted contact map tool incorporated in the MDM-TASK (284) web was then used to identify the target residue's immediate contact at a cutoff Euclidian distance of 6.7 Å. Subsequently, the contact *heatmap.py* tool from MDM-TASK-web was then used to generate contact heatmaps (32). It should be noted that in all calculations pertaining to DRN, both the drug free (open-conformation/inactivated systems) and drug bound (closed-conformation/activated systems) ensembles were considered.

3.3 Results and discussion

3.3.1 African descent related clinically significant DPD missense mutations retrieved

To obtain information on African descent related DPYD genetic variants known to encode mutated proteins, the SNPdb (407), Ensembl (381), ESP MAF (380), and 1000 Genomes project (408) databases were used. Some of these missense mutations have been linked to increased susceptibility and predisposition to disease (409). ClinVar (377) was used to map their sequence variants and human phenotype. Furthermore, the DrugBank (379) and pharmacogenetics knowledge base (PharmGKB) (410) were utilized to identify 5-FU and how missense mutations affect its pharmacokinetics and pharmacodynamics. The sequences and protein structures, on the other hand, were obtained using UniProt (411) and Protein Data Bank (412). As a result, 128 missense mutations were identified, 19 of which were linked to the African population. Using available literature, six of these mutations were identified as clinically significant (C29R, M166V, Y186C, S534N, I543V, and D949V) (Table S3.3.1 and S3.3.2), while thirteen of the remaining mutations (S201R, K259E, D342N, D432N, S492L, R592Q, A664S, G674D, A721T, V732G, T768K, R886C and L993R) were non-clinically related; meaning not yet clinically validated. Therefore, as stated earlier, clinically significant mutations are discussed in this chapter, while non-clinically significant mutations are discussed in subsequent chapters.

These mutations were located in five distinct structural domains of the DPD monomer, and their geographic distribution in the African population was determined (Figure 3.2). A subset of cofactors (FAD, FMN, and NADPH), prostatic groups ($Fe^{2+}{}_4S^{2-}{}_4$ clusters) and/or ligand (5-FU) which are responsible for proper enzyme function were located within these domains. The active site domain (domain IV) contained two mutations (S534N and I543V), whereas C29R and M166V were identified in the $Fe^{2+}{}_4S^{2-}{}_4$ clusters domain (domain I). In addition, domain II (FAD-binding domain) and domain IV (second $Fe^{2+}{}_4S^{2-}{}_4$ clusters binding domain) contained Y186C and D949V, respectively (129-131). To further evaluate our model structures, we calculated the center of mass distances (COM) of each mutation to the nearest co-factor or 5-FU.

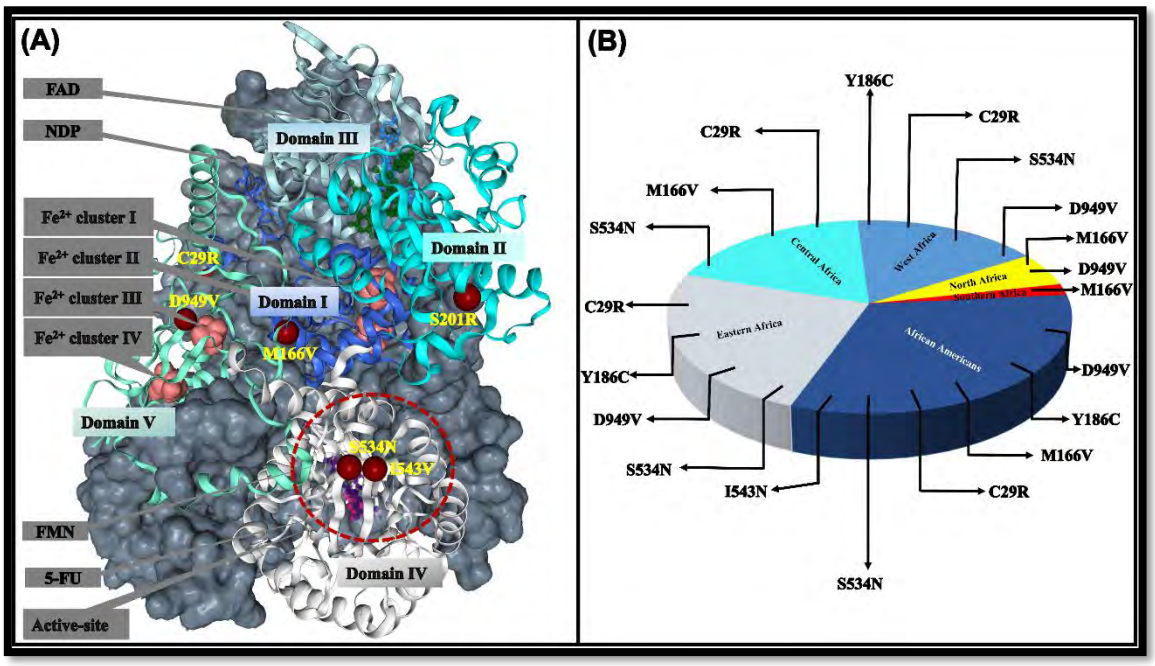


Figure 3.2: A detailed structural representation of the DPD homodimer. A) Monomer A is featured as a surface, whereas Monomer B is displayed as a ribbon, with domains I - V represented in (royal blue, cyan, blue, cream, and sea-green, respectively). The clusters of $Fe^{2+}{}_4S^{2-}{}_4$ are represented by light orange, NDP in blue, FAD in green, FMN in indigo, and 5-FU in pink. **B)** A pie chart of the geographic distribution of mutations in the African population.

Experiments involving site-directed mutagenesis show that mutations at catalytic residues, expectedly, disable enzyme function. Thus, mutations at or near enzyme active sites are frequently detrimental. Furthermore, sometimes mutations far from the catalytic site have been discovered to interfere with function (413-415). S534N was found to be closer to the 5-

FU binding site (COM distances: Monomer A 16.1 Å and Monomer B 13.6 Å) than I543V (COM distances: Monomer A 23.0 Å and Monomer B 23.4 Å). While domain I's C29R (COM distances: Monomer A 46.9 Å and Monomer B 44.6 Å) and M166V (COM distances: Monomer A 23.6 Å and Monomer B 23.9 Å) were located closer to their respective $Fe^{2+}{}_4S^{2-}{}_4$ clusters in Monomer A and B, C29R (COM distances: Monomer A 65.7 Å and Monomer B 64.3 Å) was closer to the active site than M166V (COM distances: Monomer A 73.2 Å and Monomer B 72.8 Å). Y186C, on the other hand, was closer to the FAD binding domain (COM distances of 29.0 and 28.9) of its corresponding monomers but the furthest from the 5-FU binding environment (COM distances: Monomer A 79.3 and Monomer B 78.9). In comparison to the non-active site mutations, the COM distances of domain IV's D949V were closer to the third $Fe^{2+}{}_4S^{2-}{}_4$ clusters (COM distances: Monomer A 12.3 and Monomer B 13.6), but not far from the 5-FU binding environment (COM distances: Monomer A 44.2 and Monomer B 46.3). Mutagenesis to hydrophobic, polar, charged, or aromatic residues demonstrated over- or destabilization ramifications on proteins (416-418), and this study exemplifies this. Provean, Polyphen-2, MU, and PhD-SNP programs included in Variant Analysis PORtal (VAPOR) (219), a protein structure and function prediction program, predicted mutations in this study as damaging and/or deleterious (Table S3.3.1). Similarly, both Ensembl Variant Effect Predictor (EVE) and Variant Annotation Integrator (VARITY) sequence-based variant effect prediction programs, consistently identified D949V, S534N, and M166V as highly pathogenic mutations, while Y186C, I543V, and C29R were determined to have a lower likelihood of pathogenicity (Figure S3.1.1 - S3.1.3). The pLDDT per-residue confidence score from AlphaFold2 predictions depicted higher dynamic region around residues Cys32, Ser326, Pro337, Cys671, Pro672, HIS673, Gly674, Met675, Gly676, Glu677, Arg678, Lys898, Asn901, Val902, Ala903, Phe904, Ser905, Pro906, Leu 907 and Lys 908.

Herein, human 3D WT structure were modeled utilizing pig's protein as a template (PDB ID: 1H7X) (147). The z-DOPE scores of the modeled structures were all below -1.1 (419). Afterwards, the mutant proteins were modeled by introducing mutations in each monomer using Discovery Studio version 2019. (304). Nonetheless, these modeled structures (drug free and drug bound WT and mutants) included a significant number of titratable residues for which classical MD was unable to simulate their protonation state. Thus, they were protonated at pH 7.5 in the DPD active or inactive dimeric state prior to MD simulation (59) through the H++ webserver (308). (Table S2). shows the protonation state of the human DPD WT. Protonation states of substrates and vital catalytically active residues, and

conformational variations of protein side chains and substrates, all play critical roles in chemical reactions (287). The possible impact of missense mutations on protein structure and function was then investigated in two stages: global analysis to distinguish structural changes in the entire enzyme, and local analysis to ascertain mutation-related variations at the residue level.

PHASE I: GLOBAL ANALYSIS

3.3.2 Global structural changes caused by mutations revealed

The global conformational, structural and dynamic aspects of the effect of mutations on 5-FU metabolism by the DPD enzyme were revealed through root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), number of intermolecular hydrogen bonds (H-bonds), dynamic cross-correlation (DCC), and essential dynamics. To begin with, typical RMSD versus timeline plots were deduced to assess the structure's positional divergence from the initial conformation over the simulation (Figure 3.3.A). In light of limited high-performance resources, we performed 600 ns MD simulation for duplicate WT and single mutation runs for both drug free and drug bound ensembles (Figure S3.1). The WT duplicate run results revealed that proteins attained similar conformational changes during the simulation. In both runs, equilibrium was attained at an RMSD of about 3.0. Å (Figure S3.1.A & B). The consistency of the two WT duplicate runs justified the use of single runs for further investigation. Additional RMSD calculations pertaining trajectories obtained from the proteins core (without the loop regions) for the drug free and drug bound structures were moderately lower (0.2 - 3.0 Å) than those of the full protein (0.5 - 5.0 Å) (Figure S3.1.B & C). This demonstrated that the loops were more flexible, as expected. Besides, a similar pattern to that of the full protein was seen for each monomer of the drug free and drug bound proteins. A closer look at each of these monomers using violin plots revealed a range of variation from 0.5 to 5.0 throughout all the systems (Figure 3.3.B). Apart from the drug free WT and M166V of Monomer A, which equilibrated at a single conformation (unimodal distribution), most systems equilibrated at multiple conformations (multimodal distribution). In addition, the equilibration states of the corresponding monomers (drug free: Monomer A & B; drug bound Monomer A & B) of the same system varied. In Monomer A, the drug free WT equilibrated at one conformation equilibrium and in Monomer B, it equilibrated at three conformations (trimodal distribution).

Monomer A and Monomer B were observed to equilibrate at two energy minima during catalysis (drug-binding) (bimodal distribution). Each of the other mutations had distinct variation not only per monomer but also before and after activation. D949V, S543N, I543V, and M166V displayed the most eminent differences from the native protein. We then determined the equilibrated areas of the trajectories in both monomers of the drug free and drug bound ensembles (Figure S3.2 and Table S3.2.1 and Table S3.2). When the RMSD of each frame in the trajectory was compared to the RMSD of all other trajectories, similar trends were seen (Figure 3.3.B). In each system, asymmetric behaviour of Monomer A and Monomer B was observed. Moreover, the FAD, NDP, (Figure 3.3.B). 5-FU, and FMN (Figure S3.2.3) RMSDs were highly variable in both monomers . All in all, the presence of a mutation altered the conformational stability of each inactivated (drug free) protein, which was further altered during protein activation. The conformation stability of both monomers was influenced differently in each case. We previously ascribed the dynamic changes in the RMSD to significant fluctuations in the loop regions, whilst relative stability was maintained within the protein's core (147). Additionally, the results of the successive RMSF calculation and VMD visualization stated in this study backed our previous findings.

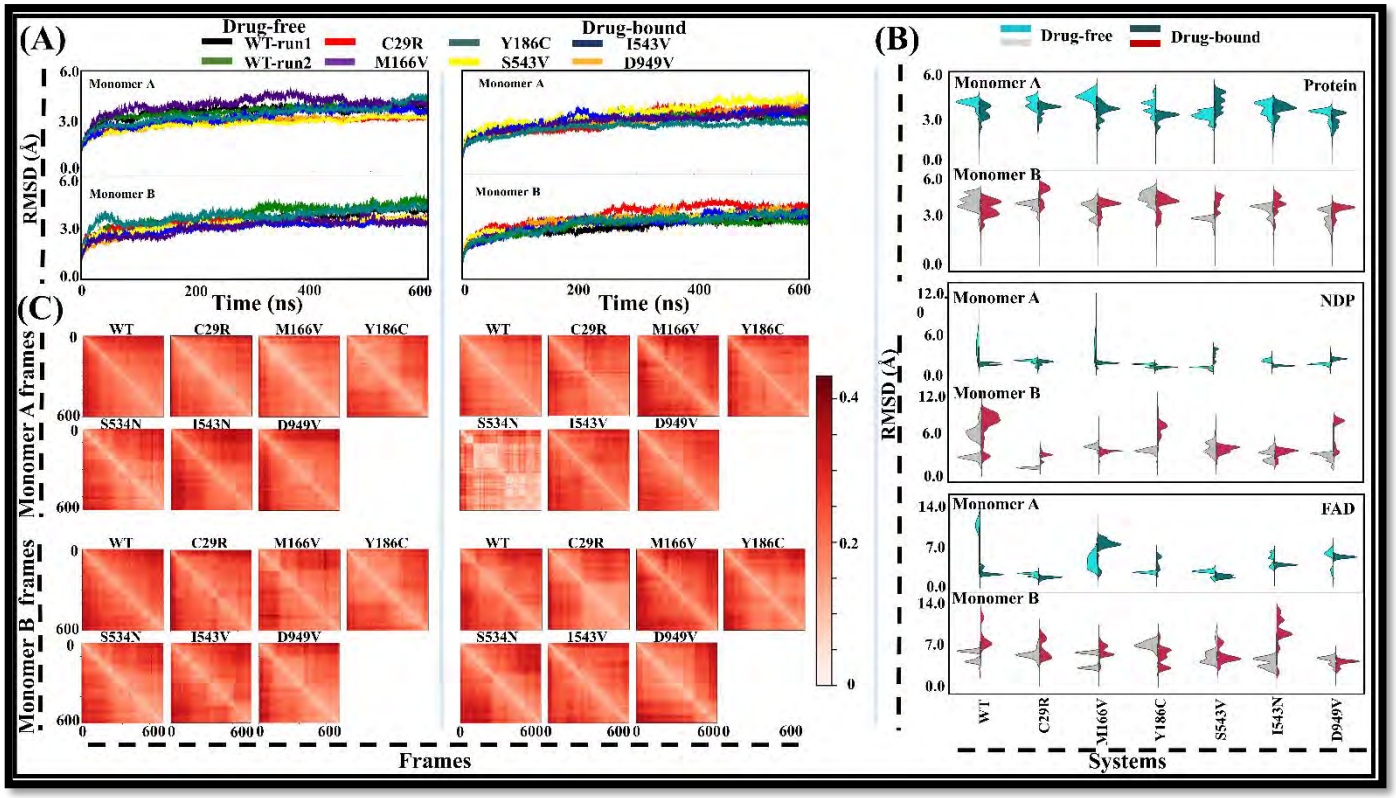


Figure 3.3: An illustration of DPD WT and mutations RMSD and all versus all RMSD. A) WT run 1 (black) and WT run 2 (green), C29R (red), M166V (indigo), Y186C (teal), S534N (yellow), I543V (blue), and D949V (orange). **A)** Line plots showing RMSD progression in reference to the initial structure across the simulation period. The *x*-axis shows the time (ns) and *y*-axis the RMSD (Å). **B)** Split violin plots of the RMSD distribution in both monomers [Monomer A: drug free (cyan), drug bound (blue-green), and Monomer B drug free (grey), drug bound (crimson)] of all systems. The *x*-axis is for the protein systems, and the *y*-axis is for the RMSD. **C)** All versus all RMSD data shown as a heatmap for each monomer of the drug free and drug bound systems. The *x*- and *y*-axes represent the frames at various time stamps (ns). The color gradient from white to dark red indicates the degree of variation. **D)** Split violin plots of the RMSD distribution of co-factors and the ligand in each monomer (NDP and FAD) of all systems [Monomer A: drug free (cyan), drug bound (blue-green), and Monomer B drug free (grey), drug bound (crimson)].

3.3.2.1 Mutation-induced asymmetric monomer changes heightened by loop fluctuation were demonstrated whereas the protein core remained stable

Further RMSF and DCC analyses were carried out to evaluate protein flexibility and the asymmetric behavior of the monomers in each system that were noticeable in the RMSD framewise results. Figure 3.4 describes fluctuation relative to the mean position of each atom (RMSF). Each drug free and drug bound protein exhibited seven major elevated fluctuating regions per monomer. The beta-sheet and adjacent loop areas between residues 40 - 52, 316 - 334, 670 - 683 (active-loop region), 901 - 922, 964 - 971, 2 - 13 (the N-terminal) and 996 - 1025 (C-terminal) fluctuated the most (Figure 3.4). Visible fluctuation was seen in S534N and D949V around residues 964 - 971 of both monomers (Figure 3.4.B). The residues surrounding the $Fe^{2+}{}_4S^{2-}{}_4$ clusters were seen to be more stable over the simulation period. Although the fluctuating regions reacted similarly in both drug free and drug bound systems, the degree of fluctuation in Monomers A and B differed remarkably. Further visualization with VMD revealed the increased fluctuation at the loop region. Next, we used DCC to determine the correlated conformational motions of residues in order to demonstrate the overall effect of mutation on each monomer of the drug free and drug bound proteins (Figure 3.4). In this instance, highly positive areas (colored yellow to red) have strong correlated motions (residue pairs move in the same direction), whilst negative regions (colored blue to green) have strong anti-cross-correlation motions (residue pairs move in the opposite direction).

The correlation between drug-free wild-type (WT) residues from both monomers exhibited a relatively stronger relationship, which diminished upon drug binding. Residues in Monomer B displayed a higher degree of anticorrelation compared to those in Monomer A. Similar patterns were observed in the case of the I543V mutation. Additionally, all mutations displayed increased residual anti-cross-correlation after activation compared to their inactive states, as expected. Notably, C29R, M166V, I543V, S534N, and D949V exhibited the most pronounced variations. Likewise, higher levels of residue anticorrelation were observed in the vicinity of the mutation site compared to the WT configuration.

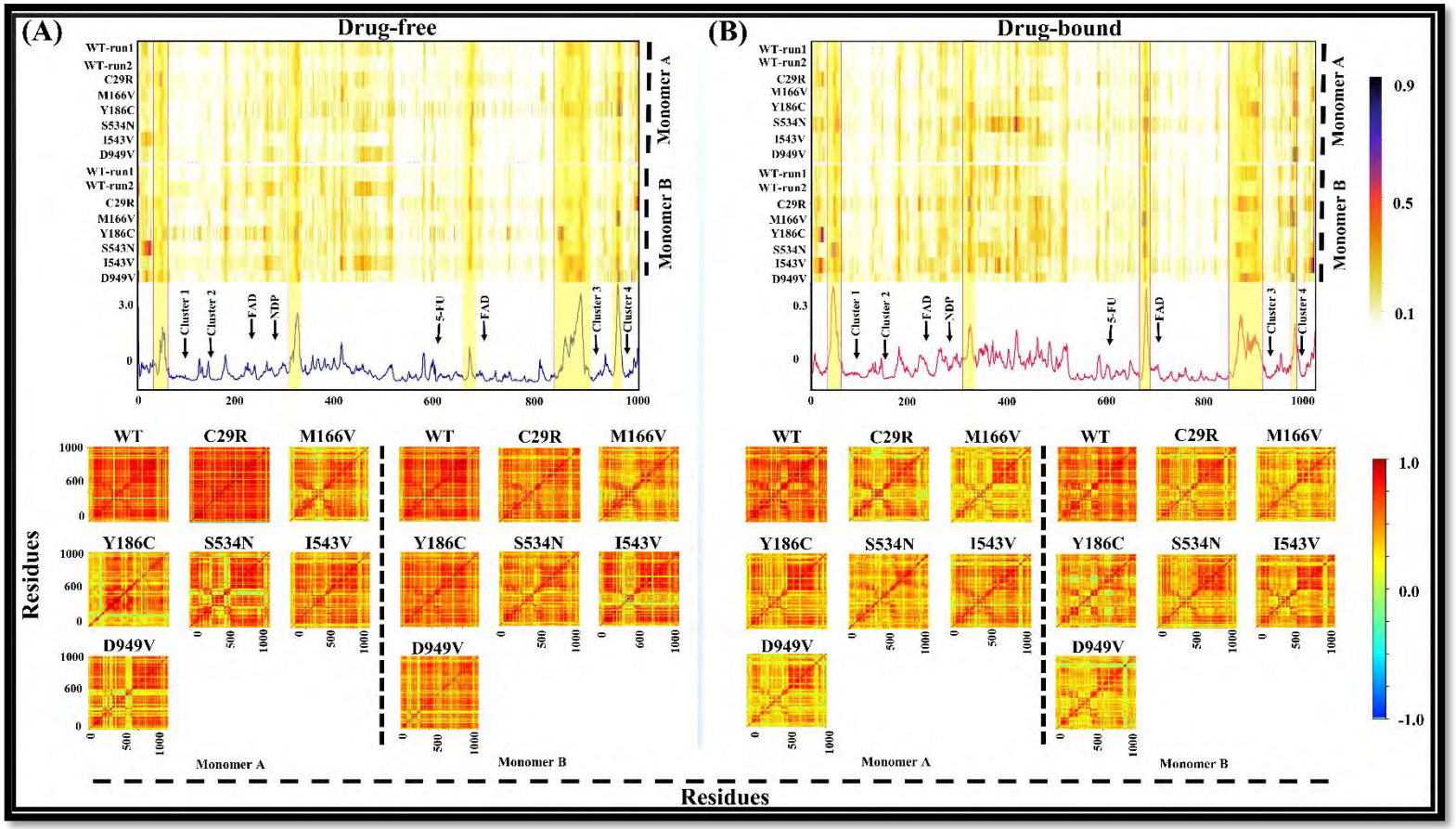


Figure 3.4: RMSF and DCC values of the WT and six mutations presented as heatmaps after 600 ns simulation. The top figure represents RMSF with x-axis indicating residues and y-axis indicating systems. The bottom figure represents DCC with x- and y-axis indicating residue numbers. **A)** The drug free systems. **B)** The drug bound systems. In RMSF graphs, high fluctuation residue areas are highlighted in box frames, and cofactor locations are indicated, whereas strongly correlated residues in DCCM are colored from yellow to red with anti-correlated residues colored from blue to green.

There was notably stronger cross-correlation motion between residues in the inactivated WT (270, 287, 301, 374, 420, 421). Osuna and colleagues (421) have shown that mutations alter the structure and organization of enzyme active sites. Moreso, it has been discovered that mutations far from the active site can disrupt function through allosteric mechanism (283, 287, 422). We proposed that DPD mutations would cause structural changes in the 5-FU binding region of the protein. Previous research demonstrated that identifying local versus global similarity represents two orthogonal directions in protein structure comparison, suggesting that structures which are the most comparable on a global scale may not be the same locally (423). Furthermore, long termini and loops, for example, are flexible and disordered fragments that may have an influence on global analysis since they are often poorly predicted and have a high possibility of jeopardizing otherwise good structural similarity (287, 423). We suppose that since loop areas account for approximately 45% of the DPD protein, they may play a role in the differences observed not only at the monomeric level, but also during DPD activation.

3.3.2.2 Essential dynamics (ED) revealed mutation influences asymmetric collective movement, resulting in distinct conformational changes per monomer

First, in a homodimer, protomer dynamics were observed to switch between identical copies of protomers; mutations affect this shift in protomers differently (283, 374). Furthermore, mutations have impacted various protomer changes during ligand binding (424). Therefore, our subsequent studies sought to investigate mutation-influenced conformational changes at the point of monomers before and after drug binding using comparative ED, one of our newly invented tools from MDM-TASK-web (36). The trajectories were aligned with one another in this method, and a single covariance matrix was calculated to prevent the production of a trajectory-specific covariance matrix that would not be accurately comparable. Regardless of the total variance from the first two PCs remaining lower than in typical PCA calculations, the single covariance matrix reflects the entire variability shared across the trajectories, so we suggest it adequately describes the differences between the systems (35). The comparison of Monomers A and B in mutated proteins to those of their respective WT showed diverse conformational changes along PC1 and PC2, in which the total variances approximately ranged between 50 to 75% and 50 to 60% in the inactive and activated systems respectively (Figure 3.5). Except for Monomer A of drug

bound D949V, which unfolded in the opposite direction, the WT-mutation comparison projection of both the drug free and drug bound systems followed the same direction of longitudinal evolution space, from unfolded state (yellow) to native state (dark-blue). While the WT and mutated proteins unfolded in the same direction, their conformational orientations were opposite in all systems.

In WT-C29R, Monomer B drug free ensembles accounted for the most global total variances (62.76%), followed by drug free Monomer A (59.93%), then drug bound, Monomer A (53.20%) and Monomer B (52.57%). In WT-M166V, on the other hand, Monomer B drug bound ensemble attributed for the least total variances (52.57%), followed by Monomer B (53.30%), then drug free, Monomer B (60.38%) and Monomer A (67.83%). The same trend was observed in WT-S534N; the Monomer B drug bound ensemble reported the lowest total variances (53.97%), followed by Monomer B (54.28%), then drug free, Monomer B (62.0%) and Monomer A (72.47%). Nonetheless, total global variances were higher in drug free Monomer B (60.32%) than in Monomer A (59.04%) and drug bound systems (Monomer A: 51.95% and Monomer B: 54.50%) of WT-I543V pair. The lowest total variances in WT-Y186C were observed in drug free Monomer B (54.29%), followed by drug bound, Monomer B (55.19%) and Monomer A (55.20%), and drug free Monomer A (57.86%). In WT-D949V, Monomer B of the drug free system accounted for the most total variances (61.64%), then Monomer A (57.31%) and the drug bound ensembles (Monomer A: 57.23% and Monomer B: 55.14%). The presence of five domains in each monomer could explain the total low global covariances. According to research, collective domain movement in multi-domain proteins is associated with low global similarity scores (58,59).

Similar studies have indicated that high confident mutation affected conformation evolution of KATG structure with more structural variability (285). In addition, significant structural disorders were observed within the heme binding pocket and the loop surrounding the binding pocket of both the WT and mutations in cryo-EM studies of W107R and T275P mutated proteins (287). Approximately 45% of DPD proteins is made of loop regions, with the most prominent being the active-loop (active-loop: residue 670 – 682) (147). The RMSF observation (Figure 16) has reported high fluctuation within these regions that influences diverse conformational changes, of which similar trends have been observed in ED results.

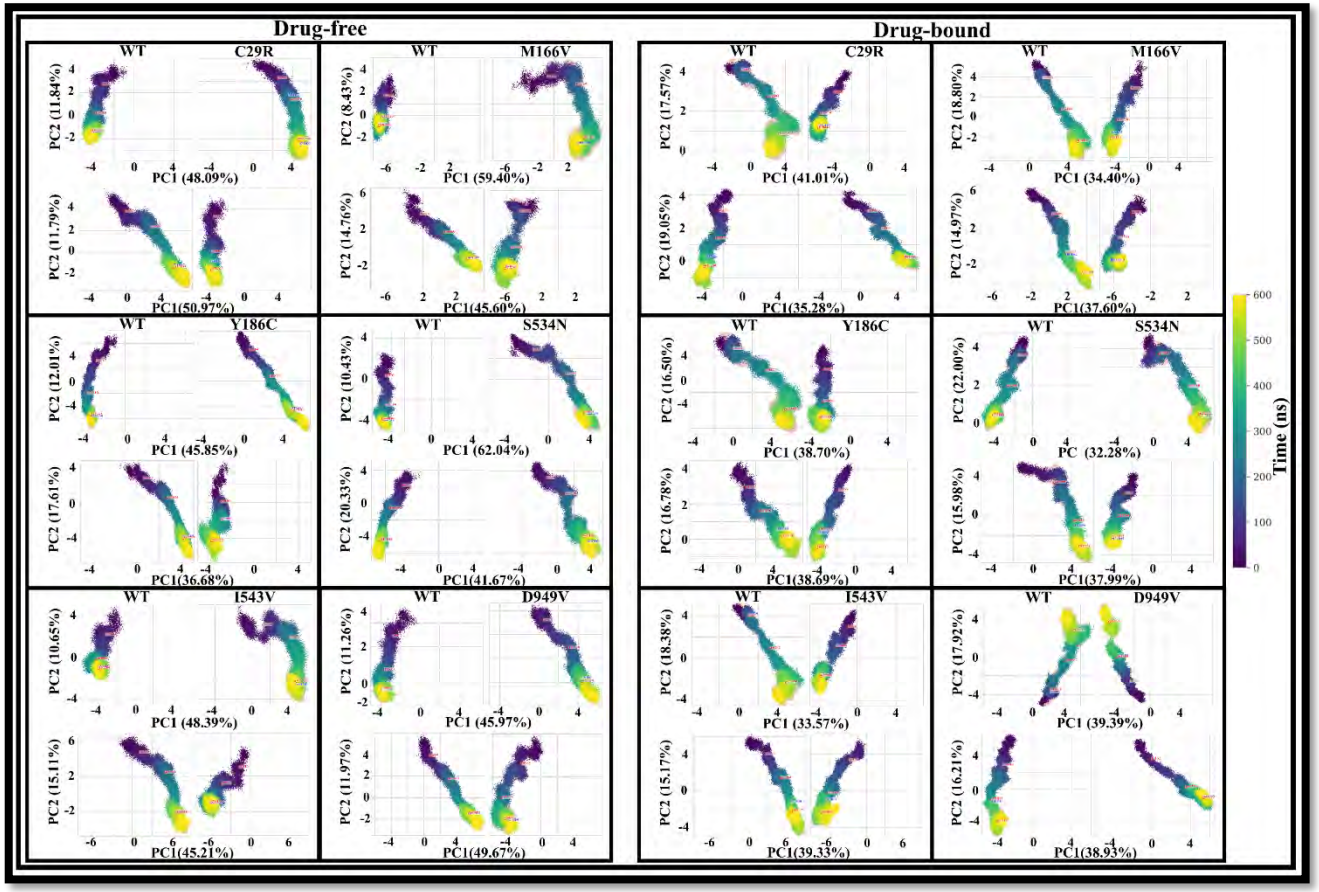


Figure 3.5: Comparative essential dynamic analysis of DPD WT and mutated proteins based on monomers. In each sub-plot, Monomer A is at the top and Monomer B is at the bottom. The WT is on the left while the mutation is to the right. The x- and y-axes represent the percentage variance explained by PC1 and PC2, respectively. The plot color code, which ranges from dark blue to yellow, represents the simulation progression in nanoseconds.

3.3.2.3 Multifaceted conformational changes in 5-FU binding environment of each monomer revealed through comparative essential dynamics

We investigated the 5-FU binding environment as a follow-up action in analyzing the observed heightened asymmetric behavior of the monomers in the presence of mutations, as well as increased active loop fluctuations around the active site. The drug free and drug bound DPD catalytic site binding environment is comprised of residues: 583, 609, 612, 613, 642, 668, 669, active-loop (670 – 682), 709, 736 and 737 (129-131). During the 600 ns simulation period (Figure 3.6), the catalytic binding environment of both drug free and drug bound structures of Monomer A reported 48.04% of the entire global structural variances (PC1: 31.16% and PC2: 16.88%). However, different global variances were observed in Monomer B (total 49.67%; PC1: 30.71%, PC2: 18.96%). The inactivated native protein occupied a conformation space around the 6 o'clock position covering a larger region of phase space along PC2 than PC1 in Monomer A. On the contrary, a more restricted lateral conformation was seen with more phase space occupied along PC1. In Monomer B, the activated native protein revealed a diagonal conformation area ranging from the 6 o'clock position towards the 3 o'clock position, and more phase space was occupied along PC1 than PC2. However, a more perpendicular conformation area was observed with more PC2 phase space than PC1. Direct comparison of either Monomer A or Monomer B drug bound (active state) and drug free (inactive state) structures revealed that a larger region of phase spaces was covered across all mutation upon activation. The most diverse conformational changes were observed in M166V, S534N, I543V, and D949V mutant proteins. The conformation variability exhibited by mutation systems could be attributed to the flexibility of the active loop region (670 – 682) (Figure 3.6.C). System S534N showed the greatest residual flexibility within the active-loop, which could be attributed to its position within the active site. As much as I543V was located within the active site, less active-loop residual flexibility was observed. M166V and D949V also demonstrated increased active-loop residue flexibility. This agreed with a previous study which showed that greatest flexibility around the loop regions of resistance mutations was associated with not only asymmetric behavior of protomers, but influenced the dynamics of the active site binding pocket in KATG protein (374).

For further analysis, we presented Rg distribution representing kernel density estimation (KDE) conformational sampling of the DPD catalytic environment during MD (Figure 3.6.D). The KDE is a non-parametric statistical technique for calculating a variable's

probability density function (PDF). KDEs are related to histograms and have the advantage of not losing information due to binning, as seen in histograms. KDEs smooth the data, making it easier to interpret and determining the distribution shape (270). All mutations showed more conformational variability in the 5-FU binding environment relative to the WT, in both inactivated and activated states, suggesting variant associated instability. Major conformation clusters were seen at 33.80 Å (Monomer A) and 33.90 Å (Monomer B) in drug free structures, whereas drug bound structures' major conformational peaks were seen at 33.95 Å (Monomer A) and 33.70 Å (Monomer B).

On the other hand, D949V conformational peaks were observed at 33.75 Å and 34.88 Å (Monomer A), as well as 34.54 Å (Monomer B) in the inactivated protein. Upon activation, major conformation clusters were seen at 33.7 Å (Monomer A) and 33.9 Å and 33.6 Å (Monomer B). Furthermore, previous studies by Offer et al., in 2014 (174) regarding the potential clinical relevance to DPD activity indicated enzymatic activity was reduced by 41% (*P=0.0031*) in the clinically significant D949V. Also, other clinical studies have reliably linked D949V to severe toxicity following 5-FU administered chemotherapy (172, 173). The M166V mutation was discovered to be significantly associated with grade III and IV toxicity in a cohort of patients with gastroesophageal, breast, or colorectal cancer who were treated with 5-FU (425). This could account for the greatest conformation differences observed in M166V versus WT in this study. Multiple conformational samples were observed before and after protein activation in all cases (activated and inactivated) (Figure 3.6.D). Additionally, Y186C exhibited conformation peaks at 33.7 Å (Monomer A) and 34.6 Å (Monomer B) in the drug free protein. When activated, the protein visited one conformation cluster at 33.8 Å and 33.75 Å in Monomer A and B, respectively. Previously, the DPD activity in peripheral blood mononuclear cells (PBMCs) in a group of Y186C African-American carriers was observed to be lower than that in non-carriers (P=2.9104), suggesting mutations impaired the catalytic activity of the protein (176). Hence, the diverse conformational structural changes exhibited by this mutation. Similarly, S534N demonstrated diverse conformational variability with three peaks at 33.70 Å, 33.80 Å, and 34.55 Å of Monomer A, as well as 33.80 Å and 34.60 Å of Monomer B before activation. Upon catalysis, major conformation clusters were observed at 33.70 Å in Monomer A and 33.80 Å in Monomer B. When compared to the WT, S534N studies revealed a 36% increase in enzymatic activity ($P=3.4\times10^{-7}$) (176). Treatment with fluoropyrimidine (5-FU) reduced DPD activity significantly in Caucasian S534N and D949V mutation carriers (426).

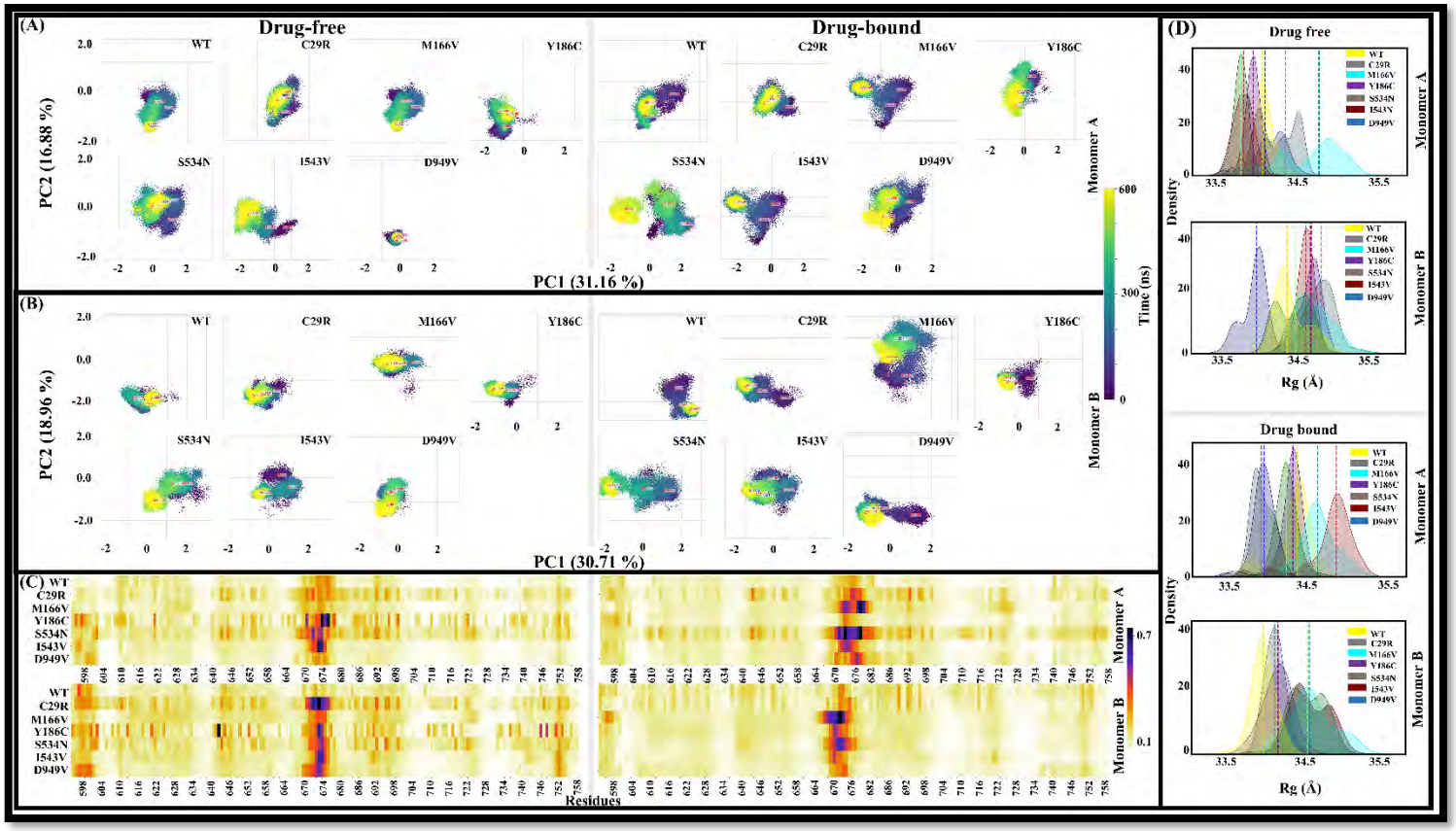


Figure 3.6: Comparative essential dynamic analysis of DPD active site. A) Scatter plots representing essential dynamics of Monomer A's 5-FU binding environment in WT and mutation drug free and drug bound proteins. **B)** Scatter plots representing essential dynamics of Monomer B's 5-FU binding environment in WT and mutation drug free and drug bound proteins. The x- and y-axes depict the percentage variance explained by PC1 and PC2, respectively. **C)** The active site RMSF heatmap represents Monomer A and Monomer B for the drug free and drug bound systems. Most fluctuation depicted around the active site loop region (residues 670 – 683). **D)** Rg distribution of the WT and mutation proteins as portrayed by different colors of kernel density estimation. The median of each kernel plot is represented by dashed lines of the same color. The x-axis depicts the sampled conformation Rgs and the y-axis peaks depict the Rgs of the most sampled conformation.

Consistent with previous studies, 71% of patients with S534N variant developed grade 3 - 4 toxicity (427). Similarly, Indian carriers showed more than grade 2 toxicity (428). The I543V mutation, which is also found in the active site as S534N, has been linked to lower DPD activity in African-Americans (176). Although the active-loop residual fluctuation was not as high as in S534N, it was higher than the WT. Conformation peaks in deactivated S534N were found at 33.70 Å and 33.90 Å in Monomer A, and 34.60 Å in Monomer B. Upon activation, the conformation peaks were at 34.70 Å in Monomer A, and 33.90 Å and 34.60 Å in Monomer B.

In addition, deactivated C29R conformation peaks were discovered at 33.70 Å and 33.90 Å in Monomer A and 34.70 Å in Monomer B. Similarly, when activated, the conformation peaks in Monomer A and B were 33.60 Å and 33.80 Å, respectively. These conformational changes were distinct from those observed in WT, suggesting that they are related to structural instability, which leads to 5-FU metabolic dysfunction toxicity. Increased grade 2 and higher toxicity was observed in Indian carriers treated with 5-FU (428). Other studies have found grade 3-4 toxicity in 46% of C29R carriers (429), and 5-FU chemotherapy has been linked to increased thrombocytopenia in Indian carriers (430).

3.3.2.4 Some monomers showed a diminished hydrogen bond pattern with loss of hydrogen occupancy of critical active site residues

The stronger cross-correlation within residues in the WT relative to the mutations, according to DCC findings, might be credited to the maintained hydrogen bonds during MD simulation. Hydrogen bonds play a crucial function in protein-ligand recognition, protein folding, and enzyme catalysis, in addition to being an important indicator of nonbonding interactions (63). Consequently, the hydrogen bonds within 5-FU binding residues in the mutations and WT during the 600 ns simulation were calculated as illustrated in Figure 3.7.

An intricate network of hydrogen bonds was maintained around the active sites of each monomer in the dug-bound WT across the simulation. Unstable hydrogen bond interaction patterns were observed in most mutation systems. A time-reliant increasing tendency for instability in hydrogen bond interactions was seen in Monomer A of M166V, S534N and D949V from 50 ns, 1 ns and 400 ns, respectively. Direct comparison to the WT revealed increased hydrogen bond instability from 45 ns and 410 ns in Monomer B of M166V and Y186C, respectively. Similarly, a marked increase in hydrogen bond instability was observed in S534N and I543V Monomer B from 480 ns and 10 ns, respectively. C29R, on the other

hand, demonstrated minimal hydrogen bond instability during the simulation in both Monomer A and B.

We then assessed the 5-FU orientation stability by measuring the center of mass (COM) distance of: 1) Monomer A and 5-FU in Monomer B; and 2) Monomer A with Monomer B and 5-FU (Figure 3.7.B). These calculations were driven by the fact that the ping-pong electron exchange from Monomer A to Monomer B active site resulted in the activation of 5-FU in Monomer B and vice versa (129-131). Interestingly, the different monomers of each mutation demonstrated the same tendencies as seen in hydrogen bond interaction in relation to the WT. Hydrogen bond instability corresponded to the change in COM distances; for example, C29R demonstrated minimal changes in hydrogen bond stability, as well as the COM distances between the 5-FU and their respective monomers, as previously observed. Increased hydrogen bond instability could have caused the bond between the three asparagine side chains (N609, N668, and N736), as well as S670 and T737, and 5-FU to dissociate or become destabilized from the protein.

Increased hydrogen bond instability observed in Monomer A's system S534N, and Monomer B's systems M166V and I543V (Figure 3.7.A), was linked to drug exit from the active sites of the corresponding monomers. 5-FU vibrated in a constant motion in the WT when observed with VMD (328); however, it dissociated from mutations that formed highly unstable hydrogen bonds after exhibiting various translation and rotation motions. The 5-FU drug exited Monomer A of S534N at 148 ns and Monomer B of M166V and I543V at 420 ns and 350 ns, respectively. In D949V the drug also demonstrated rotational movement beginning at 325 ns but did not exit the binding. Similarly, the significance of hydrogen bond formation within the key catalytic residues surrounding the active loop region cannot be overstated. Catalytic residue C671 is supported by hydrogen-forming residues H673 and E677. Hydrogen-forming residues H673 and E677 pivot on catalytic residue C671. Other side chain residues that form hydrogen bonds near the active site include M642, I613, and T575. C671 comes into contact with other 5-FU binding residues during catalysis, resulting in a catalytic reaction (165). Thus, using AmberTools20′s cpptraj (402) we compared the hydrogen bond occupancy of these residues in the mutations and WT. In Monomer A of C29R, M166V, Y186C, and I543V, and Monomer B of M166V, Y186C, S534N, I543V and D949V, reduced hydrogen bond occupancy was shown in catalytic residue C670, as well as its hinge forming residues H673 and E677 (Figure 3.7.C).

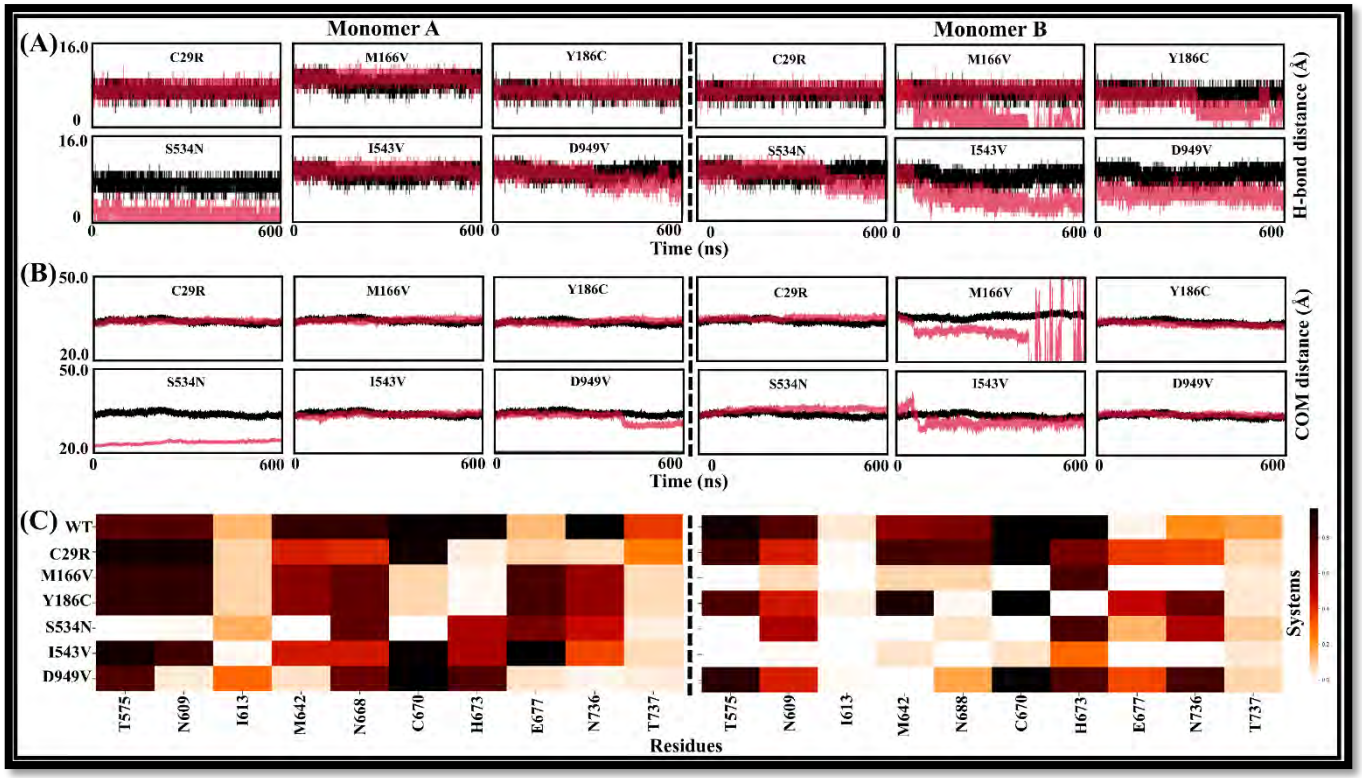


Figure 3.7: A representation of hydrogen bond interaction, center of mass distances (COM) and hydrogen bond occupancy in respect to the active site during the 600 ns simulation. A) Line graphs showing hydrogen bond interaction within the 5-FU binding residues of each monomer, x-axis represents time in ns while y-axis represents hydrogen bond distance in Å **B).** Line graphs representing COM distances between each monomer and its respective 5-FU binding environment, x-axis depicts time (ns) while y-axis shows COM distances (Å). **C)** Heatmaps representing 5-FU hydrogen bond interacting residues (ASN609, ASN668, and ASN736) and side chains of SER670 and THR737, loop pivot point hydrogen interacting residues HIS673 and GLU677. Additional side chain residues MET642, ILE613, and THR575 of Monomer A and Monomer B.

The reduced hydrogen bond network in these residues may have destabilized the 5-FU catalytic mechanism, resulting in dysfunctional drug degradation and toxicity.

It has been well documented that mutations in a protein can change the structure that the WT sequence naturally confers (431). Evidently, each mutation in this study influenced the function of each monomer in a distinctive manner, similar to results obtained by Sheik Amamuddy and colleagues (266). Missense mutations were revealed to affect the dynamics of each SARS-CoV-2 Mpro protomer/monomer differently in their studies. Other research has found that mutations influence either the loss or change of enzyme function, regardless of whether they are distal or proximal to the active site (432). Except for S534N and I543V mutations, which are found in the active site, all other mutations in this study were remotely located. Changes in their mutation topologies were evidenced at locations other than the site of mutation, including the active site. The phenomenon of structural change at distant sites from the site of perturbation is well known as allostery (433). Allosteric mechanisms permit the transmission of local structural perturbations through a network of motions or collisional contacts, which can be recognized as allosteric communication pathways in protein structure from one site to another (422).

In general, the M166V, S534N, and D949V mutations demonstrated consistent and diverse changes across multiple parameters, including RMSD, RMSD framewise analysis, RMSF, DCC, ED, and H-bond (global studies). These findings align with the investigative research conducted by Offer et al., (168, 176) focusing on the functional analysis of DPYD variants associated with the catabolism of 5-FU. The study revealed significant results, specifically, the M166V mutation demonstrating heightened enzymatic activity compared to the wild type (120%, P = 0.025). Notably, the S534N mutation displayed a significantly increased enzymatic activity (hyperactivity) relative to the wild type (36%, P = 3.4×10−7). Conversely, the D949V substitution led to a notable decline in enzyme activity by 41% (P = 0.0031), corroborating observations from clinical studies on 5-FU toxicity. This particular mutation (D949V)and C29R, M166V, S534N and I543V have been observed in both African and Caucasian populations (434). The impact on functional loss of D949V is believed to stem from its potential interference, either directly or indirectly, with cofactor binding or electron transport processes. The substitution of valine for aspartate at position 949 leads to the elimination of a hydrogen bond involving the carboxyl group, which normally interacts with the main-chain nitrogen of Glu951. This hydrogen bond plays a crucial structural role as it stabilizes a helical turn following a β-strand in the C-terminal FeS-cluster domain of DPD. Consequently, it positions the loop that carries the cluster ligands Cys953 and Cys956 in

close proximity to the cluster cFeS1. The loss of this hydrogen bond is likely to have adverse effects on cluster formation (171). Based on our structural analysis, the amino acid substitutions M166V and S534N have the potential to affect the functionality of the enzyme. These substitutions are situated within a conserved three-dimensional environment in the DPD protein. Consequently, the respective mutations can cause a rearrangement of this specific three-dimensional network, resulting in a decrease in enzymatic activity (180). Mutations such as C29R have been demonstrated to cause a slight decrease in DPD activity compared to the wild-type enzyme (179). Certain cancer patients, who were homozygous for the C29R mutation, experienced severe toxicity after undergoing 5FU treatment (435). The crystal structure analysis of pig DPD does not offer clear explanations for the lack of functionality in the recombinant mutant protein. Notably, Cys29 is not evolutionary conserved. n DPD, the histidine residue is situated in a region of the dimer interface that is accessible to solvent. Its side-chain interacts solely through a water-mediated hydrogen bond with the imidazol-ring of His94 from the other subunit in the dimer. In the human enzyme, this histidine is replaced by an asparagine, which maintains all the hydrogen-bond interactions observed for His94 in pig DPD, including a water-mediated hydrogen bond with the cysteine at position 29. As residue 29 is positioned on the protein surface, introducing a considerably larger arginine at this site does not result in unfavorable close contacts with neighboring residues, as its side-chain can be directed towards the solvent without complications (171). Despite ongoing research, the conflicting data concerning the impact of the C29R mutation on DPD function remain unresolved (435). Similarly, the amino acid position 543 is not evolutionary conserved (436). Despite the structural analysis not finding any evidence of impaired enzymatic function resulting from the specific amino acid substitutions (171), it is important to consider the potential cumulative impact of the I543V, C29R, M166V, and S534N missense mutations on destabilizing the dimeric protein complex. The presence of these mutations has led to diverse conformational changes in the DPD structure, as revealed by global analysis, suggesting that their influence on destabilization cannot be disregarded.

PHASE II: LOCAL ANALYSIS

3.3.3 Presence of missense mutations revealed a heightened accentuated asymmetric change in residue interaction pathways per monomer

The next step was to understand the shift in communication within the active site of DPD protein at the residue level by unravelling both native orthosteric and mutation allosteric communication pathways by using local analysis (DRN) and contact map network. Missense mutations in drug metabolizing enzymes have become potential biomarkers for precision medicine due to their broad effects on protein function dysregulation (9). They can alter protein active site mechanisms by introducing allosteric communication (422, 423). Recently, in the RUBi research group, the combination of MD simulations with network analysis (DRN approaches) has offered insight into the mechanisms of interaction between protein residues, including allosteric communication pathways (270, 273, 314, 358). Residues are normally referred to as nodes in a protein communication network, and non-covalent interactions between nodes are referred to as edges (437). The name "centrality" refers to the significance of a node in a communication network. Changes in residue side chains caused by mutations can cause reorganization of protein network patterns and thus change node centrality (374).

As with our previous studies (284, 329), we used five different centrality metrics [*betweenness centrality (BC), closeness centrality (CC), degree centrality (DC), Eigen-centrality (EC), and Katz centrality (KC)*] to identify high centrality residues in WT versus mutation protein systems. Each centrality metric calculation required a more stable (equilibrated) area of the trajectory for accuracy and reproducibility. Maximum stability was not achieved by the end of the 600 ns MD simulation due to the large size of the protein in our case. Furthermore, the drug exited some systems before 600 ns. As a result, where applicable, samples were extracted from the same time frame of an equilibrated 30 ns section of trajectory each monomer on a protein system prior to drug exit (Figure S3.2; Tables S3.2 and S3.3).

The overall depiction of high *BC, CC, DC, EC,* and *KC* (Figures S3.3.1 & S3.3.2) demonstrated a similar pattern of high centrality residue concentration in each system. In both the drug free (inactivated) and drug bound (activated) states of each metric, there were variations in the high centrality residues between each mutation system and the WT. Components involved in the redox reaction cascade (NADPH, FAD, $Fe^{2+}{}_4S^{2-}{}_4$ clusters, FMN, and 5-FU) which also reflected the interface residues (Figure S3.4 and Table S3.4) had increased residual communication in both drug free and drug bound systems of both

monomers. As expected, these were more noticeable in the *DC*, followed by the *CC, BC, KC*, and *EC*, in that order.

We then aimed to improve on the selectivity of the previous results' predictive accuracy by identifying the top global 2.0%, 3.0%, and 4.0% of high centrality significant residues in the five metrics using stringent measures (Figures S3.5.1, S3.5.2 & S3.5.3, respectively). Prior to determining the top global high centrality residue, symmetry was corrected by aligning a single equilibrium conformation of each monomer high centrality sample conferring to its metric. The global 4.0% high centrality significant residue was chosen for further analysis over the 2.0% and 3.0% residues because it provided sufficient information without increased noise. Additionally, we showed that each metric offers exceptional viewpoint to the network, rendering collective information significant. Based on these analyses, some new terminologies were introduced by our research group, for instance, the terms "*hubs"*, "*persistent hubs*" and "*super-persistent hubs*" (266). Any node that was part of the set of highest centrality nodes was defined as a "*hub*". This was based on the top 5% centrality nodes measured across all related samples for any given averaged centrality metric. Similarly, the term "*persistent hub"* was defined as a hub that existed across all systems compared. "*Super-persistent hubs*" were persistent hubs found across all penta-metrics (266). Herein, our approaches were used to identify metric-specific hubs in DPD protein and to investigate the impact of missense mutations on its communication profile in both the inactive (drug free) and active (drug bound) states. As previously indicated, the heatmap (287, 374) displayed 4% hub data for each metric. The subsequent subsections give an account of the heatmaps analyses in detail. Moreover, to gain deeper understanding into changes in hubs caused by mutations, we mapped the uniquely observed hubs of each mutation protein in relation to the WT hubs (*BC, CC, DC, EC, KC*) as well as the rest of the commonly shared hubs with the WT for each metric.

3.3.3.1 The *Betweenness Centrality (BC)* identified a pair of distinct compensatory communication pathways from NADPH to 5-FU along the interior of the oxidoreduction catalytic channel

BC is an indicator of a residue's usage (and thus importance) in protein communication, and it signifies the number of shortest paths across all other node pairs that pass via a particular node of importance (329). This has been evidenced to be one of the most insightful metrics in a variety of studies. It has been employed in identification of critical

allosteric residues (287, 420, 422), and ligand and mutation-induced network switch (287, 374, 438, 439).

Intermonomer and intersystem differences were observed in Figure 3.8.1 of the 4 % averaged *BC* calculations, in the presence of mutations. This was consistent with previous findings regarding the asymmetric monomer behavior of a dimeric protein (266, 285). From the heatmap (Figure 3.8.1) Leu155 and Val373 hub residues were unchanged from the reference protein (WT) in the presence of DPD missense mutations in both drug free and drug bound systems. Hence, they were identified as persistent hubs. Leu155 is found in domain I's α-helix, connecting to Gln156, which forms a one-of-a-kind bond with Fe^{2+} ion. This is an important component of electron transport chain to the active site (129-131). Val373 is found in the α-helix of the NADPH binding domain III and participates in the NADPH cofactor interaction (129, 130). The NADPH-nicotinamide moiety is positioned between the FAD isoalloxazine ring, and one of Asp342's carboxyl oxygens is hydrogen bonded to the backbone amide of Val373, hence important in NADPH stability (129, 130). Residues at positions Ile116 and Leu142 were among the additional *BC* hubs found in at least five systems. These are domain I α-helix interphase residues that are an extension of the $Fe^{2+}{}_4S^{2-}{}_4$ cluster contacts, which are critical for electron transfer during catalysis (130, 131). Therefore, any mutations that occurs closer to the cofactor or the $Fe^{2+}{}_4S^{2-}{}_4$ clusters, may disrupt electron transport to the active site (436). Remarkably, the residues Cys82 and Val493 were exclusively hubs drug bound mutant systems, particularly in Monomer A (Table S4). Cys82 forms a sulfur bond with the first $Fe^{2+}{}_4S^{2-}{}_4$ clusters (147), while Val493 is a α-helix FAD binding domain II residue in contact with FAD binding residues (131). Short-range contact using contact maps analysis discussed later in this study (Chapter 3.3.4), showed that new contacts were gained with residue Cys82 (Cys82: Leu80, Ala83, Ala84, Ala85, Pro86, Cys87 and Asn141) and Val493 (Val493: Pro92, Thr93, Asp119, Asn120, Pro121, Leu122, Cys126, Gly127, Pro197, Ala198, Val490, Glu491, Ser492 and Lys497), thus, observed increased centrality of these hubs. Barozi and colleagues, (285) had similar finding. The observed increase in centrality for these residues could be a compensatory mechanism for upholding the mutants functional. Hence, the partial loss of function displayed by the protein in carrier patients (168, 176).

The mapping of collective drug bound *BC* hubs to their corresponding monomers suggested a long-distance proximal clustering communication pathway along the NADPH to 5-FU oxidoreduction catalytic channel per monomer (Figures 3.8.2 and S3.6.1). Through a chain of $Fe^{2+}{}_4S^{2-}{}_4$ clusters, this channel transports electrons from the NADPH binding site in

domain III to the catalytic site for pyrimidine (5-fluorouracil) reduction in domain IV (129-131, 147). Unlike our previous studies (283, 287), which depicted a well-defined communication path, this study and Barozi et al reported a clustering communication region of large *BC* hub ensembles (285, 286). The following were some of the common hubs that formed these ensembles from the NADPH to the active site in drug bound system of native protein (Table S3.5.1):

- Monomer A: Ile61, Thr65, Leu80, Asn85, Ile97, Ile101, Thr102, Asn108, Leu114, Ile116, Phe117, Leu122, Gly123, Thr125, Cys126, Cys130, Thr132, Ser133, Val137, Cys140, Leu142, Leu155, Leu200, Ser227, Ile231, Phe234, Leu310, Ala372, Val373, Pro374, Glu375, Gly597, Met598, Gln604, Ser605, Thr768, Leu775, Arg776, Asp797, Ala799, Ley803, Gln804, Cys816, Thr825, Lys958, Thr962, Cys996, Lys997 = 48
- Monomer B: Thr22, Asn44, Ile62, Glu69, Asn76, Met77, Pro92, Leu95, Ile97, Ile101, Thr102, Ala113, Ile116, Asn120, Leu124, Thr125, Gly127, Pro131, Ser133, Leu135, Asn137, Gly138, Asn141, Leu142, Thr145, Pro149, Leu155, Pro228, Ile231, Phe234, Ile370, Ala372, Val373, Pro374, Ala380, Glu383, Thr595, Phe602, Gly603, Sey605, Leu775, Ala799, Gln822, Thr853 and Thr856 = 45

These *BC* hubs were mostly found in the interface areas, around the $Fe^{2+}{}_4S^{2-}{}_4$ clusters and FMN (Table S3.4). In both the drug free and drug bound systems, the large hub ensembles that were detected in the WT were lost in the majority of the mutant systems, and new ones were gained, particularly around cofactors and the substrate (Table S3.5.1). For instance, mutant systems gained importance (form new connections not seen in reference WT) with either of these hub residues Ala144 (Monomer A: C29R, M166V, & Y186C and Monomer B: All except C29R & S534N); Glu146 (Monomer A: All except D949V and Monomer B: C29R); Glu147 (Monomer A: All except C29R & S534N and Monomer B: All except D949V). Nonetheless, these *BC* hubs (Ala144, Glu146, Glu147, Ile150, and Gly153) have been introduced in domain I around the $Fe^{2+}{}_4S^{2-}{}_4$. They have increased interaction with residue connecting Cys140 and Gln156 in comparison to the WT. Cys140 and Gln150 form disulfide bridges with $Fe^{2+}{}_4S^{2-}{}_4$ clusters (147). Protein interface residues, as well as $Fe^{2+}{}_4S^{2-}{}_4$ clusters are known to be involved in inter- and intra-protein communication as well as providing structural stability (440). Metalloproteins also take part in electron transfer, oxidation, oxygenation and oxygen transport (441). The presence of these hubs may indicate the existence of a compensatory communication path to the active site as a result of mutations

which may disrupt the redox process. Furthermore, the majority of mutations formed new contact with the active site hub and FMN interacting residues Gln604 (Monomer A: All except M166V and Monomer B: All except C29R); Pro772 (Monomer A: M166V and Monomer B: S534N and D949V), but lost contact with interacting residues Thr768 and Ala (Monomer A: All systems), as well as Ser605 (Monomer A: All except M166V & Y186C and Monomer B: All except I543V) with respect to the WT. Similarly, the presence of new hubs may indicate a compensatory communication pathway that may disrupt active loop functionality, resulting in increased fluctuation in the presence of mutation, as indicated by RMSF results (Figure 3.6C). In turn, this may impair protein structure and functionality. For instance, partially impaired functionality has been reported in Caucasian D949V carriers (162). CPIC has classified this group as intermediate metabolizers (164). On the other hand, the S534N mutation has been classified as normal metabolizers by CPIC (164). However this variant has been reported to have lower activity in the Caucasian population, 71% of carriers have been found to develop grade 3-4 toxicity (426, 427). In addition, increased drug toxicity was observed in Indian carriers (442). Surprisingly, CPIC classified C29R as having no effect on DPD activity, despite clinical findings to the contrary (164). The carrier of C29R (ethnicity unknown) has been reported to have grade 3-4 toxicity. There was an increase in 5-FU concentration three hours after administration (429). Similarly, South Indian carriers have an increased risk of thrombocytopenia (430). Furthermore, African American homozygous carriers had 127% DPD activity compared to the WT (168). Despite the fact that M166V was not mentioned in CPIC (164), increased DPD activity (120%) (168), as well as increased grade > 2 toxicity in Indians has been reported (443). Similarly, 5-FU administration caused grade 2 toxicity in I543V Indian carriers (428). The partial loss of functionality seen in all of these mutations (168, 176), could be due to a compensatory mechanism used by the protein to maintain functionality. For instance, compensatory gain was observed with the newly introduced hub Lys381, which had the highest *BC* centrality (0.047) among all drug-bound systems. Contact map analysis, discussed later in this chapter, revealed that this hub formed connections with Pro45, Ile62, His64, Thr132, Pro304, Phe367, Val368, Met377, Glu378, Leu379, Ala380, Lys381, Glu383, Lys384, Cys385 and Phe387, hence, the high hub centrality status (Figure 3.3.4).

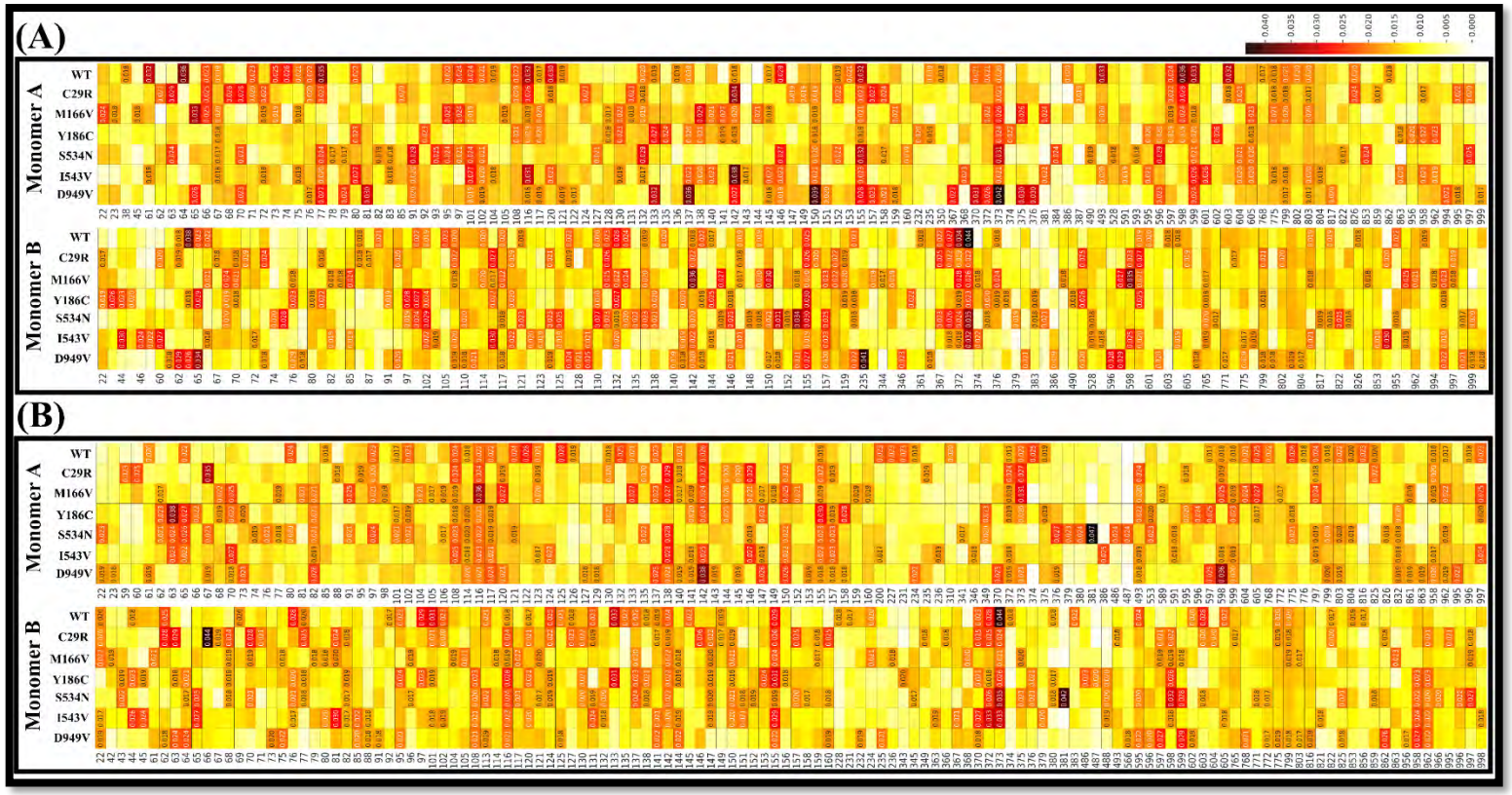


Figure 3.8.1: Averaged *BC* metric heatmap of the A) drug free and B) drug bound ensemble showing the top 4.0 % global significant hubs (x-axis) of the WT and six mutations (y-axis). The hubs are color-coded from low to high value (light-yellow to dark-red) and annotated accordingly.

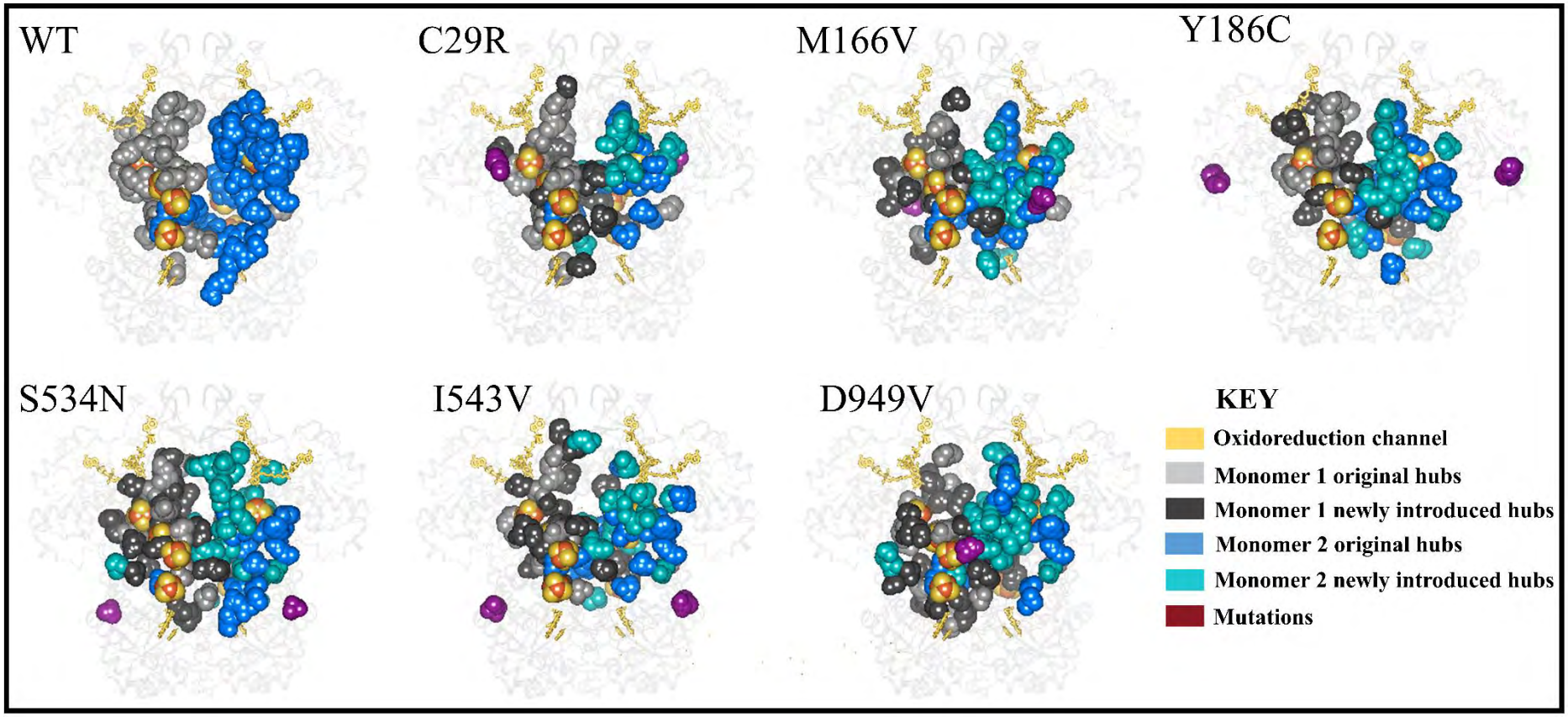


Figure 3.8.2: 3D representation of averaged *betweenness centrality* hubs of the activated (drug bound) ensemble showing a pair of communication pathways proximal to oxidoreduction channel (Monomer A NADPH (yellow) to FAD (yellow) to 2 x $Fe^{2+}{}_4S^{2-}{}_4$ (yellow) clusters, proceeding to Monomer B 2 x $Fe^{2+}{}_4S^{2-}{}_4$ (yellow) to FMN (yellow) for the reduction of 5-FU (yellow)) and vice-versa. The mutations are colored in deep-purple.

3.3.3.2 The *degree of centrality (DC)* identified a pair of distinct compensatory communication pathways from NADPH to 5-FU along the exterior of the oxidoreduction catalytic channel.

DC is determined by the number of adjacent nodes to a given node; consequently, these hubs are considered critical at the local level in regards to their functionality (284, 374). The data presentation of the intermonomer and intersystem heatmaps revealed that the monomers have moderately asymmetric behavior (Figure 3.9.1), which was also observed in the *BC* data (Figure 3.8.1). Considering the average top 4.0% *DC* hubs in DPD, persistent averaged *DC* hubs comprised of residues Asn108, Thr110, Ala113, Leu155, Ala203, Thr554, Val625, Val778*, Thr779, Ile781, Ile790, Ser808, Ala810, Leu813 and Thr832. According to other studies, *DC* had the fewest persistent hubs. (285, 286). In comparison to other metrics such as averaged *BC*, *CC, EC,* and *KC*, this metric comprised the second most persistent hubs in this study (Figure 3.9.1). This could be attributed to the importance of the oxidoreduction catalytic channel (NADPH, FAD, $Fe^{2+}{}_4S^{2-}{}_4$ clusters, FMN, and 5-FU) components which are distributed cross - sectionally in the protein.

3D visualization of this metric showed that the averaged *DC* hubs revealed distinct communication pathway per monomer in both drug free and drug bound. Most strikingly, these communication pathways were accentuated along the exterior of oxidoreduction catalytic channel (NADPH, FAD, $Fe^{2+}{}_4S^{2-}{}_4$ clusters, FMN, and 5-FU) (Figure 3.9.2 and Figure S3.6.2), more externally unlike the *BC* pathway. In respect to the WT, these are some of the hubs that formed ensembles from NADPH to the active site (5-FU) (Table S3.5.2):

- Monomer A: Ala72, Asn108, Tyr110, Ala113, Leu155, Gln157, Asn199, Ile200, Asn203, Leu206, Asn207, Glu218, Pro232, Val313, Ala314, Val335, Ser350, Ala351, Asn478, Gly480, Phe542, Thr554, Ser605, Val625, Leu628, Ile641, Leu653, Asn654, Leu665, Glu666, Ile707, Val714, Ala729, Val732, Gln740, Val778, Thr779, Ile781, Ile790, His807, Ser808, Leu813, Gly833, Leu837 and Asn928 = 45.
- Monomer B: Ala72, Asn108, Tyr110, Ala113, Gly138, Gly139, Gly153, Leu155, Phe159, Ala203, Ala207, Gly225, Val241, Val337, Ala356, Ala478, Trp568, Thr595, Gly603, Se605, Val625, Leu653, Ser654, Ile694, Ile707, Ala729, Val732, Val778, Thr779, Ile781, Ile790, Ile796, His807, Ser808, Thr832, Gly833, Ala928 and Thr962 = 38.

In reference to the WT, the presence of the mutation resulted in a loss of contact with Pro232, Val313, Val335, Ser350, Ala351, Phe542, Ile641, and Val714 of Monomer A, as well as Val241, Gly603 and Ile796 of Monomer B. The fact that the number of hubs and communication pattern (gain and loss of hubs) in both monomers were not completely identical, as shown above, confirms the asymmetric behavior of monomers/protomers in dimeric proteins. Sheik Amamuddy et al., previously reported on this concept (266). Furthermore, compensatory gain of *DC* hubs was observed in the presence of mutations relative to the WT around the following cofactors:

i) FAD in C29R (Monomer A: Ala198, Ser201 & Cys202 and Monomer B: Glu218), M166V (Monomer A: Ala198, Ser201 & Cys202), Y186C (Monomer A: Ala198, Ser201 & Cys202 and Monomer B: Ser199 & Glu218), S534N (Monomer A: Ala198, Ser201 & Cys202), I543V (Monomer A: Ala198, Ser201 & Cys202 and Monomer B: Pro197, Ala198, Ser201, Cys202 & Glu218) and D949V (Monomer A: Ala198, Ser201 & Cys202 and Monomer B: Cys202).

ii) FMN in C29R (Monomer A: Ala810 Monomer B: Leu813), M166V (Monomer A: Ile770 and Monomer B: Ile770 & Leu813), Y186C (Monomer B: Leu775 & Leu813), S534N (Monomer B: Leu775 & Leu813), I543V (Monomer B: Leu813) and D949V (Monomer A: Leu775 & Ala810 and Monomer B: Ala810 & Leu813).

iii) NADPH in C29R (Monomer A: Ile482, Ala486, Ser492, Val493, Asp495 & Gly496 and Monomer B: Ile482), Y186C (Monomer A: Ile482, Ala486), S534N (Monomer A: Ile482, Val493, Asp495 & Gly496 and Monomer B: Ile482 & Gly496), I543V (Monomer A: Ile482, Ala486 & Gly496 and Monomer B: Ile482, Ala486, Val493 & Asp495) and D949V (Monomer A: Val493 and Monomer B: Ile482, Ala486 & Asp495).

iv) 5-FU in C29R (Monomer A: Thr595 and Monomer B: Ala618 & Leu710), M166V (Monomer A: Thr595), Y186C (Monomer A: Thr595 & Leu710 and Monomer B: Leu608), S534N (Monomer A: Thr595 & Val718 and Monomer B: Val691), I543V (Monomer A: Thr595 & 710) and D949V (Monomer A: Thr595, Pro672, Gly681 and Monomer B: Ser606).

Asn108, Leu155, and Val778 were identified as persistent hubs with the greatest connectivity. Strikingly, Asn108 was discovered to connect three domains: domain I (first set of $Fe^{2+}{}_4S^{2-}{}_4$ clusters and residue 27-172); active site domain IV (FMN, 5-FU, and residues 525-847); and domain V (second set of $Fe^{2+}{}_4S^{2-}{}_4$ clusters and residues 848-1025). Leu155 is a super-persistent hub (recognized in several metrics) in domain I's α-helix, connecting to

Gln156, which forms a one-of-a-kind bond with Fe^{2+} ion. Met599 is also noteworthy as an interface residue between domain I and the active site (domain IV). The evolutionarily conserved α-helix high significant hub Val778, in contrast, plays an important role within the active site area, forming contacts with FMN and 5-FU connecting residues. Val778 basically acts as an intermediary within the active site area. Ala499 and Gly603 were exclusively drug bound mutant system hubs, especially in monomer A. While domain III α-helix residue Ala499 was newly introduced hub in the mutant systems of Monomer A, residue Gly603 located within the active site domain lost contacts with all the mutant, specifically in Monomer B. Furthermore S534N, I543V and D949V exhibited reduced residue connectivity with Trp568, Gly603, SER605 and SER606, which are active site residues. It should be noted that S534N and I543V are active site mutations, whereas D949V is a domain V mutation, all of which have clinically shown partial loss of function in carrier patients (428, 444, 445). Mutations provoked change in inter- and intra-residue interaction resulting in the formation of new communication patterns to the active site. In the process, a compensatory residue-residue communication mechanism that ensures cellular process is maintained ensued. Studies have reported such mechanism in the context of *BC* (287, 374, 446).

Overall, we demonstrated for the first time that: i) the asymmetric behavior of DPD protein, ii) two distinct communication pathways per monomer were observed using both *BC* and *DC*. The *BC* pathway was located anterior to the oxidoreduction channel, while the *DC* pathway was located exterior to the oxidoreduction pathway. Furthermore, A direct cross-correlation analysis was conducted to investigate the relationship *BC* and *DC* hub residues in both the drug free and drug bound state of DPD enzyme. The results revealed a significant positive correlation between these two measures. indicating their pivotal role in maintaining functional integrity of the network (Figures S3.7.3 and Figures S3.7.4). iii) The compensatory gain in *DC* hubs along the oxidoreduction pathway to the active site may point to a mechanism for maintaining DPD functionality.

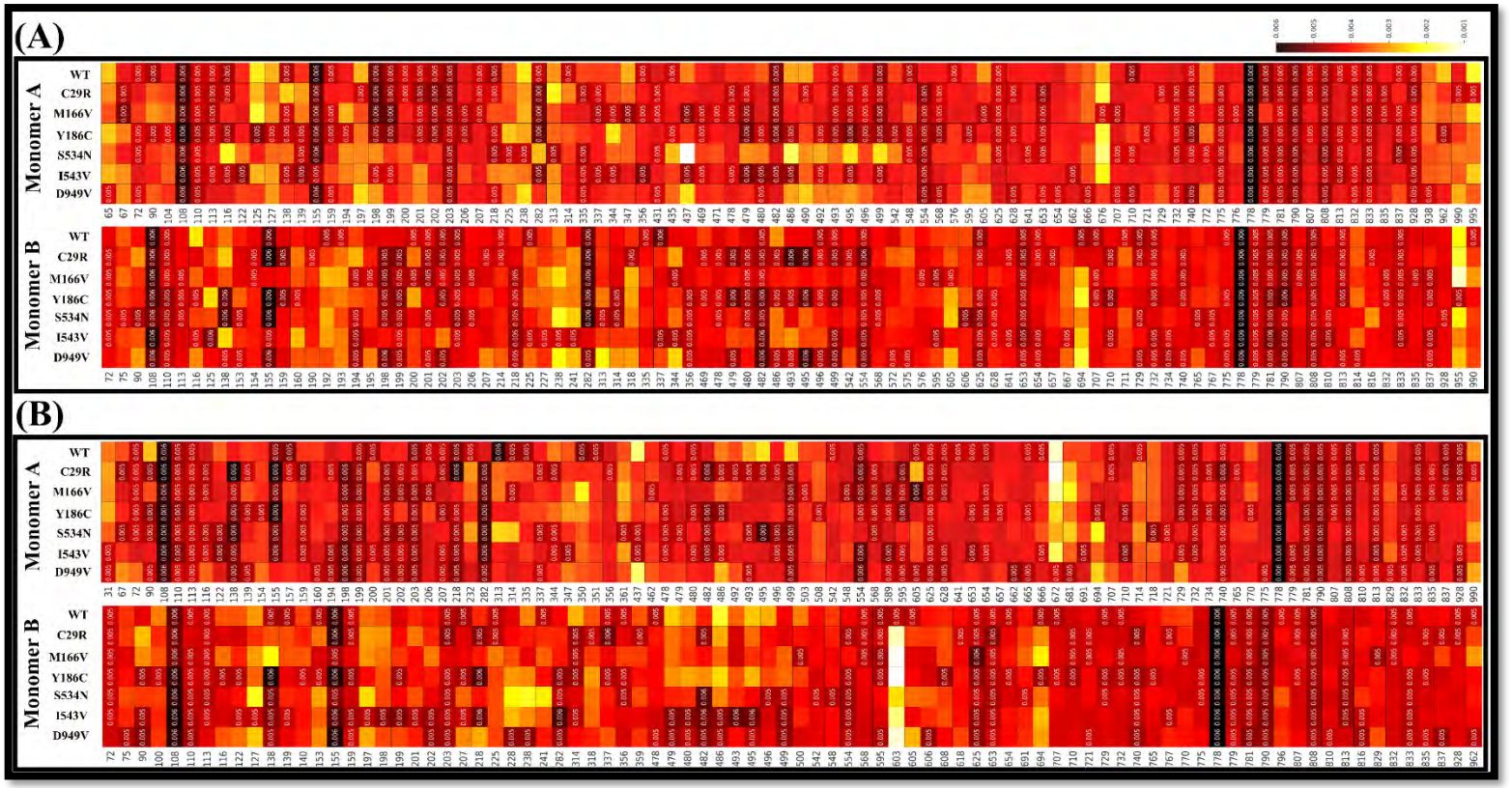


Figure 3.9.1: Averaged *DC* metric heatmap of the A) drug free and B) drug bound ensemble showing the top 4.0% global significant hubs (x-axis) of the WT and six mutations (y-axis). The hubs are color-coded from low to high value (light-yellow to dark-red) and annotated accordingly.

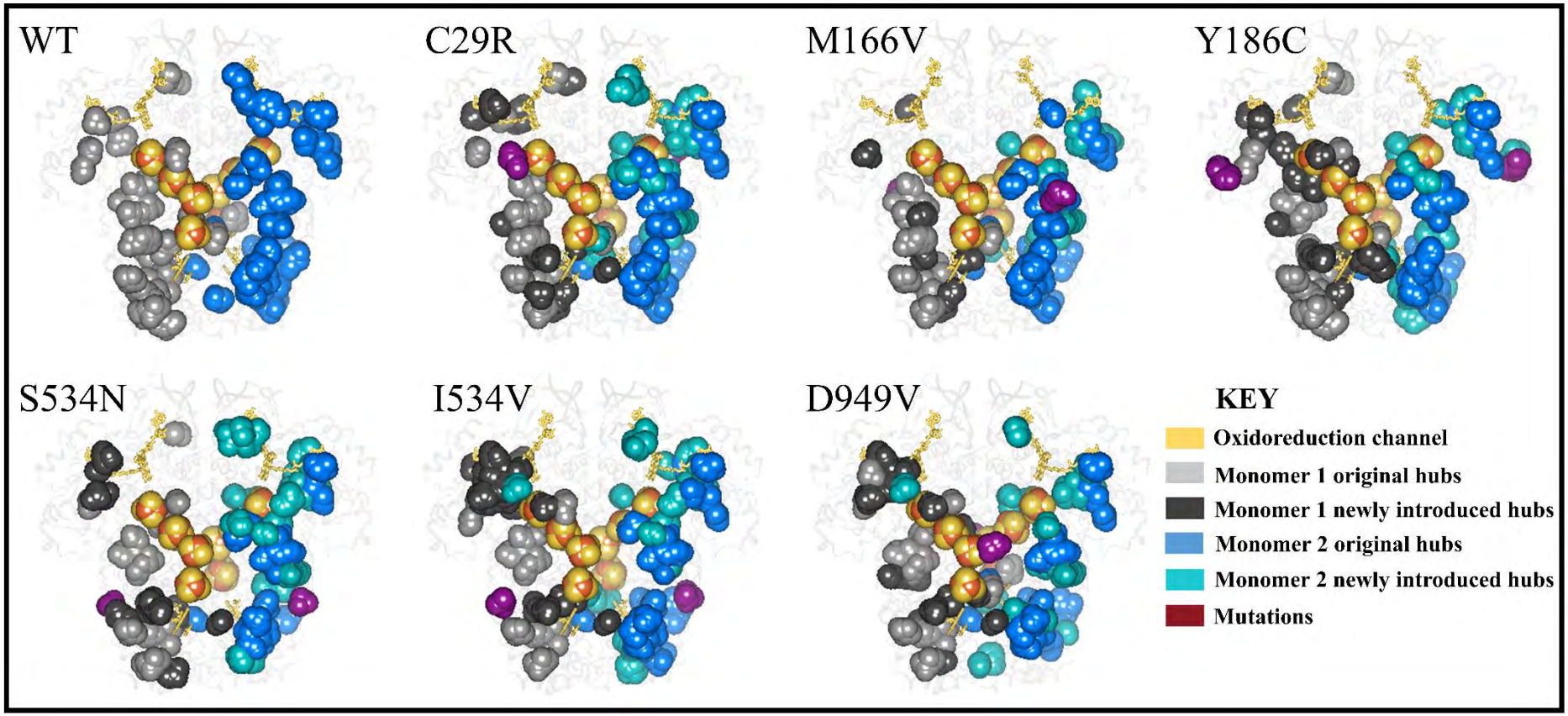


Figure 3.9.2: 3D representation of averaged *degree centrality* hubs of the activated (drug bound) ensemble showing a pair of communication pathways proximal to oxidoreduction channel (Monomer A NADPH (yellow) to FAD (yellow) to 2 x $Fe^{2+}{}_4S^{2-}{}_4$ (yellow) clusters, proceeding to Monomer B 2 x $Fe^{2+}{}_4S^{2-}{}_4$ (yellow) to FMN (yellow) for the reduction of 5-FU (yellow)) and vice-versa. The mutations are colored in deep-purple.

3.3.3.3 The *Closeness centrality (CC)* revealed the dynamic dimeric hub network behavior at the proteins cores in the presence of mutations

CC defines the central nodes that are closest to the majority of the other nodes. It is determined as the inverse of the average for the shortest distances from one node to every other node. Our preceding studies have demonstrated that high *CC* values occur in and around the protein core, specifically around the $Fe^{2+}{}_4S^{2-}{}_4$ clusters (283, 287, 374). Figure 3.10.1 depicts a heatmap of the top 4.0% averaged *CC* hubs, while the 3D mapping was illustrated in Figure 3.10.2 (drug bound) and Figure S3.6.2 (drug free). The intermonomer and intersystem heatmap presentation of the data revealed that the monomers exhibit slightly asymmetric behavior (Figure S6), which was also reported in the *BC and DC* data (Figure 3.10.1).Visualization of these residues revealed that they were at the core of the DPD protein (Figures 3.10.2 and S3.6.3), which was in accordance with previous research on other proteins (39). These hubs were mainly around the $Fe^{2+}{}_4S^{2-}{}_4$ clusters and active site (Figure 3.10.2 and Figure S3.6.2). In comparison to *BC, DC, KC*, and *EC*, this metric (*CC*) had the most persistent hubs (G71, Ala72, Leu73, Ala74, Glu75, Ala76, Cys79, Leu80, Ala101, Ile104, Leu142, Ile150, Leu155, Pro598, Met599, Pro602, and Ser605) (Table S3.5.3). Further analysis of *CC* hubs revealed most of these interface residue are located in the stable part of the dimer, a concept that has been reported before (147). Of interest is persistent hub Leu155 which was identified as a persistent *BC, DC* and *CC* hub hence, earning it the term "super-persistent hub" (266). This hub connects to the following residues: Leu95, Phe100, Ile116, Asn120, Gly123, Met128, Ile152, Gly154, Gly155, Gln157, Phe158, Ala159 and Thr160, forming a domain I inter-residue communication sub-network that relays interdomain information around the $Fe^{2+}{}_4S^{2-}{}_4$ clusters. Thus, the portrayed stability around the $Fe^{2+}{}_4S^{2-}{}_4$ clusters (147). Previous studies have shown the correlation between intermonomer distance measure and the number of *CC* hubs due to presence of mutation (285). Except for systems S534N, M166V, and D949V, where a slight (<10.0 Å) increase in COM distance in comparison to the WT was noted (Figure 3.7), no apparent increase in COM distances was observed in the remaining systems. These three mutant systems have exhibited fewer *CC* hubs compared to other systems (Figure 3.10.2: C, E and G). Furthermore comparative ED analysis demonstrated great conformation variation in relation to the WT, which is in agreement with other prior studies (266, 285-287).

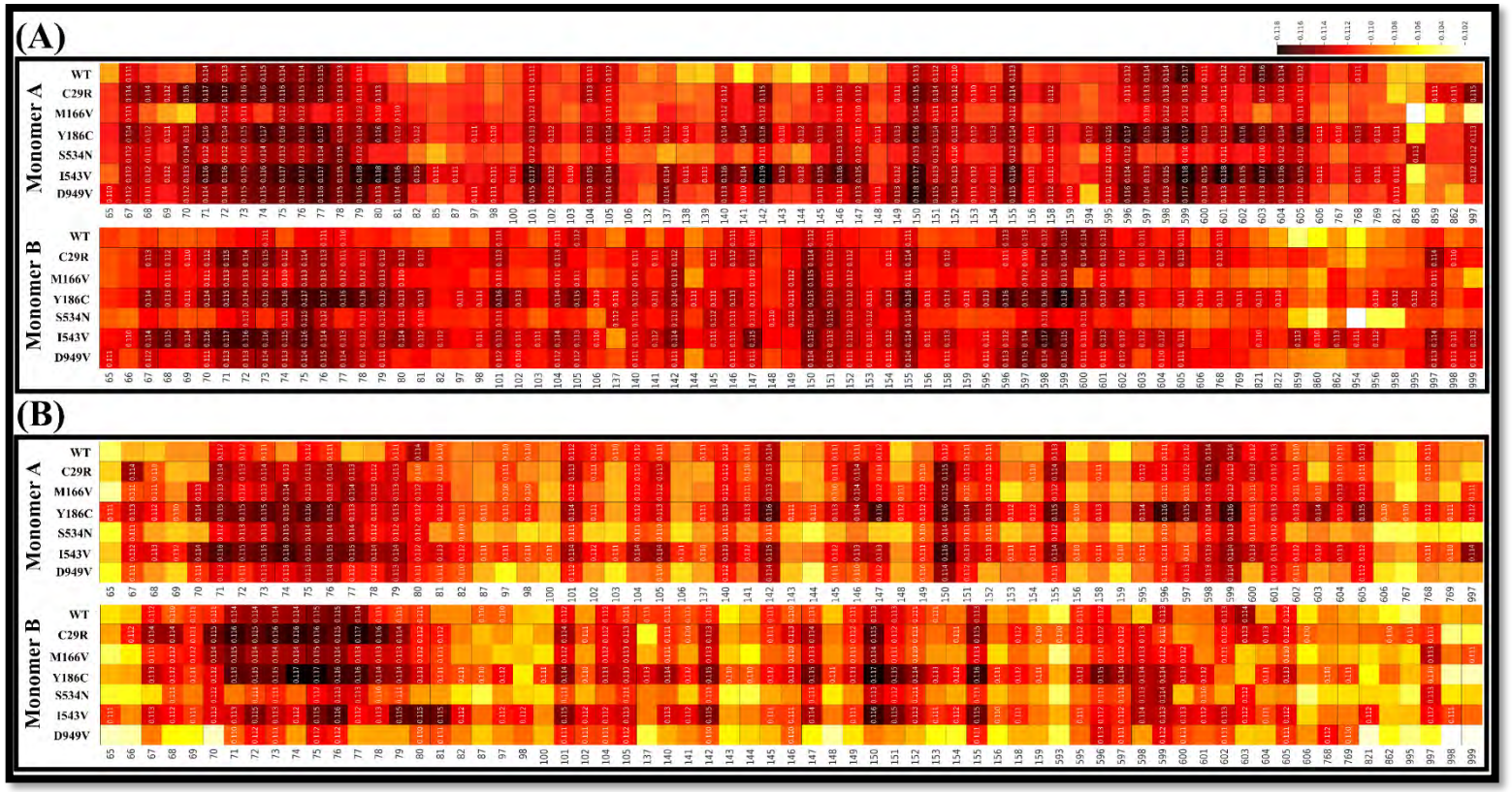


Figure 3.10.1: Averaged *CC* metric heatmap of the A) drug free and B) drug bound ensemble showing the top 4.0% global significant hubs (x-axis) of the WT and six mutations (y-axis). The hubs are color-coded from low to high value (light-yellow to dark-red) and annotated accordingly.

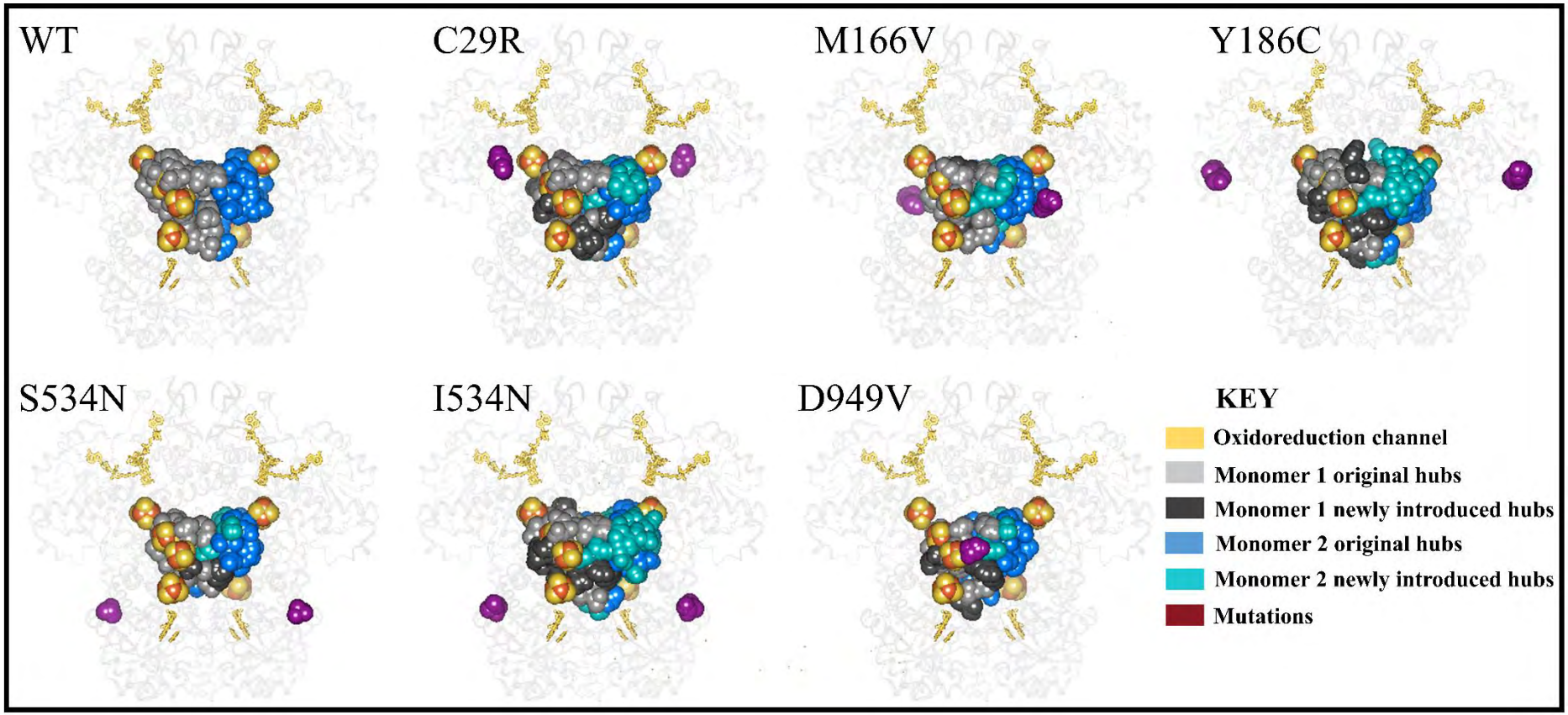


Figure 3.10.2: 3D representation of averaged *closeness centrality* hubs of the activated (drug bound) ensemble showing a pair of communication pathways proximal to oxidoreduction channel (Monomer A NADPH (yellow) to FAD (yellow) to 2 x $Fe^{2+}{}_4S^{2-}{}_4$ (yellow) clusters, proceeding to Monomer B 2 x $Fe^{2+}{}_4S^{2-}{}_4$ (yellow) to FMN (yellow) for the reduction of 5-FU (yellow)) and vice-versa. The mutations are colored in deep-purple.

3.3.3.4 The *Eigencentrality (EC)* revealed the dynamic dimeric hub network behaviour at the active site in the presence of mutations

EC is used to identify high connectivity nodes which are adjacent to other high connectivity nodes and as such are reliant on the residue neighborhood (284, 287, 374). Similar to other metrics, the clear asymmetric behavior of each monomer was demonstrated (Figure 3.11.1). There were no persistent hubs of averaged *EC* for DPD reference protein. In the WT, *EC* hubs (which were identified mostly in the active site) included residues at position:

- 103-115, 159, 605-608, 707-712, 724, 732, 734, 772, 774, 736, 776-795, 799-812, 829-837, 852, 928 of Monomer A.
- 595, 707-710, 721, 732, 734, 736, 740-741, 744, 775-792 and 805-813 of Monomer B

Previously, a high concentration of high *EC* hubs in a domain was associated with its stability(285), and could be used to identify ligand-binding site (447). While the WT had the most *EC* hubs in Monomer A (69), mutant S534N (69) had the most hubs in monomer B compared to WT and other mutant systems, indicating asymmetric DPD behaviour. Asymmetric behaviour was also observed in C29R, M166V, and I543V, which lacked *EC* hubs in the active site of Monomer A (C29R) and Monomer B (M166V and I543V), respectively. Additionally, loss of active site *EC* hubs in the mutant system may be associated in the active site domain instability, which may account for drug exit in at least one monomer of some mutant systems (Monomer A: S534N and Monomer B: I543N and M166V). Hence, the partial loss of function depicted by these mutants (168, 186).

The 3-dimensional depiction of the WT structure (Figures 3.11.2 and S3.6.4) concerning the hubs that were present after DPD activation (Table S3.5.4) indicated that (1) central placement of *EC* hub clusters around the active site region, indicating its importance in DPD functionality (Figure S3.11.2), and (2) the DPD monomers demonstrated asymmetric behavior regardless of the presence of mutations. This is in agreement of previous *EC* studies on other proteins (266, 285).

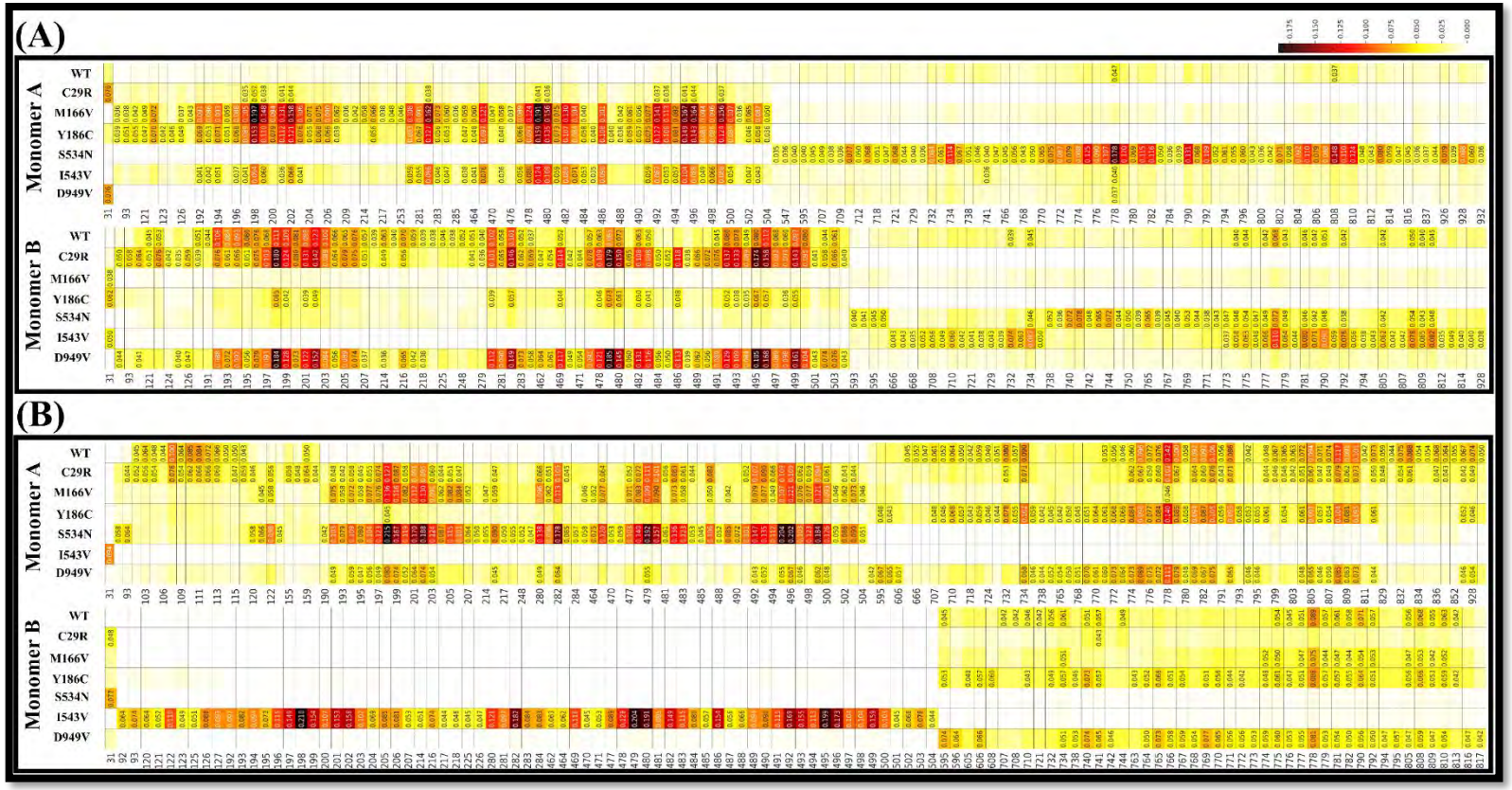


Figure 3.11.1: Averaged *EC* metric heatmap of the A) drug free and B) drug bound ensemble showing the top 4.0% global significant hubs (x-axis) of the WT and six mutations (y-axis). The hubs are color-coded from low to high value (light-yellow to dark-red) and annotated accordingly.

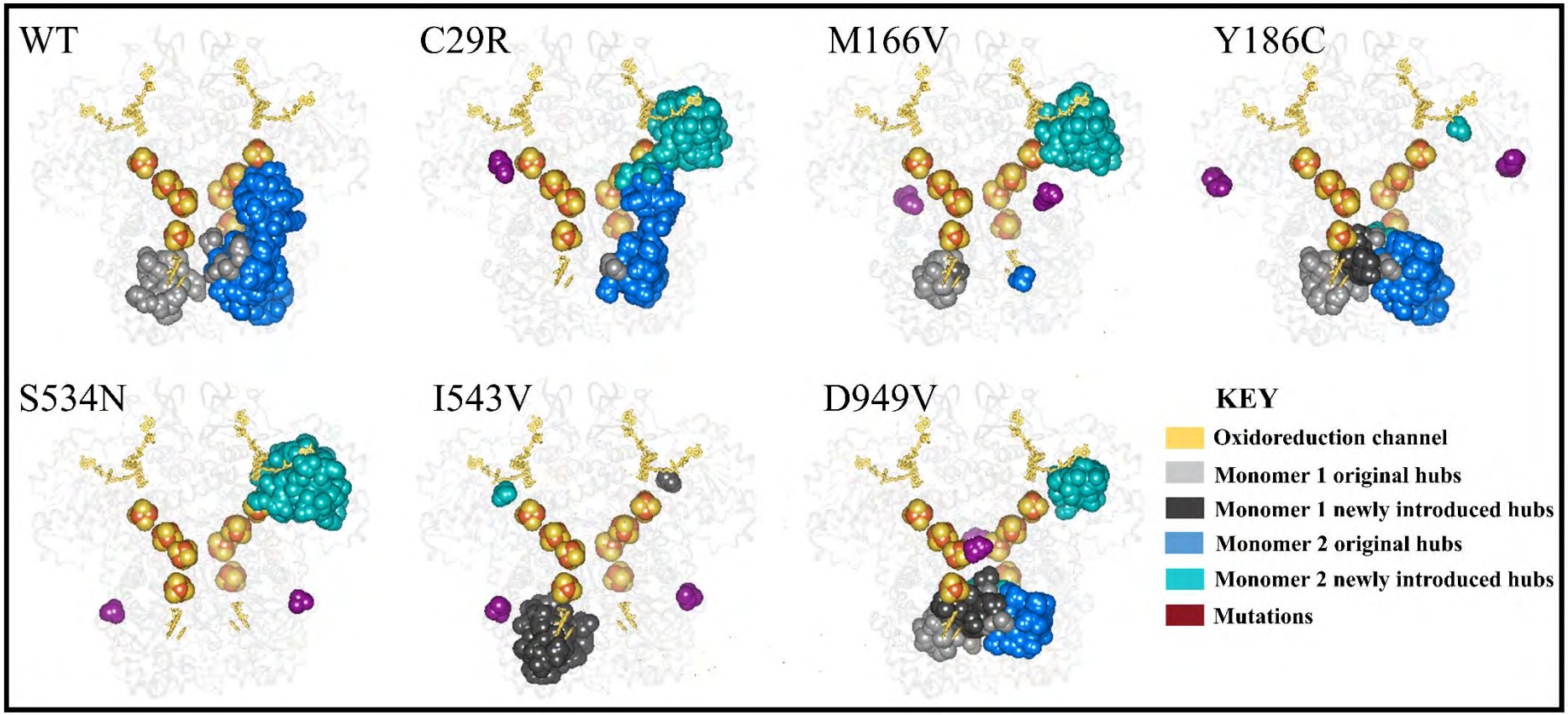


Figure 3.11.2: 3D representation of averaged *eigen centrality* hubs of the activated (drug bound) ensemble showing a pair of communication pathways proximal to oxidoreduction channel (Monomer A NADPH (yellow) to FAD (yellow) to 2 x $Fe^{2+}{}_4S^{2-}{}_4$ (yellow) clusters, proceeding to Monomer B 2 x $Fe^{2+}{}_4S^{2-}{}_4$ (yellow) to FMN (yellow) for the reduction of 5-FU (yellow)) and vice-versa. The mutations are colored in deep-purple.

3.3.3.5 The *Katz centrality (KC)* revealed the dynamic dimeric hub network behavior around the cofactors and active site in the presence of mutations

KC is defined by the relative influence of each node in a specified network by considering both its immediate nodes and non-immediate nodes that are linked via immediate neighbours. Basically, *KC* is an extension of *EC* (284, 287, 374). When comparing *Katz centrality* to *EC*, a study by Barozi et al., (374) observed a more uniform distribution of *KC* value, whereas a study by Okeke et al., (287) discovered that key hubs, particularly of the mutant proteins, showed relatively similar results to their *EC* values. Likewise, in our study, *KC* hubs were an extension of *EC*, emphasizing not only the importance of active site-like *EC* values but also cofactors. Intersystem and intermonomer heatmaps illustrated the asymmetric behaviour of the dimeric DPD (Figure 3.12.1). In drug bound system, Monomer A and Monomer B of the WT had 40 and 24 hubs, respectively. The drug free systems had 36 and 46 hubs in Monomer A and B, respectively. Similar trends were witnessed, in the mutant systems (Table S3.5.5). For instance, I543V gained most newly formed hubs in Monomer A (33) than any other mutant system. Also, S534N gained 33 and 5 newly formed hubs in Monomer A and B respectively.

Figure 3.12.1 depicts a heatmap of the top 4.0% *KC* hubs, with 15 key persistent hubs (Asn108, Thr110, Gly111, Leu155, Leu775, Asn777, Val778, Thr779, Ile781, Asn782, Ile790, Phe805, Ser808, Ala810, and Gly833) identified (Table S3.5.5). The super-persistent hubs Asn108, Leu155 and V778 were present in *BC, CC and DC* metric too.

3D visualization of key *KC* hubs communication network (Figures 3.12.2 and S3.6.5) identified four important central locations of these hubs (iron clusters; FAD, FMN and the 5-FU binding site). In the WT, some of the *KC* hubs included the following residues:

- Monomer A: Ser103, Ile104, Tyr109, Tyr110, Gly111, Ala112, Ala113, Leu155, Ser199, Ala203, Ser605, Ser640, Ile641, Leu665, Glu666, Leu667, Phe707, Leu710, Val732, Ala734, 775, 776, Ala777, Val778, Ther779, Thr781, Ile782, Ile790, Ala792, Phe805, Ser808, Gly809, Ala810, Leu813, Thr832, Gly833, Ala836, Leu837 and Ala928.
- Monomer B: Asn108, Tyr110, Gly111, Leu155, Ser159, Thr595, Phe707, Ala708, Leu710, Val732, Ala734, Leu775, Ala777, Val778, Thr779, Thr781, Ile782, Ile790, Ala792, Phe805, Ser808, Gly809, Ala810 and GLy833.

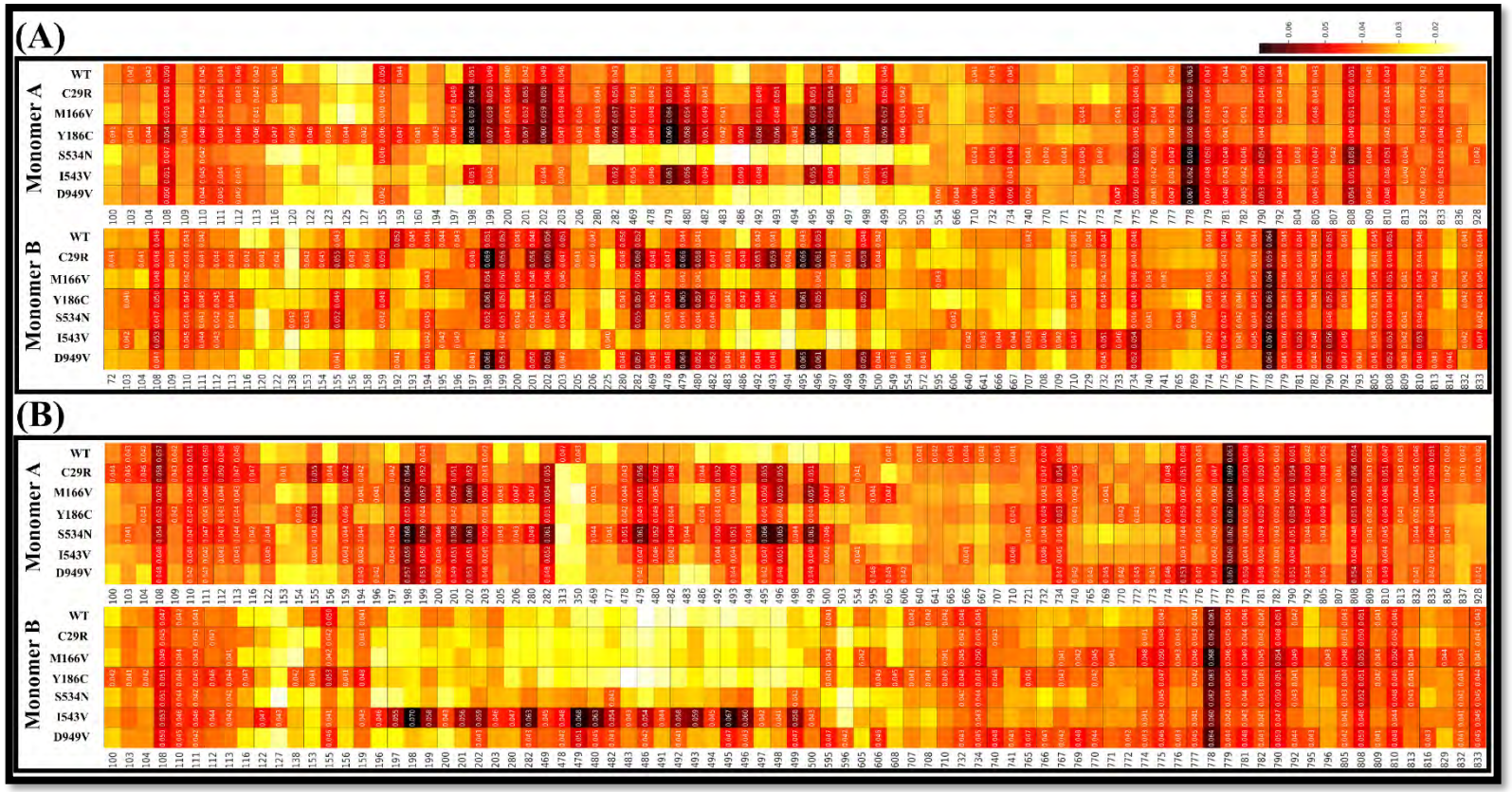


Figure 3.12.1: Averaged *KC* metric heatmap of the A) drug free and B) drug bound ensemble showing the top 4.0% global significant hubs (x-axis) of the WT and six mutations (y-axis). The hubs are color-coded from low to high value (light-yellow to dark red) and annotated accordingly.

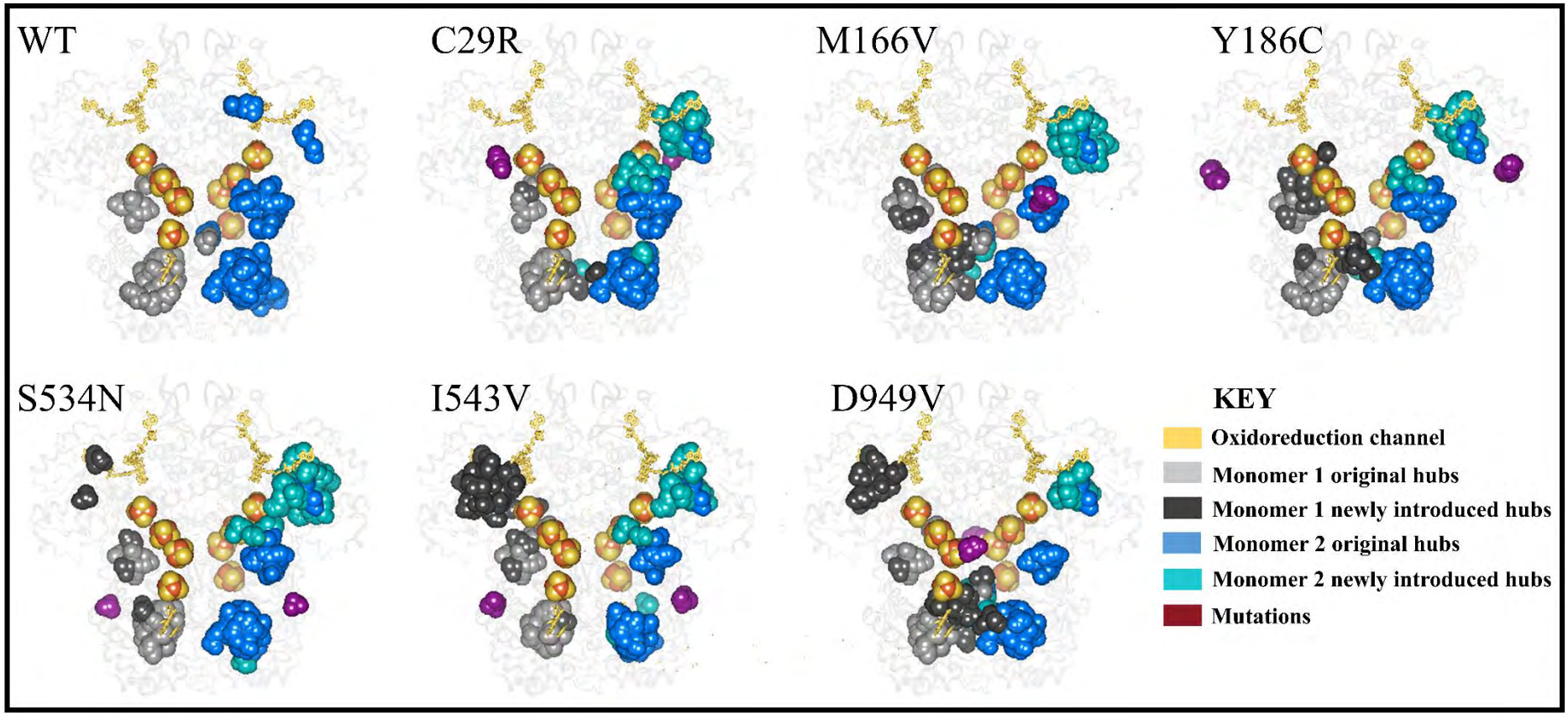


Figure 3.12.2: 3D representation of averaged *KC* hubs of the activated (drug bound) ensemble showing a pair of communication pathways proximal to oxidoreduction channel (Monomer A NADPH (yellow) to FAD (yellow) to 2 x $Fe^{2+}{}_4S^{2-}{}_4$ (yellow) clusters, proceeding to Monomer B 2 x $Fe^{2+}{}_4S^{2-}{}_4$ (yellow) to FMN (yellow) for the reduction of 5-FU (yellow)) and vice-versa. The mutations are colored in deep-purple.

Residues 108-113 and 155 are interconnecting residue to domain I $Fe^{2+}_4S^{2-}_4$ clusters, residues 199, 203, 313 and 350 are interconnecting residues to FAD, residues 605, 640, 641, 665, 666, 667, 707, 710, 732 and 733 are active site residues (residues involving 5-FU), residue 775-805, 809-813, 832 and 833 are interconnecting residues to FMN. These findings further support the vitality of the oxidoreduction channel components (NADPH, FAD, $Fe^{2+}_4S^{2-}_4$ clusters, FMN and 5-FU) and identify specific regions of interest in these domains.

3.3.4 Mutational impacts on residue interactions as determined by weighted contact maps

3.3.4.1

In section 3.3 (averaged DRN metric hub differences due to mutations were investigated). The impact of mutations on the DPD structure by identifying DRN hubs from penta network metrics showed: asymmetric inter-monomer communication network, as well as compensatory gain in hub communication network. This gave us insight into the centrality hubs and how they differ in different mutant protein systems. During catalysis, residual rearrangement transpires not solely at the point of mutation, but also at the specific cofactor level. The loss of the residue-residue interaction and the formation of the new connections triggered a change in monomeric residue interaction in the specific cofactors, which varied from mutation to mutation (266, 448). The degree of residue interactions with closest neighbors (within 6.7 Euclidean distance), graded from 0 to 1 (presence or absence) for (1) the mutated residues, (2) cofactors, (3) 5-FU binding sites, and (4) persistent hubs (Figure 3.13), was defined using the MDM-TASK weighted contact map tool. This method has previously been used to investigate the short-range mutational residue-residue change in interactions (286, 287, 374, 438, 446, 448). Figure 3.13 (consider Figure S3.7.1, S3.7.2, S3.7.3 & S3.7.4 for amplified version) revealed minor differences in residue contacts between monomers, which were ascribed to local asymmetry within chain conformations, which has been found to be common in complexes (285, 449).

NADPH binding domain III is required for reduction, hence, providing electrons for transport to the active site in domain IV (450, 451). The amino acid sequences of these domains are highly conserved across animal species, and the ability to form dimers is paramount in DPD activity because it is activated by dimerization and electron transport through the $Fe^{2+}_4S^{2-}_4$ clusters (130, 131). Despite the fact that none of the clinically studied mutations (C29R, M166V, Y186C, S534N, I543V and D949V) are found in NADPH binding

domain, change in the inter-residue communication relative to the WT was noted. NADPH comprises of binding residues Ala340, Thr343, Arg364, Lys365, Arg371, Gly484, and Ala486. Based on the WT, enhanced slightly asymmetrical NADPH-protein residue interaction was observed in both monomers (Figure 3.13, Figure S3.7.1 and Table 3.6.1). The following WT residues showed enhanced interactions:

- Monomer A: Asn48, Arg235, Gly339, Ala340, Gly341, Asp342, Thr343, Ala344, Asp346, Cys347, Ile361, Phe363, Arg364, Lys365, Gly366, Asn369, Ile370, Arg371, Ala372, Phe390, Leu391, Ser292, Thr411, Trp419, Gly480, Asp481, Val483, Gly484, Leu485, Ala486, Asn487, Thr488, Glu491 and Ser492.
- Monomer B: Gly339, Gly341, Asp342, Ala344, Asp346, Cys347, Val362, Phe363, Arg364, Lys365, Gly366, Asn369, Ile370, Ala372, Ser392, Thr411, Ala437, Cys480, Asp481, Val483, Gly484, Leu485, Ala486, Asn487 and Ser492.

Of these residues, Gly366, Asn369, Ala372, Phe390, Ala486 and Glu491 are interface residue critical for FAD co-ordination and stabilization (285, 440). While maintained interactions of these residues were observed in Monomer A, Monomer B, lost contact with Arg371 and Phe390 in presence of mutation. This supports the asymmetric interaction behavior of DPD, specifically in C29R, M166V, Y186C, and D949V (non-active site domain mutations). Furthermore, Ala344, Ile347, and Gly480 have been identified as high centrality *BC* or *DC* hubs, and they may play an important role in the inter-monomer compensatory allosteric communication mechanism that maintains protein function. Ala344 was identified in Monomer A of S534N and I543V but not in Monomer B, similarly Ile347 was identified in Monomer A of S534N but nor in Monomer B. Gly480 was identified in the WT, C29R, Y186C and I543V, but only in S534N and I543V of Monomer B.

FAD is reduced in the event of NADPH binding to DPD protein, the reduced equivalents are then passed to FMN via $Fe^{2+}{}_4S^{2-}{}_4$ clusters (129). Point mutations at the region closer to functional components of a protein, are known to hinder protein structure and function (452, 453). Y186C is a mutation located in the FMN binding domain. There were no noticeable changes in the inter-residue contact frequencies in Y186C, with the exception of residues Pro179, Glu184, Ala188 and Leu209 in Monomer A, which increased by more than 40%, and Glu184 and Leu209 in Monomer B, which increases by more than 40%. In respect to the WT (Figure S3.8.1). The replacement of tyrosine, which has a bulkier side chain than cystine, could have caused aberrant dimer crosslinking (176, 454) due to changes in inter-

residue relationships. The WT FAD comprises of binding residues: Ala198, Val129, Lys219, Leu226, Leu261, Asp481 and Thr489 with the following contacting residues:

- Monomer A: Thr125, Cys130, Met128, Val129, Pro198, Ser199, Ser201, Glu218, Gln220, Gly225, Ser227, Glu230, Ile231, Pro232, Phe233, Leu236, Ser260, Ser262, Met266, Thr343, Asp346, Ala450, Gly479, Gly480, Val482, Gly484, Ala486, Thr488, Val490 and Ser492
- Monomer B: Asp61, Cys126, Gly127, Cys130, Phe193, Gly194, Ala195, Pro197, Ala198, Ile200, Phe217, Leu226, Thr228, Glu230, Ile231, Pro232, Arg235, Cys257, Ser260, Ser262, Met266, Gly282, Cys284, Leu285, Ser446, Val447, Ala450, Gly480, Val482, Gly484, Ala486, Thr488, Val490, Ser492, Val493

Cys130 is an important residue that interacts with the iron cluster, forming a covalent bond, while Ala486 is an interphase residue important for protein stability(131, 440). In the WT, Cys130 portrayed no monomer differences. In contrast to I543V, S534N, M166V, and D949V exhibited asymmetric interaction behavior with increased or retention in monomer A in relation to monomer B, an opposite of C29R and Y186C. Seeing as C29R (domain I), M166V (domain I), Y186C (domain II), and D949V (domain IV) are all located in the non-active site domain, the change in FAD-protein interaction on monomer A in comparison to monomer B could be an implication of allosteric signaling. Furthermore, Lys497 is a good example of a residue that shows interaction differences between the two monomers. Due to the mutations, monomer B FAD lost contacts with Lys497. Furthermore, from the heatmap, FAD contacting residue in monomer B exhibited total loss of inter-residue interactions than those of monomer A when compared to the WT (Figure 3.13, Figure S3.7.2 and Table 3.6.2). There was a total loss of interaction between Ile283 and Gly479, in all systems. In contrast, a gain in H-bonds was observed between Ala198 and Ile281 in D949V, Val482 and Ala489, as well as Thr481 with Ala486 and Ser492. Monomer A's newly formed residue-residue interaction differed from that of Monomer B, supporting the asymmetric behavior of the monomers observed previously in this study. Interestingly, Ala198, Val482 and Ala486 have been identified as high centrality *BC* or *DC* hubs, and may be very vital in inter-monomer compensatory allosteric communication mechanism that maintain protein function (Figure 3.8.1 and 3.9.1). This could explain the partial loss of functionality in DPD degradation observed by carriers of the discussed missense mutation such as Y186C (455). Saif et al., discovered that African-American Y186C variant carriers experienced severe toxicity following 5-FU chemotherapy administration, which was linked to partial 5-FU medication

degradation (177). Further research study in a group of healthy Americans reported that DPD enzyme activity was 46% lower in PBMCs from Y186C carriers than in non-carriers (P=2.9104), implying that the mutation partially hampered the enzyme's catalytic activity (456).

Prior investigation, successfully identified and characterized an inter-residue compensatory allosteric communication mechanism in the ligand binding environment, employing the penta-DRN metric (59, 285, 286). Additionally, our study revealed the existence of comparable allosteric communication occurring at the NADPH and FAD levels. Building upon these findings, we further explored the localized effects of mutations specifically at the FMN level, aiming to elucidate their impact on the allosteric communication network. We investigated the local impact of mutation at the FMN level. This cofactor receives electrons from FAD via a series of $Fe^{2+}{}_4S^{2-}{}_4$ clusters which in turn reduces 5-FU (130). The native FMN contains the following residues: Lys574, Thr575, Lys709, Gly767, Thr793, Gly816 and Ser817, and it is found at the active-site (residues 525-847) domain IV. Similarly, the FMN-protein interaction was asymmetric, with increased FMN interaction on the following residues (Figure 3.13, Figure S3.7.3 and Table S3.6.3):

- Monomer A: Ala549, Ser550, Thr552, Ala554, Thr555, Lys574, Thr575, Phe576, Ser577, Ser606, Ile613, Ser614, Ser640, Met642, Glu666, Leu667, Asn668, Leu669, Ser670, Ala708, Leu710, Thr733, Thr735, Ser766, Gly767, Thr768, Ala769, Ile770, Arg771, Ala774, Ala792, Thr793, Gly794, Gly795, Ile795, Ile796, Asp797, Ser801, Leu813, Gln814, Val815, Cys816, Ser817, Ile819, Gln820 and Asn821 of the WT.
- Monomer B: Ala549, Ser550, Ala551, Thr555, Thr573, Lys574, Thr575, Phe576, Ser577, Ser606, Phe607, Ile613, Ser614, Lys616, Ser640, Ile641, Met642, Glu666, Leu667, Asn668, Leu669, Ser670, Cys671, Ala708, Leu710, Thr773, Asn736, Ser766, Gly767, Ile770, Arg771, Ala774, Ala792, Thr793, Gly794, Gly795, Ile796, Asp797, Gln814, Val815, Cys816, Ser817, Ala818 and Gln820 of the WT.

Some of the key inter-residue relationships within the active site include: catalytic residues; Ser670 and Cys671, as well as active site Asn668 and Asn736 that are in contact with FAD and interface residues Phe607, Val819, Gln820 and Asn821. Collectively these residues are important in ensuring protein stability (130, 131, 440). In respect to the WT, Cys671 exhibited different monomer behavior in presence of mutations, it lost contact with all systems in Monomer A. Based on RMSF results, this catalytic residue is located in the active site loop region, which was characterized by high fluctuation in the presence of mutation.

Additionally, using ED, diverse conformational variability was also noted, which was associated with enhanced active loop area fluctuation, particularly on active-site domain mutations (S534N and I543V), as well as D949V, and M166V. (Figure 3.6C). The asymmetric behaviors of monomer were observed when the WT and D949V increased residue-residue contact while the other systems lost these connections. Similar to Y186C, the active site domain mutations (S534N and I543V) have also been clinically implicated in DPD deficiency that partially impair protein function (146, 168, 445). These mutations (S534N and I543) are located directly opposite each other at the short α-hairpin-loop which forms the bottom of the $(\alpha/\beta)_{8-}$ or TIM-barrel of the active site domain IV. The side chain of S534 is stabilized with the carboxyl-group of Asp532, whereas I543 forms Van-der-Waals interaction with the backbone of G569, as such both residues appear to be fixed in their position in the native structure (180, 425) S534N formed new compensatory H-bonds with Glu536 and Asn544 (Monomer A) Phe542, pro545 and Thr920 (Monomer B), thus, the concept of allosteric signaling (285, 286). Whereas increased residue-residue connections were found in both monomers of S534N (Monomer A: Glu536, Asn544, and Lys541 and Monomer B: Asn544), no other significant differences were found when compared to the WT (Figure S3.8.2). Asn544 and Leu813 has been identified as a high *DC* hub, exclusive to Monomer B of mutations, resulting in compensatory allosteric communication mechanism. Similarly, inter-residue interaction within mutation I543V in comparison to the WT showed newly formed H-bond with Gly567 and Ser534 (Monomer A). There was loss of interaction with Phe570 and Asn635 (Monomer B), as well as reduced interaction of up to 2 folds in Asp634 (Monomer A) and Lys541 and Asp534 (see Figure S3.8.3). This might be due to the fact that valine was replaced by isoleucine. Although the two residues are aliphatic-hydrophobic amino acids with the same bulkiness, suggesting that they have the same properties, change in inter-residue communication was noted at the point of mutation (457). Further comparison of inter-residue interactions around Monomer A of FMN using the WT as a reference revealed that, M166V (7 interactions) and Y186C (7 interactions) lost the most inter-residue interactions than the other mutations [C29R (4 interactions), S534N (4 interactions), I543V (5 interactions), and D949V (5 interactions)] (Figure 3.13 and Table S3.6.3). New inter-residue interactions were formed: Monomer A [S534N, C29R and M166V (Asn552- Cys816); I543V, D949V (Ser670-Lys709) and C29R, Y186C and I543V (Arg771-Gly767)] and Monomer B: [Y186C (Ser817-Ala549) and all mutant systems (R605-Gly767). Loss of inter-residue interaction were seen in: Monomer A [all mutant systems except I543V(Gly767-Cys816); in all mutant systems (Leu813-Thr793) and in all mutant systems except (Cys816-

Thr993)] Clearly, a compensatory mechanism of inter-residue communication was depicted to maintain protein function that was unique to each system per monomer. This concept has been reported in other studies (59, 285, 286).

Equally important is the 5-FU chemotherapeutic drug, which binds in domain IV of the DPD protein and is the final recipient of electrons from FMN for oxidation to active metabolites (131). Both orthostatic (mutation within the active site) and allosteric (mutation outside the active site) mutations are known to have an impact on substrate binding, either directly or indirectly (59). They operate by altering or reducing the contacts required for substrate binding to the enzyme, which can reduce catalytic efficiency or enzyme fitness (59, 285). Herein, we also investigated the inter-residue interaction of 5-FU binding environment comprising of residues N609, N668, G764, T737 and N736, E611, L612, I613 using contact maps/heat maps and the WT as a reference. These residues increased inter-residue contact was as follows: (Figure 3.13, Figure 3.6, Figure S3.7.4 and Table 3.6.4):

- Monomer A: Thr575, Val583, Asn585, Arg589, Ile590, Ile591, Asn609, Ile610, Glu611, Leu612, Ile613, Ser640, Ile641, Met642, Leu667, Asn668, Leu669, Ser670, Cys671, Pro672, His673, Gly674, Ala683, Cys684, Gly685, Gln686, Leu690, Lys709, Leu710, Thr711, Pro712, Thr735, Asn736, Thr737, Val738, Gly764, Val765, Ser766 of the WT.
- Monomer B: Thr575, Val586, Arg589, Leu608, Asn609, Ile610, Glu611, Leu612, Ile613, Ser614, Ser640, Ile641, Met642, Glu666, Leu667, Asn668, Leu669, Ser670, Cys671, Pro672, Ala683, Cys684, Gly685, Lys709, Leu710, Thr711, Pro712, Thr735, Asn736, Thr737, Val738, Gly764, Val765, Ser766 and Phe935.

DPD active site residues that are crucial in protein function include, FMN binding residue Lys709 and substrate/inhibitor binding residue Asn736 bridged by Ser670, which forms strong hydrogen bonds to the carbonyl of Lys709 and the carboxamide nitrogen of Asn736, respectively (123, 124). The fundamental role of H-bonds in target-drug affinities has received considerable attention (431, 444, 445). Furthermore, H-bond formation drives protein conformation, which optimizes hydrophobic interactions and increases complex molecule binding affinities (444). Therefore, the different global conformation changes due to mutations might have been influenced by the asymmetric H-bond compensatory mechanism across all mutant system. These could have been influenced by mutant-induced increased residual fluctuation in the loop area, which included the active loop region. While catalytic residue Cys671 is supported by hydrogen forming residue His677 and Glu673, other

residues such as Met642, Ile613 and T575 form hydrogen bonds near the active site. Catalytic reaction ensues when Cys671 (located in the active loop area) come in contact with other 5-FU binding residues (125). Major loss and compensatory gain in inter-residue connection were observed around catalytic residue Cys671 in I543V and D949V (Monomer A) and M166V, S534N and D949V (Monomer B). The ED analysis of these systems also exhibited the most diverse conformation differences. These systems also exhibited increased H-bond instability in the ligand binding site, particularly S534N and D949V (Monomer A) and I543V and M166V. (Monomer B). This was characterized by observed increased COM distances (Figure 3.7). and a loss of *CC* hubs when compared to the WT of their respective monomers (Figure 3.10.2). Same tendencies have been reported by a recent study (285). Although D949Vand Y186C showed a compensatory gain of the most *EC* hubs, a total loss *EC* hub was observed in the catalytic site domain mutations (S534N and I543V), as well as Monomer B of M166V in our study (Figure 3.11.2). Interestingly, some of the 5-FU-protein interacting residues were identified as high *DC* compensatory communicating hubs, such as Arg589, which was found in Monomer A M166V, Y186C, S534N, and I543V but not Monomer B. On the other hand, Lys710 was identified in Y186C and I543V of Monomer A, and C29R in M166V and Y186C of Monomer B. Similarly, Val765 in C29R and Y186 of Monomer A and B, respectively (Figure 3.9.2). Despite the rarity of I543V research, two clinical reports have linked I543V to 5-FU toxicity, due to partial loss of DPD function (458, 459). Interestingly, the highly conserved residue S534 (436) as well as I543 are positioned directly opposite to each other at the hinge of a short β-hairpin-loop which forms the bottom of the (α/β)8- or TIM-barrel of the active site domain IV. While the S534 side chain is stabilized by the carboxyl-group of D532 (2.6 Å), I543 shows a Van-der-Waals-interaction to the backbone of G569, so that in the WT structure both residues seem to be fixed on their positions. The alteration of both amino acid positions would distort the structural integrity of the β-hairpin-loop that covers the channel of the TIM-barrel from the opposite direction of the active site (436). However, the presence of mutated I543V affected the Van-der-Waals-interaction to the backbone of G569 (Figure S3.10).

Similarly, partial loss of function has also been detected in M166V, D949V and C29R carriers. For instance, a clinical case-control study suggested that M166V may contribute to 5-FU-related toxicities in gastroesophageal and breast cancers, but not in colorectal cancer (425). In M166V, methionine has been replaced by valine. Despite the two residues being hydrophobic, the unbranched methionine side chain is stacked deeply and with great flexibility into the inner compartment. Besides, the polarization and redox versatility of its

sulfur atom are distinct features of methionine that cannot be replaced by valine (460). Herein, the previously mentioned global conformation variability was observed in M166V. Additional investigation using the activated WT as a reference revealed that 100% of local inter-residue contacts were preserved in four (Phe163, Alq165, Ser167, and Ile168) residues in both WT and mutation (Figure S3.8.4). Despite the reduced contact in three cases (Thr110, Lue846, and Gln910), there were no significant differences between the WT and mutation.

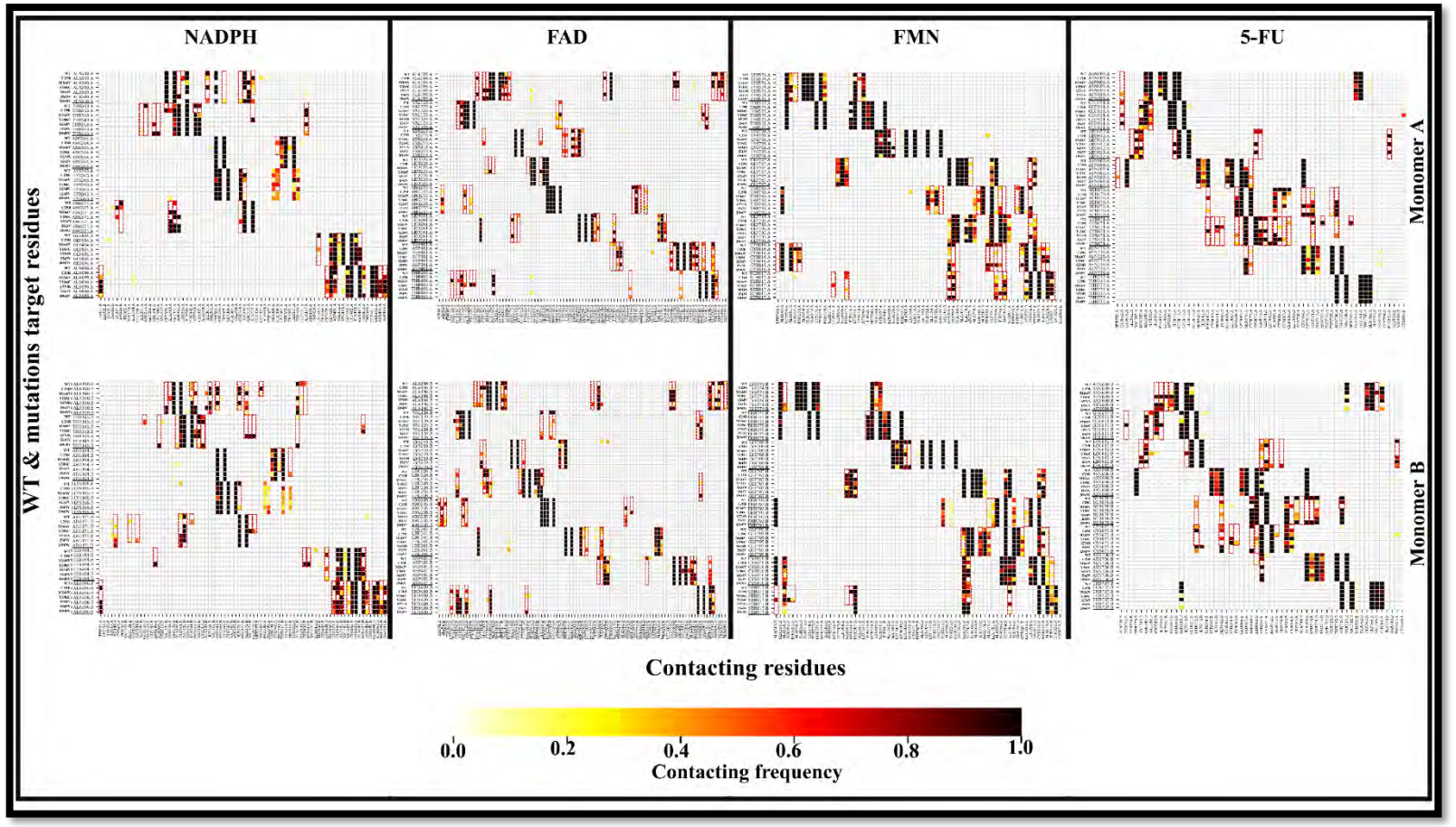


Figure 3.13: Activated DPD protein residue contact heatmap showing the interaction between Monomer A and Monomer B of NADPH, FAD, FMN, and 5-FU cofactors. The shaded boxes in each frame represent significant differences in the frequency of interaction between the cofactor residues and the surrounding contacting residues in the WT versus the mutation. White (0) to dark-red (1.0) represent no-contact (0) to strong-contact (1.0).

Conversely, the mutated residue formed compensatory gain in H-bond with Phe912. In terms of Monomer B residue-residue interaction, four residues (F163, A165, S167, and I168) had 100% interaction in both the mutation and the WT. Mutated residues formed new weak interactions with Ser853 and Pro854, and contacts with Val162 and Leu164 decreased while they increased with Thr110 and Lys846 (Figure S3.8.4), indicating allosteric signaling.

Regarding D949V, as shown in Figure S3.8.5, the higher hydrophobicity (Val) of activated D949V, rather than a charged group (Asp), altered the residue interaction. There was a loss of interactions (Monomer A: R100: 21% and Monomer B: S858:7% & R867: 14%) and compensatory gain in inter-residue contacts (Monomer A: D873: 12%, A946: 1%, & L1002: 98% and Monomer B: V865: 7.2%, T856: 6%, & I998: 29%) was observed when compared to the WT. The D949V mutation was discovered in pediatric patients with complete DPD deficiency (461). It has been proposed that D949D interferes with electron transport or cofactor directly (461, 462). Additionally, treatment with 5-FU resulted in grade 3 toxicity in Caucasian patients with D949V (463).

C29R has been identified in patients with DPD activity that was nearly normal (23). However, a study by Gentile et al., (464) strongly associated C29R to reduced fluorouracil degradation and severe toxicity. Moreso, when compared to the WT using contact map, C29R caused only minor differences in inter-residue contacts. However, total loss of residue-residue contact was observed in Gly498 (Monomer A and B) and Asn494 (Monomer A). A compensatory gain in inter-residue interaction was exhibited in Lys497 of both monomers (Figure. S3.8.6). This could be related to the substitution of a charged cysteine residue for a charged arginine residue. Gly498 and Asn494 are interphase residues that have been associated with protein stability.

The heat map showed changes inter-residue contact relationship at 5-FU binding site in presence of mutations. To support our findings, we visualized 2D representation of protein-5-FU (ligand) interactions in both monomers using LigPlot+ (398) (Figure S3.9). We extracted the drug bound structures from the lowest energy wells before the drug exited and visualized the substrate interactions. It is well acknowledged that biomolecular such as ligands recognition is influenced by weak, noncovalent chemical interactions such as hydrogen bonds (465-467). Apart from Monomer A of S534N and Monomer B of M166V and S534N, the rest of the system retained 5-FU during simulation. Monomer B M166V lost hydrogen bond interactions with Asn609, Asn668, and Ser670, whereas Monomer A lost interaction with Ser670. Although both monomers lost H-bond interaction with Ser670 in S534N, Monomer A and Monomer B lost H-bond interactions with Asn609 and Asn668

respectively. Furthermore, C29R, I543V, Y186C, and D949 lost hydrogen interactions with Ser670 in Monomer A and Monomer B. Ser670 is responsible for protein stability within the active site by forming strong H-bond with Asn736 (5-FU binding residue) and Lys709 (FMN binding residue). The loss of hydrogen bonds in these key binding residues including Ser670, particularly in S534N (Monomer A) and I543V and M166V (Monomer B) may have hampered 5-FU stability resulting in drug exit. Hence, dysfunctional degradation observed in mutation carrier patients (158, 160, 168, 428).

3.3.5 The importance of hub maintenance as a predictive tool for allosteric hot spots and cold spots

Previously we pinpointed key hubs that were maintained throughout the simulation using five different heatmaps of each metric (Figures 3.12.1, 3.12.2, 3.12.3, 3.12.4 & 3.12.5). Subsequently, we identified mutations' allosteric hot spot and cold spot regions. In both the inactivated (drug free) and activated (drug bound) systems, mutation allosteric hot spots were defined as regions that were prone to being occupied by non-conserved hubs. Persistent hub areas, on the other hand, were least affected by all mutations and thus designated as mutation allosteric cold spots (468). Surprisingly, point mutations influenced residues not only in their immediate surroundings, but also at the cofactor (NADPH, FAD, FMN) and active site (5-FU). Change in protein-NADPH, protein-FAD, protein-FMN and protein-5-FU on Monomer A in comparison to Monomer B may be an indication of allosteric signaling (59, 285). We proposed some allosteric hot spot regions around co-factors using synchronized information from penta-metric and weighed contact heat maps. This included (Figure 3.8.1, 3.9.1, 3.10.1, 3.11.1, 3.12.1 and 3.13):

- NADPH: Ala344, Ile347, Gly480 and Val482.
- FAD: Lys497, Ala198, Val 482, Val486
- FMN: Asn544 and Leu813
- 5-FU: Arg589, Lys710 and Val765

It should be noted that these areas were deemed least favorable for drug design because they were most disrupted by mutations. Cold spot areas, in contrast, which were predominantly centrally located within the protein core beginning from the first $Fe^{2+}{}_4S^{2-}{}_4$ cluster towards the active site (Figure 3.14), were identified as better targets for drug design. This is in line with previous studies (468). Contact analysis for the persistent hubs discovered that the high

centrality noted at the protein center core consisting of $Fe^{2+}_4S^{2-}_4$ cluster regions was due to increased interaction between the $Fe^{2+}_4S^{2-}_4$ clusters environment with the hub and neighboring residues. Notwithstanding, we discovered four key hubs (Asn108, Leu155, Met599, and Val778) that were not only positioned in cold spot regions but also had the greatest centrality values when the WT and mutation were accounted for (Figure 3.15, S3.10.1, S3.10.2 and S3.10.3). Furthermore, the importance of each these hubs was explained using penta-metric centrality analysis (Section 3.3.3).

In the pursuit of novel or repurposed/repositioned allosteric drug modulators, researchers are focusing their attention on "mutation cold spots" (59, 283). These regions, characterized by minimal mutation impact and the presence of resilient hubs, offer a promising avenue for exploration of drug binding site (283, 286, 469). Small molecule drugs with the potential to restore activity in hypomorphic mutants are currently garnering attention as a promising aspect of precision medicine. Hypomorphic mutants are genetic variations that lead to partial loss of gene function or reduced expression, impacting protein activity. Notably, there have been successful examples of small molecules restoring the wild-type conformation of p53, effectively rebuilding its proper function (470). Promising agents such as PRIMA-1, MIRA-1, and various derivatives of the thiosemicarbazone family have shown potential in this regard (470). Furthermore, A recent study conducted by Tang et al., (471) focused on SMAD4, a key player in the TGF-β pathway that is frequently mutated in various tumors. The researchers developed a TR-FRET technology that accurately replicated the dynamic and differential interaction between SMAD4 and SMAD4R361H with SMAD3. Through this technology, they identified a bisindolylmaleimide derivative called Ro-31-8220, which showed the ability to induce the interaction between SMAD4R361H and SMAD3. Remarkably, Ro-31-8220 was found to reactivate the dormant transcriptional activity mediated by SMAD4R361H and restore the tumor-suppressing effects induced by TGF-β in cancer cells with SMAD4 mutations. This discovery highlights Ro-31-8220 as an example of a SMAD4R361H/SMAD3 interaction inducer and demonstrates a general strategy to reverse the functional loss observed in tumor suppressors carrying hypomorphic mutations. Furthermore, this finding supports a systematic approach to developing small-molecule, which can offer valuable insights into biological processes and contribute to the discovery of therapeutic interventions.

.

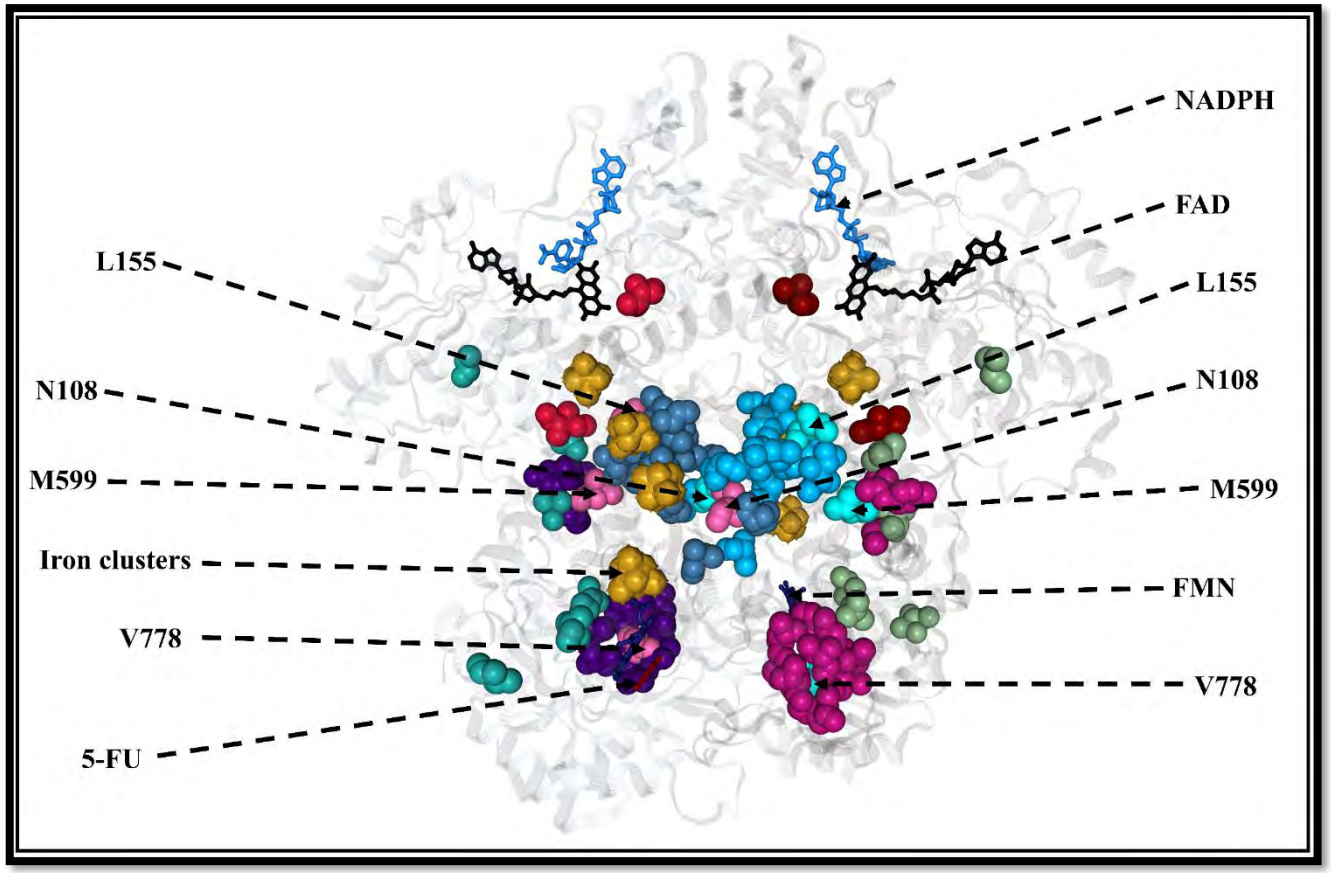


Figure 3.14: 3D representation of collective persistent hubs from Monomer A (*BC*, red; *CC*, teal; *DC*, blue-green; *KC*, Indigo) and Monomer B (*BC*, maroon; *CC*, light-blue; *DC*, green; *KC* violet) of the DPD enzyme. Iron clusters are shown in gold, while super-persistent hubs and hubs with very high centrality are shown in light pink.

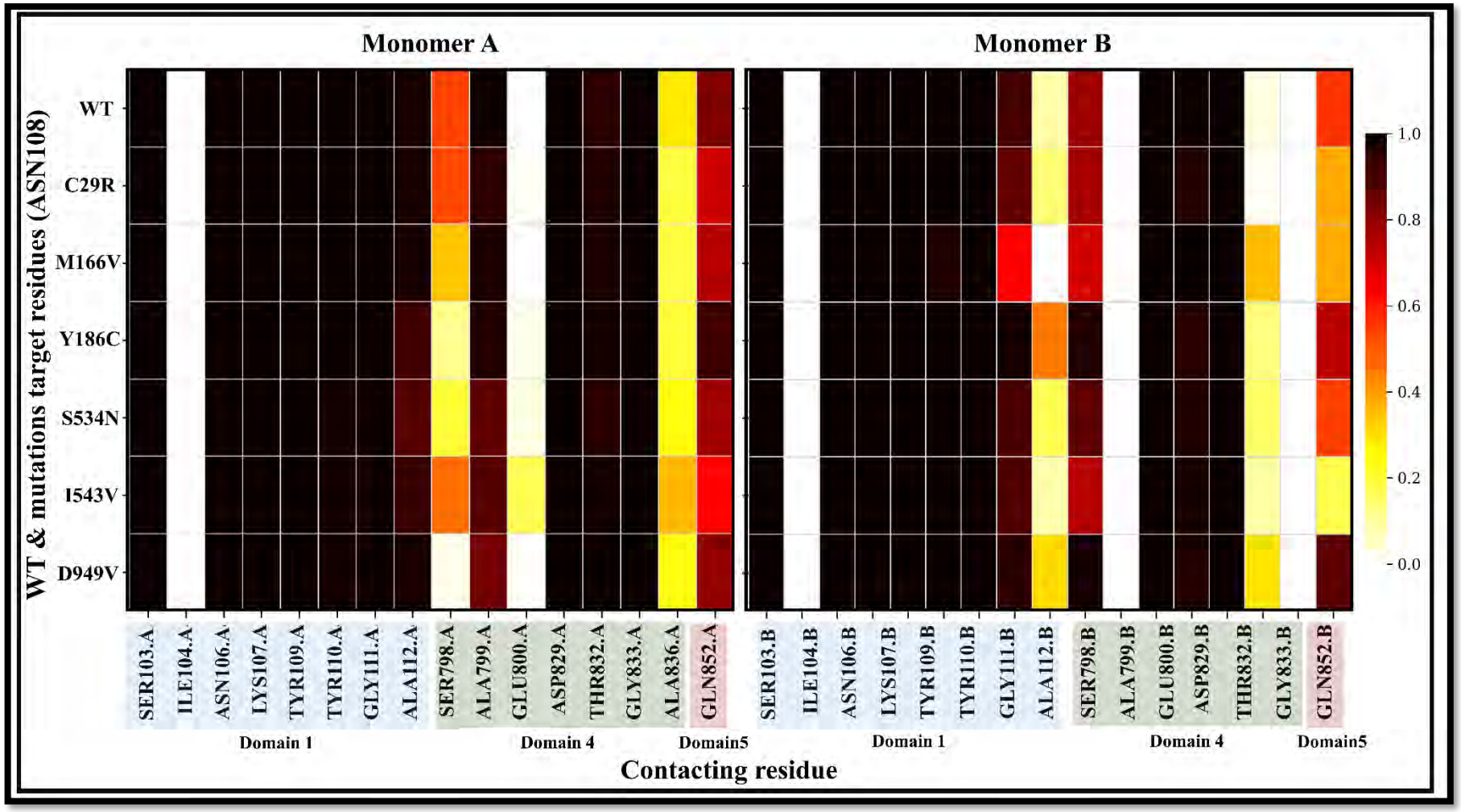


Figure 3.15: Contact heatmap representing contacting residue to high significance hub L108 which interlinks residues from $Fe^{2+}_4S^{2-}_4$ clusters domain I & domain V and FMN-binding domain IV.

3.4 Conclusion

In this study, we performed a comprehensive analysis to distinguish structural changes caused by missense mutations by combining atomistic simulations with post-MD trajectory analysis. In addition to the standard trajectory analysis approaches (RMSD, RMSF, Rg, H-bond), we employed a few of our recently developed methods, such as comparative ED (284) and combined DRN metric analysis in these investigations (266, 283, 284, 287). We also backed our findings with LigPlot+ results. Two of the clinically significant mutations in these investigations (S534N and I543V) are in the active site domain, while the majority (C29R, M166V, Y186C, and D949V) are found in the other four domains. As a result, we discovered mutations mediated local impacts at the point of residues, as well as long-range allosteric implications, which may explain 5-FU cancer drug dysfunctional metabolism by DPD mutated protein.

The following are the specific major findings from this investigation: (1) The RMSD, all versus all, RMSF, DCC, and comparative ED analyses indicated that the mutated proteins had noteworthy active loop flexibility within the 5-FU binding environment and accentuated asymmetric monomer behavior relative to the native protein. A recent cryo-EM article (12) found similar results for two other resistance mutations. Previous investigations on KATG resistant mutations (374) reported similar findings. Nevertheless, additional studies discovered profound (asymmetric) changes in the 5-FU binding environment in comparison to the WT. Consequently, we deduced that, irrespective of the location of these mutations, they affect DPD sensitivity to 5-FU through changes in the active site binding environment. (2) RMSF results revealed that the protein core of both monomers, mainly made of iron clusters, remained stable in all systems while enhanced flexibility was seen at the loop regions. Metal ions within metalloproteins have been found to play a fundamental role in many biological redox processes and can help the protein maintain structural stability (147, 472). As a result of the absence of structural changes caused by mutations, our findings support the concept of iron clusters providing structural stability to the DPD active site. Moreover, the collective global top 4% persistent hubs across all averaged metrics revealed intriguing findings. Mapping these persistent hub residues found that the majority of these (persistent) hubs are in the protein's core (comprising of iron clusters), denoting their importance in DPD functionality. (3) Furthermore, we examined the 5-FU contact residues per monomer during the MD simulations centered on

atomic contact frequency. In the presence of mutations, there was a loss and/or gain of interactions in one or both monomers relative to the WT. Catalytic residues Cys671 and Ser670, which are required for catalysis within the DPD active site, showed differences in interactions, including contact loss in some instances, in the various monomers of mutated proteins. (4) Regarding DRN analysis, we discovered that 5-FU binding activates communication pathways centered on the oxidoreduction electron transport chain, which includes all cofactors (NADPH, FAD, $Fe^{2+}{}_4S^{2-}{}_4$ clusters, FMN) and 5-FU. Using the WT as a reference, clearly different changes in inter-residue network patterns around different mutations' cofactors were observed. The *BC* hubs showed mutation-induced large hub ensembles forming two distinct proximal inter- and intra-monomer allosteric communication pathways from NADPH to the active site. Similarly, a pair of allosteric communication pathways from NADPH to the active site were depicted as clustered *DC* and *BC* hubs. Both the *BC* and *DC* pathways began from NADPH to FAD of Monomer A and B to the protein core ($Fe^{2+}{}_4S^{2-}{}_4$ clusters), then finally to the active site (FMN and 5-FU). This was referred to as the oxidoreduction pathway. In comparison to the WT, some of these hubs were lost while others were gained. This could indicate a compensatory allosteric path to the active site caused by mutation, which might destabilize the active-loop and 5-FU binding environment. It was fascinating to see how one pathway dominated the other in the number of hubs, revealing the asymmetric behavior of the monomers. This was in contrast to our previous findings, in which the allosteric communication network formed by *BC* (287) or *EC* (283) to the active site was clearly defined as a single communication pathway. However, one of these studies (374) revealed a large hub ensemble communication pattern, which was similar to what we observed in our study. As a result, none of the six mutations resulted in total loss of function (164, 168, 173, 176). D949V has been classified as an intermediate metabolizer by the Clinical Pharmacogenetics Implementation Consortium (CPIC) (164). We discovered that the D949V mutation caused significant inter-residue crosslinking around the Monomer A, FAD, and active site regions of both monomers based on *CC, EC*, and *KC* results. Furthermore, using newly introduced and retained residues, a novel allosteric communication pathway from NADPH to the active site region was discovered. We proposed that this is one of the most distinguishing features, resulting in a 50% loss of D949V function (164). However, due to the lack of clinical evidence presented by the other variants (C29R, M166V, Y186C, S534N, and I543V), no dose has been established. (5) The *EC* hubs were profoundly located in the active site in the WT, but

all mutated proteins appear to have lost this importance. *KC*, on the other hand, highlighted the significance of co-factors as well as the active site. (6) In addition, we identified the following noteworthy residues: L108 was found in the interface of domain I and domain V; L155 is a super-persistent hub in domain I's α-helix that formed a unique bond between Q156 and the Fe^{2+} ions (147). Non-conservative hubs, in contrast, were discovered in mutation hot spots, primarily around cofactor regions, whereas the persistent hubs were mostly found in $Fe^{2+}{}_4S^{2-}{}_4$ clusters.

As a result, we advocate a search for new or repurposed/repositioned allosteric drug modulators that target persistent hubs in areas least affected by mutations (dubbed "mutation cold spots"). This will aid in the restoration of proper 5-FU metabolism and the prevention of toxicity. Overall, we discovered that DPD missense mutations are strictly regulated by a dimeric asymmetric intra- and inter-monomer allosteric network pathway. Despite the paucity of clinical data for C29R, M166V, Y186C, I543V, and S534N, our analysis revealed that these mutations impair function. More clinical research on missense mutations in people of African ancestry should be conducted, and an appropriate treatment protocol, including dose recommendations, should be implemented. This is an important consideration in underserved African precision oncology medicine.

CHAPTER 4
THE IMPLICATION OF VARIANTS OF UNCERTAIN SIGNIFICANCE DPD MUTATION ASSOCIATED WITH AFRICAN DESCENT POPULATION IN DYSFUNCTION METABOLISM OF DPD

4.0 General Background

As the body of genomic and pharmacologic data on cancer grows exponentially, data analysis is the main stumbling block in translating such information into biologically significant and clinically relevant alterations (173, 473, 474). Although various approaches have been implemented lately in the analysis of such datasets (475, 476), the majority of them, notably those dealing with mutation data (476, 477), use a protein-centric perspective (emphasis on changes at the whole-protein level), failing to account for the given position of the different mutations within a protein (478). These methodologies have proven useful in a variety of applications; regrettably, they are incapable of dealing with situations in which different mutations in the same protein have different effects depending on which region of the protein is altered (479). This notion is easily clarified by the fact that often proteins are complex, comprising of several distinct domains and/or functional regions (478). Protein functional regions (PFR) are defined as complex multi-domain protein areas that provide cryptic but vital cellular signaling functions, resulting in intricate temporal and signaling molecules as well as their signaling network (480).

For instance, the dimeric DPD protein has five domains, all of which are involved in the electron transport chain, resulting in the degradation of 5-FU at the active site. Mutations can affect any of these domains, including N-terminal $Fe^{2+}{}_4S^{2-}{}_4$ cluster domain I, FAD-binding domain II, NADPH-binding domain III, FMN/pyrimidine-binding domain IV (the active site) and C-terminal Fe–S domain V (129, 130). As previously described (147, 450), there is a two-site ping-pong mechanism in which NADPH binds and reduces FAD, then, through the Fe-S clusters, the reducing equivalents are transferred to the FMN at the other end of the protein, as suggested by steady-state DPD analysis. This indicates DPD is active only as a dimer. Dihydropyrimidine and oxidized FMN are formed when reduced FMN reacts with pyrimidine (129, 130, 450). Therefore, any mutation-induced changes at these PFRs could alter DPD

structure and function. This was well illustrated by the findings of our earlier studies of clinically relevant missense mutations, two of which were located in the active site domain (S534N and I543V) and the rest in the remaining non-active site domains (C29R, M166V, Y186C, and D919V). As such, we proposed that, beyond revealing the general effect of loss of function in a whole protein due to existence of missense mutations, analyses can be narrowed down to the protein domain level.

Our subsequent studies aimed to predict the effects of region-drug association of missense mutations located in either: the non-active site (Chapter 4) or active site (Chapter 5) domains using available *in vitro* data (non-clinical data). This was based on atomic studies on clinical data from DPD pathological mutations, which revealed a distinct variation in 5-FU metabolism. In each section, missense mutations were identified in either the non-active site (remote) or catalytic domain(s) based on their location on the DPD protein.

4.1 Introduction

Allostery is recognized as a crucial regulatory system of protein activity given the dynamic nature of protein structures (433, 481). Its regulation can be influenced by the binding of effective ligands, mutations, or covalent modifications (referred to as allosteric effectors), which perturb sites distant from the signal initiation site (59, 287, 433). Hence, allostery has been defined as a biological concept that refers to any sort of long-distance communication ("allosteric signaling") in proteins that occurs as a result of perturbations at one site ("allosteric site") causing structural and/or dynamic changes at another site ("active site/orthosteric site") on a protein (59, 69, 269, 420, 422). It is now clear that mutational perturbations consistently affect the packing and dynamics of a significant fraction of protein residues, including those located more than 10-20 Å from the mutated site (482). Through these mechanisms (allosteric modulation of function), the protein's stability and native conformation in its functional regions, such as the active site, are significantly altered (59, 287). Historically, allosteric regulation was associated with quaternary structures (for example, hemoglobin), but it is now accepted as an intrinsic property of all proteins, including monomeric structures (483, 484). Therefore, the following chapter describes the allosteric mechanism of DPD remote missense mutations associated with African ancestry populations.

Research manuscript draft:

INSIGHT INTO THE MECHANISM OF ACTION OF REMOTE MUTATION-INDUCED ALLOSTERIC EFFECT ON DIHYDROPYRIMIDINE DEHYDROGENASE PROTEIN STRUCTURE AND FUNCTIONALITY, WHICH CAUSES THE DYSFUNCTIONAL METABOLISM OF 5-FLUOROURACIL ANTI-CANCER DRUG

This section discusses remote mutations obtained from non-clinically significant data (*in vitro* and *ex vivo* studies). Included in this chapter is the concept of allosteric communication pattern influenced by the presence of missense mutations highlighted through a hybrid of computational approaches.

The first draft manuscript for this chapter is available and is awaiting approval from all parties involved before being submitted for publication.

Contribution: This study was conducted under the supervision of Özlem Tastan Bishop and Thommas Mutemi Musyoka. **Maureen Bilinga Tendwa** performed all global and local calculations, including the DRN calculations. However, the script for the identification of global top 4% high significant multi-metric (*BC, CC, DC, EC* and *KC*) DRN residues was generated by Olivier Sheikh Amamuddy (for another study); and DRN hub calculations were performed by Victor Barozi in this study. The paper draft was written by Maureen Bilinga Tendwa.

4.2 Material and methods

It is worth noting that this chapter used similar investigative approaches to those applied in clinical data (**Chapter 3.2).** However, the following remote mutations were found in this case: S201R, K259E, D342N, D432N, S492L, R886C, and L993R. To recap, prior to simulation, the previously generated metal ion parameters for DPD protein (Chapter 2) were considered. Thus, we inferred these AMBER Fe^{2+} force fields into our remote mutated and protonated model structures and ran a 600 ns MD simulation. Regardless, the effects of the mutations at the global and residue levels were investigated using post-MD analysis using the aforementioned combined approaches.

4.3 Result and discussion

4.3.1 DPD remote missense mutations associated with the African ancestry population identified through *in vitro* and *ex vivo* study

Data on DPD pathological mutations associated with the African lineage subpopulation that had never been studied clinically before was obtained from literature (149, 162, 165, 168, 176, 450, 451, 485). Missense mutations in this population were then identified using the following web servers: the NHLBI Exome Variant Server based on Minor Allele Frequencies (ESP MAF) (486), 1000 Genomes Project (487) (Table S4.1), high coverage African ADME Dataset-HAAD (488), Database of Single Nucleotide Polymorphisms (dbSNP) (489), ensemble (434) and UniProt (490). Thereafter, seven non-active site mutations (S201R, K259E, D342N, D432N, S492L, R886C, and L993R) of interest were filtered using combined *in vitro* data from PharmKGB (491) and ClinVar (492), as well as consensus from *in silico* pathogenicity prediction results from Variant Analysis PORtal (VAPOR) (219) which incorporates: Proven (493), Polyphen-2 (494), PhD-SNP (495), and I-Mutant (217). Similarly, both Ensembl Variant Effect Predictor (EVE) and Variant Annotation Integrator (VARITY), as sequence-based variant effect prediction programs, concurred in their predictions. They classified the mutations S201R, K259E, D342N, D432N, S492L, R886C, and L993R) to be likely pathogenic. (Figure S4.1.1 and S4.1.2).

To visualize their relative locations, the identified mutations were mapped onto a previously modeled DPD human three-dimensional structure (Figure 4.1). We determined the distances between each mutation and the nearest co-factor or active site. This was based on the fact that each DPD domain was a critical protein functional region containing a cofactor essential in electron transport (oxidoreduction chain) from NADPH to the active site. Three mutations (K259E, S201R, and S492L) were identified in the FAD binding domain II, with K259E positioned furthest from the active site (Monomer A: 76.4 Å and Monomer B: 78.6 Å). S492L (Monomer A: 13.3 Å and Monomer B: 14.8 Å) was located closer to FAD than S201R (Monomer A: 17.7 Å and Monomer B: 15.2 Å) and K259E (Monomer A: 15.5 Å and Monomer B: 14.4 Å). On the other hand, D342N and D432N were located in the NADPH-binding domain III, although, D432N was positioned slightly closer (COM distances: Monomer A;72.7 Å and Monomer B; 63.7 Å) to the active site than D342N (COM distances: Monomer A; 78.3 Å and Monomer B; 65.4 Å). Equally, two mutations (R886C and L993R) were seen on the C-terminal

$Fe^{2+}{}_4S^{2-}{}_4$ cluster binding domain V, where R886C (Monomer A: 47.8 and Monomer B: 48.3) was closer to the metal cluster than L993R (Monomer A: 19.5 and Monomer B: 20.0) (Table S4.2). Despite being located away from the active site, previous research has shown that such mutations can affect protein structure and function via the phenomenon of allosteric communication.(59, 287, 420, 422). Therefore, to elucidate the mutation-induced allosteric effects on DPD structure and function, the subsequent analysis was conducted in two parts. The first part was centered on global analysis which determined the overall structural changes in DPD, whereas the second part focused on local analysis to determine the effects of mutations at residue level.

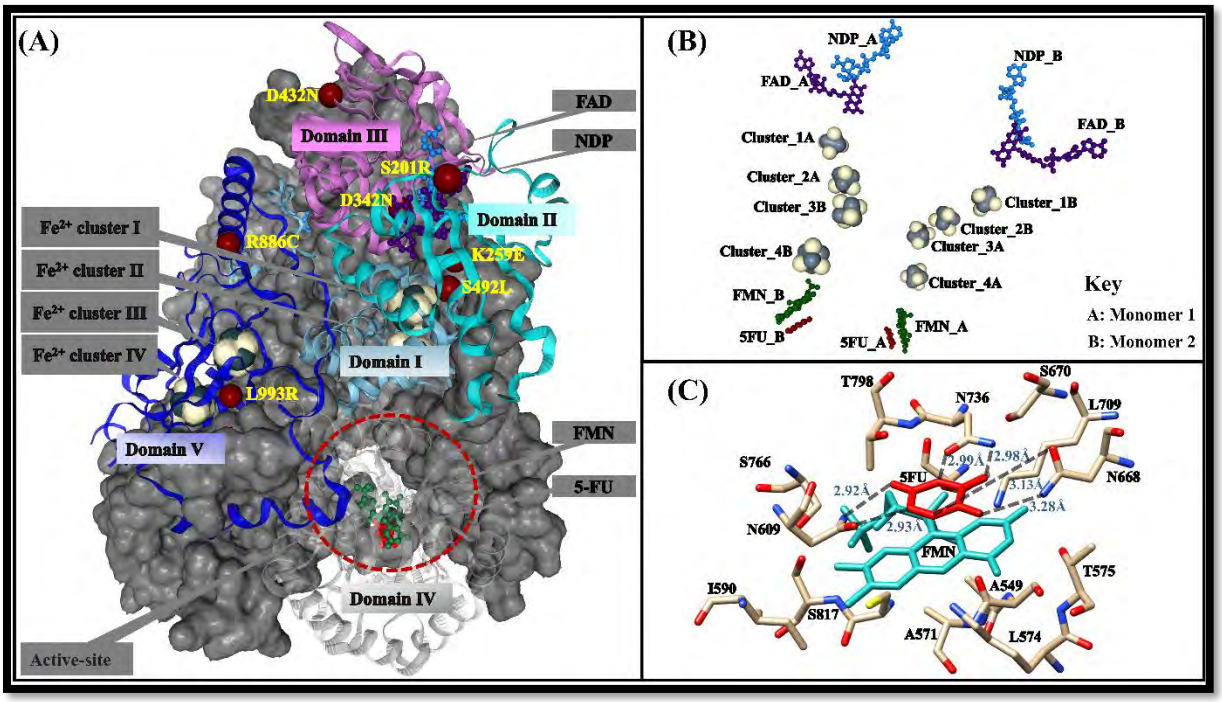


Figure 4.1: A detailed structural description of DPD protein. A) 3D Mapping of the seven non-clinical missense mutations (S201R, K259E, D342N, D432N, S492L, R886C, and L993R) onto modeled human DPD structure. Monomer A is denoted as a grey surface, whereas Monomer B is illustrated as a ribbon, with domains I-V represented in sky blue, cyan, sea-green, purple, cream and royal blue, respectively. The red spheres represent various mutations. The clusters of $Fe^{2+}{}_4S^{2-}{}_4$ are shown in ivory and teal, NDP in blue, FAD in green, FMN in indigo, and 5-FU in red. **B)** The visual representation of oxidoreduction channel. **C)** The active site surrounding and interacting residues.

PART I

4.3.2 Global Analysis

From the previously homology modeled human DPD structure (147), we utilized Discovery Studio 19 (304) to create a 3D model of each inactivated (drug free) and activated (drug bound) mutation. All ensembles were created by mutating each specified residue as desired. Following validation with PROCHECK (209), ERRAT (496), and ProSA (210) (Table S4.3), the protein was found to be native-like in all drug free and drug bound models. Whilst the founded models were of high quality and suitable for subsequent MD simulation, they did contain a significant number of titratable residues for which conventional MD could not indicate their protonation states (266). As a result, before simulation, the modeled drug free and drug bound ensembles were protonated at pH 7.5 of the DPD active dimeric state (130) using H++ web server (390). Deprotonation/protonation of active site atoms, bound substrate, and/or cofactors can affect structural stability, cause conformational changes, as well as decrease or increase enzyme activity (266). As a result, it has been acknowledged that characterizing protein structures, notably accurately identifying proton locations, is fundamental in chemically understanding enzyme catalytic processes (147). MD simulation can provide more information about such protein dynamics (279). As a result, we ran 600 ns MD simulations on eight DPD protein systems (WT and seven mutations) in their drug free and drug bound states (398). This was based on the observation by Lohkamp and colleagues (129, 249) that DPD protein exists in two states: drug free (protein inactive state) and drug bound (protein active state). Therefore, the framewise RMSD analysis, root mean square of deviation (RMSD), radius of gyration (Rg), dynamic cross-correlation (DCC), root mean square of fluctuation (RMSF), number of intermolecular hydrogen bonds (H-bonds), and comparative essential dynamics (ED) incorporating principal component analysis (PCA) then provided detailed structural and protein dynamic information in global analysis. These assessments were performed on each monomer of the drug free and drug bound systems. Line graphs, violin plots, heatmaps, as well as density maps were used to visualize the information.

4.3.2.1 C-α RMSD demonstrated protein conformational changes caused by mutations

In our previous study (Chapter 3: Clinically relevant Data), the remote mutations C29R, M166V, Y186C, and D949V revealed that the WT and most mutant proteins frequented multimodal conformations and displayed asymmetric protein behavior. Therefore, we sought to determine the effect of remote mutation on protein structure and function using non-clinical data. According to offer et al., 2014 (174), among the potentially harmful missense variants, S201R were among those mutations with little to no residual enzymatic activity (12%). A significant decline in activity between 12.5-25% of the native protein were observed for R592Q, D342N and S492L. K259E has been predicted to be delirious (497), and the substitution of a basic to acidic amino acid (K259E) results in little change in the proteins secondary structures (498). Consensus from *in silico* pathogenicity prediction results from Variant Analysis PORtal (VAPOR) results classified D432N, R886C, and L993R mutations as dangerous (219).

To see how these mutations impacted on protein structure and function, we used C-α RMSD to assess the global conformation differences between the mutated proteins and the WT structure. The information was conveyed using various types of RMSD analysis. Initially, the most commonly used RMSD against timeline plots were computed, showing the divergence of a given structure with time relative to its initial conformation (Figure 4.2.A). The mutated and WT proteins deviated from their starting structure in different patterns until 350 ns, with backbone RMSD ranging from 1.0 Å to 4.3 Å in both drug free and drug bound systems. Certain mutations exhibited extreme variability patterns when compared to the WT after 350 ns and continued to do so until the simulation ended. In their drug bound state, R886C (4.3 Å) and L993R (2.7 Å) of Monomer A differed the most from the native (3.2 Å), whilst S201R (2.7 Å) and L993R (3.0 Å) of Monomer B varied the most from the native (3.4 Å). However, the drug bound systems, L993R (2.8 Å) and S201R (2.8 Å) of Monomer A, as well as L993R (2.8 Å) and S201R (2.5 Å) of Monomer B, showed noticeable divergence from their respective WT (3.4 Å). The RMSD calculations for each monomer were then represented as violin diagrams (Figure 4.2.B). Apart from S201R in the drug free and drug bound states of both monomers, which equilibrated at a single conformation (unimodal distribution), most ensembles in each monomer equilibrated at multiple conformations (multimodal distribution) and displayed asymmetric behavior. This accords with clinical RMSD data in cases of remote mutation. Moreover, the distribution patterns of the respective monomers of the same system (Monomer A and B: drug free and Monomer A and B: drug bound) differed from one another and also from their drug free and

drug bound states. The drug free WT was equilibrated at two conformation equilibria in both monomers (bimodal distribution). After activation (drug binding), however, both monomers visited multiple conformation equilibrium states (multimodal distribution). S201R, R886C, and L993R had the most noticeable differences from the native. When the RMSD of each frame in the trajectory was compared to the RMSD of all other trajectories, similar trends were observed (Figure 4.2.B). Additionally, NDP, FAD, FMN, and 5-FU RMSDs were not exceptions (Figure 4.2.B and Figure S4.2.2). Both monomers' conformation stability was clearly influenced differently in WT and mutations in both drug free and drug bound states. In each system, conformation drift was observed during the transition from drug-free to drug-bound protein states, with monomeric asymmetric protein behavior. This was attributed to increased flexibility in the protein loop region, including the active loop area; a similar trend was observed in our parameterization paper (147). Other studies have found that the presence of mutations causes a changes in protein conformation variability, which has been linked to, increased protein flexibility in the loop areas. (285, 287). Our clinical data analysis and previous research indicated that mutations caused monomeric asymmetrical behavior (266, 285, 381).

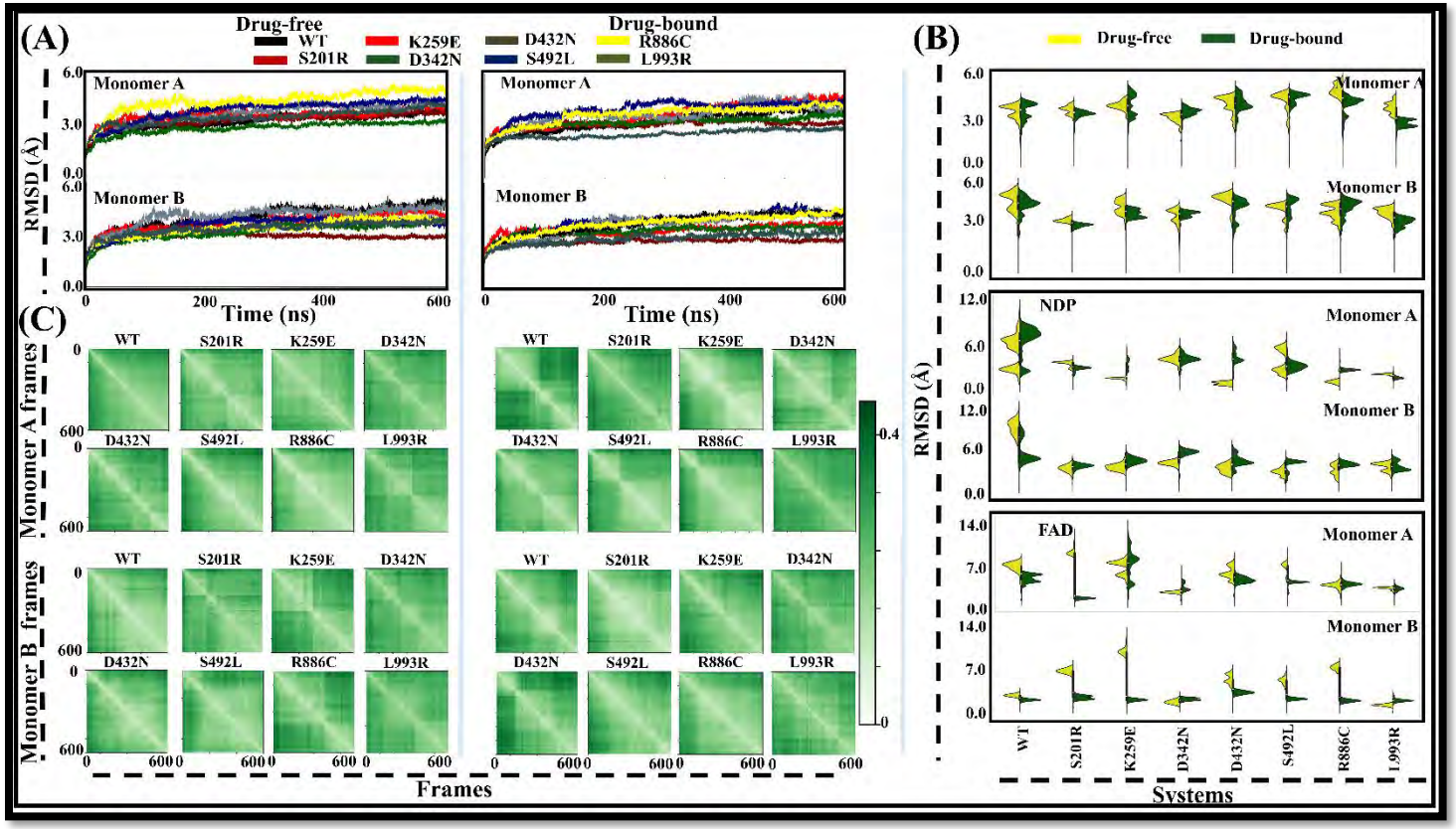


Figure 4.2: RMSD visualization of each monomer (WT and mutated proteins), its co-factors and substrate (5-FU) as they transition from open to closed conformation state. A) Violin representations of Monomer A (yellow) and Monomer B (green) in drug free (open) and drug bound (closed conformation) systems. **B)** Framewise (green) representation of Monomer A and Monomer B in drug free (open conformation) and drug bound (closed) systems. **C)** Violin representations of NDP and FAD drug free and drug bound state of Monomer A (yellow) and Monomer B (green).

4.3.2.2 RMSF and DCCM highlight the asymmetric behavior of DPD monomers in the WT and mutated proteins.

Following that, we utilized RMSF and DCCM analysis to illustrate the asymmetry of monomers in each system and how mutations affect it, as revealed by RMSD (violin plots, line plots, and framewise) results. Figure 4.3.A and B show mobility as a fluctuation comparative to each atom's mean position (RMSF). Each monomer per system, both drug free and drug bound, showcased six main greater mobility regions. The most fluctuating areas were the beta-sheet and consecutive loop areas around residues 40 - 52, 316 - 334, 736 - 771, 901 - 922, 964 - 971, and also the N-terminal (2 - 12) and C-terminal (996 - 1025). (Figure 4.3.A & B). Certain mutations showed more mobility areas than others; noticeable fluctuation occurred in most residues of Monomer A's S201R, D342N, and R886C, along with Monomer B's K259E, D432N, and R886C in their drug free state. Equally, K259E of both monomers had more fluctuating residues in its drug bound than WT. The protein core region ($Fe^{2+}_4S^{2-}_4$ clusters) oscillated the least, suggesting that residues in this environment were comparatively more stable during the simulation process. This is a trend that we have observed in our previous studies (147), as well as our clinical data results. Despite the fact that the mobile regions behaved similarly in both drug free and drug bound systems, the degree of fluctuation in Monomer A and B differed markedly.

DCCM was then applied to ascertain the relationship between the residue on each monomer of the drug free and drug bound proteins (Figure 4.3.C and D). According to the matrices, highly positive regions (colored yellow to red) have strong correlated motions (residue pairs move in the same direction), whereas negative regions (colored blue to green) have strong anti-cross-correlation motions (residue pairs move in opposite directions). Although WT residues showed relatively stronger cross-correlation in the drug free state of both monomers, this lessened in the drug bound. Moreover, across all systems, active site region residues had higher cross-correlation than residues from other domains.

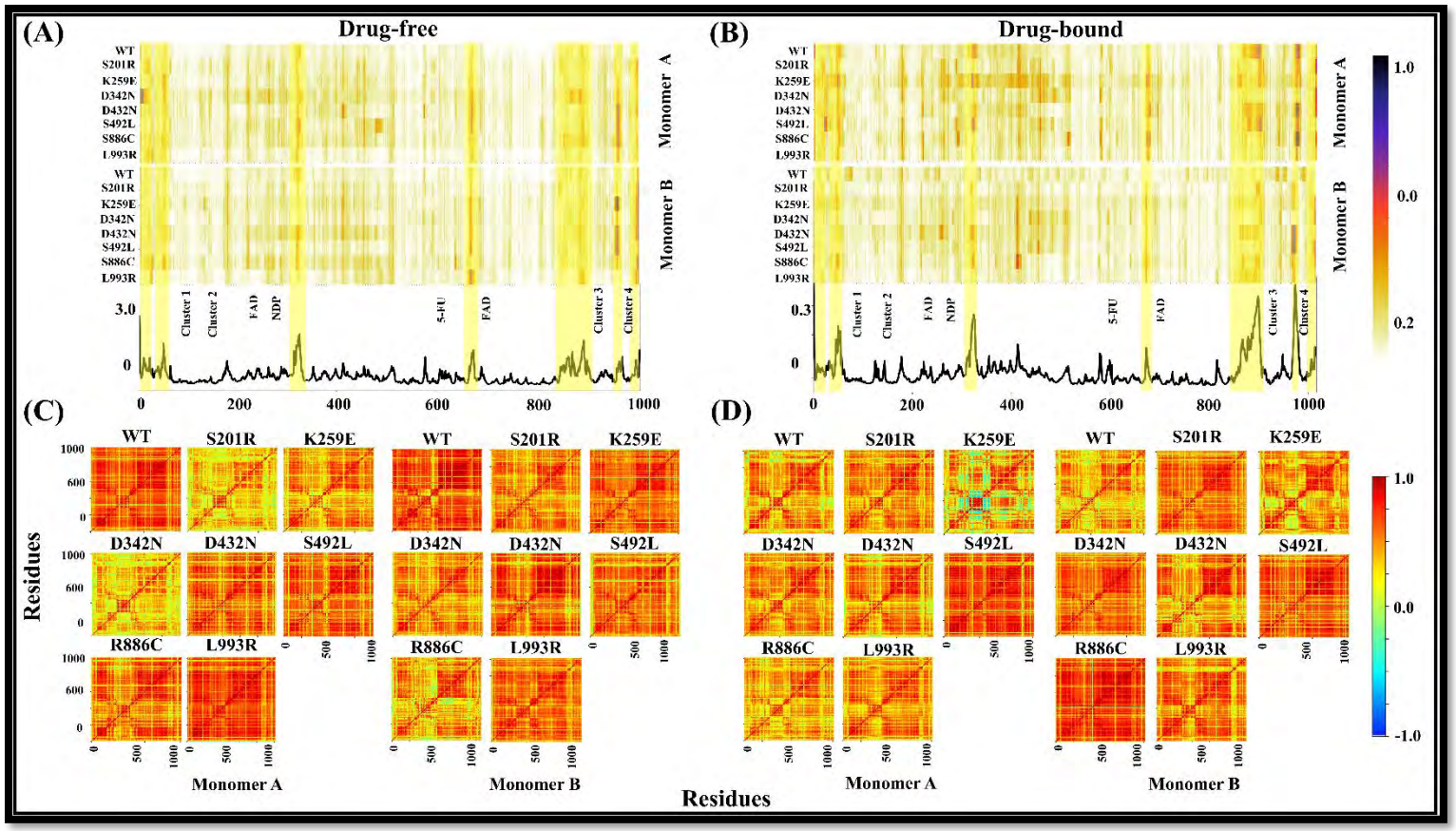


Figure 4.3: A heatmap portrayal of the WT and seven mutants' root means square fluctuation (RMSF) and dynamic cross-correlation matrix (DCCM) after 600 ns simulation. The RMSF is represented as a fluctuation relative to the mean position of each atom, with a highlight of regions with high fluctuation seen in light brown box-frames. The locations of cofactors are also indicated. DCCM colors highly correlated residues yellow to green, while anti-correlated residues are blue to green. A) denotes drug free RMSF proteins, whereas B) denotes drug bound RMSF proteins. C) denotes drug free DCCM proteins, whereas D) denotes drug bound DCCM proteins.

Similarly, increased levels of residue anticorrelation were detected near the mutation site in comparison to the WT configuration However, the correlations between active site residues were lowest in drug free Monomer A D342N and S201R, and in drug bound Monomer B K259E and R886C. In general, the DCCM results agreed with all versus all RMSD analysis, which also revealed that each mutated system displayed asymmetric behavior per monomer that differed noticeably from the WT. As mentioned earlier, this was linked heightened fluctuation at the loop areas including active loop region (147).

4.3.2.3 Comparative essential dynamics revealed asymmetric conformational changes in the presence of mutations

As with the clinically identified mutations, we used ED to analyze mutation-caused conformational changes at the monomer level (Figure 4.4). This was executed because protomer dynamics in a homodimer can switch between identical copies of a protomer (266); which was also supported by the RMSF calculations. (Figure 4.3). Therefore, comparative analysis using MDM-TASK-web tools were made between WT and each mutated protein for each monomer (499). Samples were extracted from the same time frame of an equilibrated 30 ns section of trajectory each monomer on a protein system prior to drug exit. Considering the monomers, the total variance ranged from approximately 52 to 80% and 55 to 64% for drug free and drug bound systems, respectively. S201R, K259E, D342N, D432N, S492L, R886C, and L993R mutated systems sampled a diverse conformational space in comparison to the WT across PC1 and PC2 axes. PC1 repeatedly represented the majority of conformational variance in drug free systems, with the lowest percentage (53.75%) in K259E and the highest (79.26%) in R886C of Monomer B. However, upon drug binding, Monomer A of D432N and R886C had the lowest (55.61%) and highest percentages (63.16%), respectively. A distinctively constricted conformational space in contrast to the WT was noted in S201R and L993R in both monomers and along both PCs in the drug bound state. It might imply that the mutant systems with a more diverse conformation space than the WT (S201R, K259E, D342N, D432N, S492L, R886C, and L993R) sampled more conformations prior to equilibration, which is consistent with both the RMSD and the DCC. More importantly, as in previous analyses, as well as, the clinical data findings asymmetric behavior of both monomers was observed (266, 374).

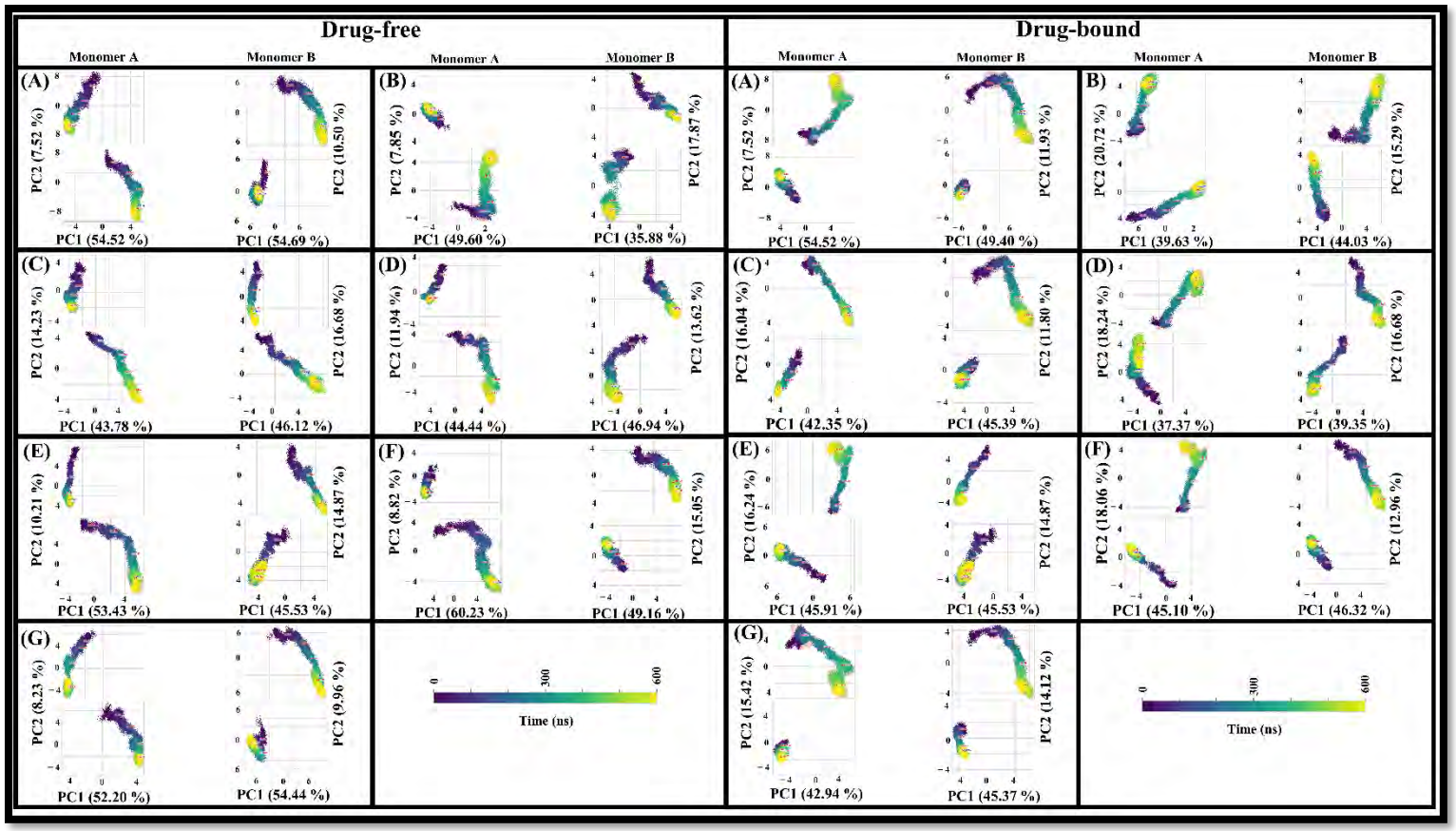


Figure 4.4: Comparative essential dynamics analysis of the mutated proteins in respect to the WT at monomer level for both drug free and drug bound systems. In each subplot, the WT is at the top and the mutation is at the bottom, with Monomer A on the left and Monomer B on the right. The x- and y-axes depict the percentage variance explained by PC1 and PC2, respectively. The plot color code from dark blue to yellow represents the simulation progression in nanoseconds. Letter A) S201R, B) K259E, C) D342N, D) D432N, E) S492L, F) R886C, and G) L993R represent mutations in both drug free and drug bound states.

4.3.2.4 Mutation-related changes in the 5-FU binding environment were revealed by comparative ED, Rg and RMSF

We then focused on the 5-FU binding environment of each monomer per protein. These residues include: N609, N668, G764, T737, and N736 (E611, L612, I613). Also of importance in this region is the active-loop containing residues: 670 – 682. We demonstrated in the preceding sections that these loops become more flexible in the presence of missense mutations.

As seen in Figure 4.5.A and B, comparative ED analyses of Monomer A and Monomer B active site environments in either their drug free or drug bound state showed diverse conformation variability as expected. Monomer A accounted for 49.94% of the widely sampled global conformational subspaces in both drug free and drug bound conformation. This was lower than the one displayed in Monomer B (Total: 53.56%; PC1: 35.91%; PC2: 17.65%). Monomer A of the WT depicted a more widespread longitudinal conformation subspace in its drug free conformation state than the restricted conformation subspace displayed in its drug bound conformation state. In drug free states, convergent conformation subspaces that vary slightly from their drug bound ensembles were observed. Moreover, distributed conformation subspaces were noted in S201R, K259E, D342N, S492L, R886C, and L993R of Monomer A's drug bound, with the exception of D432N.

Though each Monomer B system had different global structural variability in both drug free and drug bound conformation, most drug free systems had more distributed conformation subspaces (WT, S201R, K259E, D342N, D432N, S492L, R886C, and L993R). In the drug bound conformation state of WT, S201R, K259E, D342N, D432N, and L993R, this tendency was maintained. Furthermore, extreme conformation drift from drug free to drug bound state was noted in K259E. Clearly, the drug free WT protein conformation dispersion of Monomer A deviated not only from that of the drug bound WT, but also from that of the mutations in the same group. Monomer B showed a similar pattern. The high fluctuation in the active-loop area most probably triggered the conformation variability in the active site environment, that could further affect the binding of 5-FU to the catalytic site.

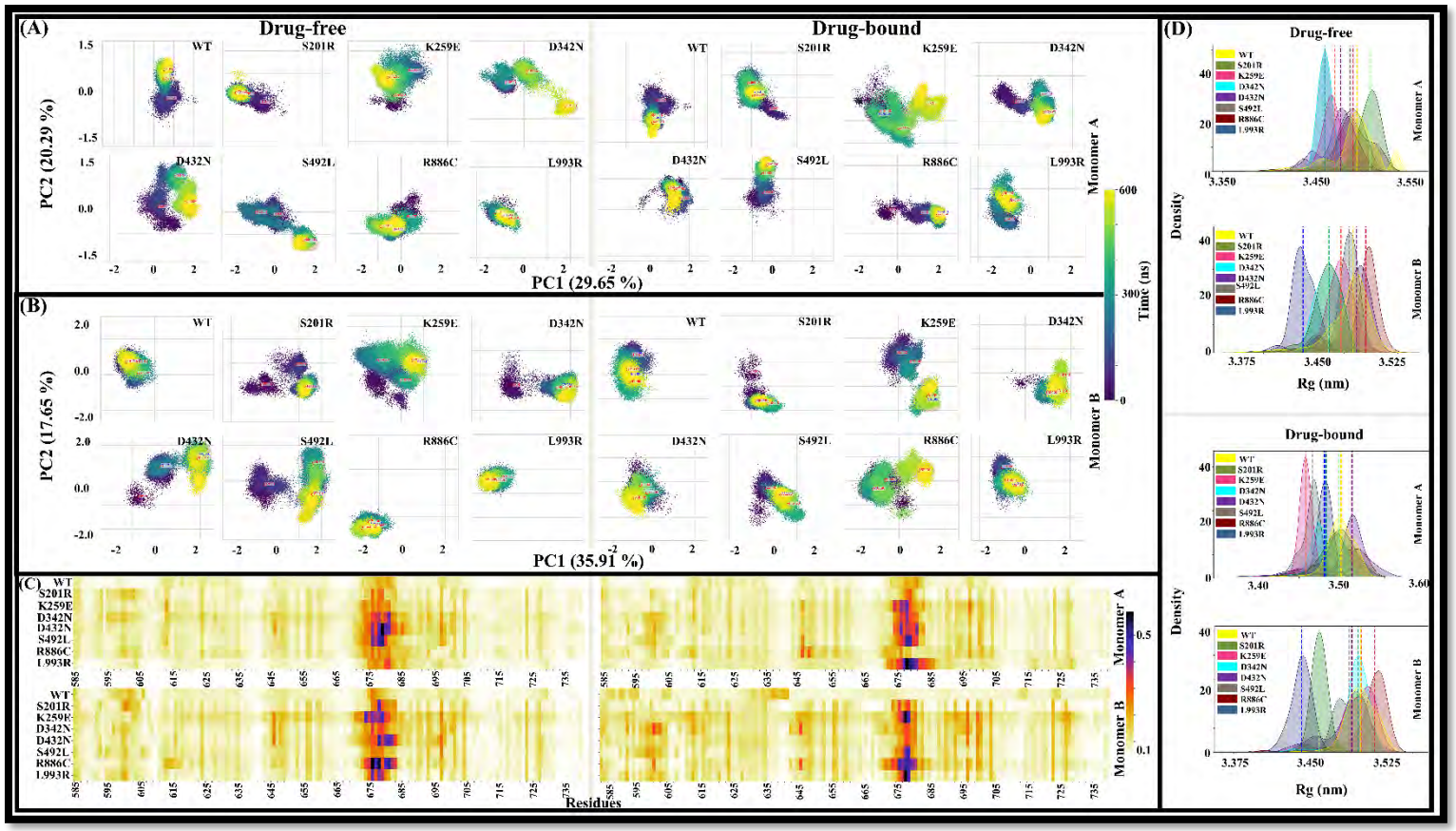


Figure 4.5: Essential dynamics (ED), RMSF, and radius of gyration (Rg) for DPD WT and 7 mutations active site environment. The visualization of comparative essential dynamics (ED) is based on principal component analysis (PCA) of enzyme ensembles. A) drug free and drug bound states of Monomer A. A total of 46.94% (PC1: 29.65% and PC2: 20.29%) global conformation subspaces are sampled. B) Monomer B samples a total of 53.13% (PC1: 33.50% and PC2: 119.63%) global structural variances in both C) drug free and D) drug bound states. The protein's unfolding (yellow) to folding (dark blue) evolution state is depicted over 600 ns. E) Violin plots representing the radius of gyration (Rg) of Monomer A (Monomer A) and Monomer B (Monomer B) in the proteins' drug free and drug bound states. The drug free group is represented as blue (Monomer A) and grey (Monomer B), whereas yellow (Monomer A) and cyan (Monomer B) represent the drug bound group of proteins.

This was evidenced by RMSF analysis of active site residues (Figure 4.5.C), which indicated that the majority of fluctuations happened around the active site loop (residues 670 - 682) of each mutation before and after activation in comparison to their respective WT. Similar observations were made by Barozi and colleagues (374), where higher flexible regions noted in the KATG mutants N-terminal domain were thought to affect the general dynamics of the protomer, as well as the binding pocket kinetics. Further Rg analysis (Figure 4.5.D and Tables S4.4.1 – S4.4.4) demonstrated that the mutations' diverse impact on protein compactness per monomer may be due to varying degrees of flexibility around the active site region. With the exception of Monomer B S492L, mutations showed a greater degree of variation across all systems, and clinical data revealed similar results using both RMSF and Rg analysis.

These findings are consistent with previous research, which discovered that the shapes of numerous enzyme active sites influence the subtle changes in substrate caused by mutations located far from the active site induced mechanism (25, 32). As reported previously by Osuna et al., (93), both catalytic and noncatalytic mutations had a significant impact on a number of factors, including the reorganization of catalytic and other active site residues, dynamic motion, and water penetration, via such allosteric communication.

4.3.2.5 The relationship between active site environment hydrogen bond interaction and monomer centers of mass distances was uncovered.

Hydrogen bonds not only serve as a significant aspect of nonbonding interactions, but they play a pivotal role in protein folding, protein-ligand recognition, and enzyme catalysis (500). As a result, they make a significant contribution to the preservation of proteins' native conformation. In DPD, 5-FU forms covalent bonds, van der Waals interactions, and hydrogen bonds with the protein residues (129, 130, 398). The general H-bond instability in the active site region, which corresponded to an increase in COM between the active sites in Monomer A and Monomer B to their respective monomers, was characterized by drug exit in the affected monomers. This included: domain I M66V mutation, as well as active site mutations S534N and I543V of the clinically relevant missense mutation.

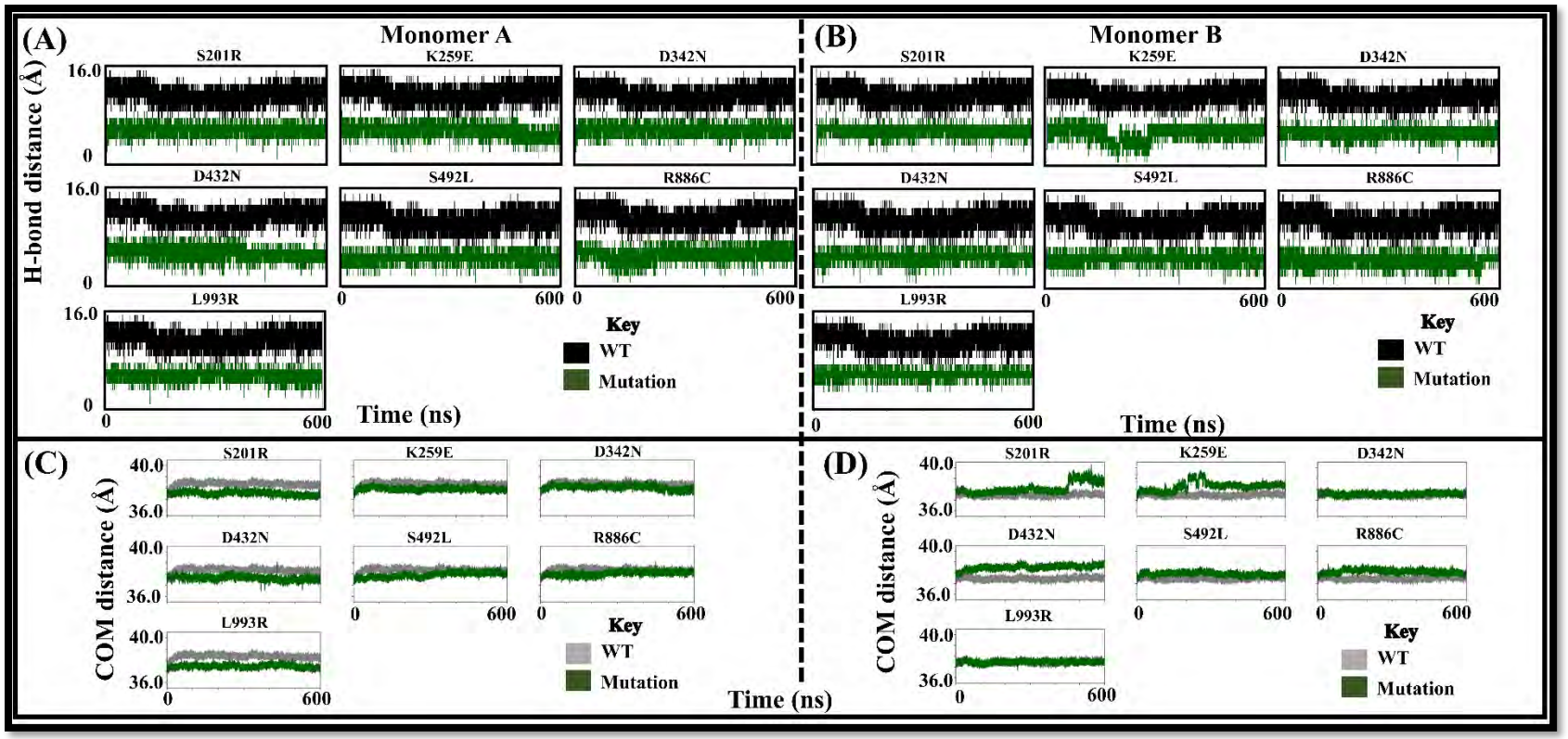


Figure 4.6: A comparison of the 5-FU binding environment in WT (in grey & black) and mutations (in green) using line plots. Hydrogen bond analysis during the 600 ns simulation period of 5-FU binding environment of A) Monomer A and B) Monomer B. Center of mass (COM) distance between 5-FU and C) Monomer A and D) Monomer B. The x-axis illustrates time, and the y-axis illustrates the number of H-bonds formed and COM distances, respectively.

Next, the ligand's (5-FU) positional stability was investigated by measuring the distance between the COM of the 5-FU and the COM of each monomer active site per protein system (Figure 4.6C & D). This was due to noticeable variation in the active site of different mutant proteins. Previous research suggested that this measurement was more accurate than ligand RMSD with respect to the reference structure's active site (285). The active site residues obtained previously from literature in the essential dynamic section were used for active site COM calculations in this section. An intricate network of hydrogen bonds was maintained within the active sites of the individual monomers in the activated WT across the simulation. Although all mutation ensembles had fewer H-bond interactions than the WT, the fewest interactions were found in Monomer A's D432N and L993R, as well as Monomer B's K259E. On the other hand, the COM distances between Monomer A and its corresponding active site B and vice versa revealed an increase in systems S201R, D432N, and L993R of Monomer A, as well as S201R, K259E, and D432N of Monomer B (Figure 4.6.C and D). These systems also showed reduced 5-FU H-bond interactions within the active site. As demonstrated in the next section, the increased COM distance may have resulted in a lower contact frequency between 5-FU and the active site interacting residue, as seen in D432N and L993R (Monomer A) and K259E (Monomer B). Unlike our clinically relevant data, where drug exit was observed due to reduced hydrogen stability and increased COM distance for active site domain mutations (S534N and I543V) and domain I M166V (remote) mutation. None of the remote mutations in this section result in the loss of the ligand. As discussed in the following section, this could be attributed to a reduction in the contact frequency between the 5-FU and coordinating residues in the active site binding environment. Hence, inhibiting proper binding and metabolism of 5-FU results in reduced enzymatic activities as demonstrated by *in vitro* studies (174, 176).

Considering the comprehensive global analysis encompassing RMSD, RMSD framewise analysis, RMSF, DCC, ED, and H-bond, the mutations S201R, D342N, S492L, and L993R consistently yielded notable results in most of these analyses. In other studies, the S201R mutation demonstrated no detectable enzymatic residual activity (< 12.5%, $P = 3.0 \times 10^{-7}$). However, reduced enzyme activities ranging between 12.5% and 25% were observed in the D342N ($P = 1.9 \times 10^{-6}$) and S492L ($P = 2.3 \times 10^{-4}$) mutations. Surprisingly, despite its location in Fe-S cluster coordinating domains, the L993R mutation did not exhibit a significant lack of enzymatic activity. Conversely, considerable reductions in enzyme activities were observed in

other mutations associated with the Fe-S cluster coordinating domains, not featured in this study, such as K958E, H978R, and V995F. Meanwhile, no significant differences in enzymatic activity changes were reported in K259E, D432N and R886C (168, 176).

PART II

4.3.3 Local analysis

In this study, we focused on the atomic role of transitioning between drug free and drug bound protein ensembles in the regulatory control of 5-FU metabolism. Weighted residue contact-map tools, as well as multi-centrality metrics such as *betweenness centrality (BC), closeness centrality (CC), degree centrality (DC), eigenvector centrality (EC), and katz centrality (KC)* were used to achieve this (33, 95). It should be noted that DPD equilibration was not completed in the majority of systems after 600 ns of MD simulation. This could be due to the large protein size (1017 residues per monomer) as well as the large area occupied by the loop regions, which account for approximately 45 percent of the total protein. Hence, the different asymmetric conformational changes per monomer were observed from RMSD, Rg, ED and RMSF results. Additionally, the DRN metric results from clinically relevant data, including remote mutations C29R, M166V, Y186C, and D949V, demonstrated how the loss and gain of hubs altered the orthosteric (displayed by WT) communication pathway via a compensatory mechanism that the protein used to maintain its functionality. As a result, we used the same approaches, such as contact map and heat map, as well as the DRN metric approach, to elaborate DPD functionality at the residue level. As a result, for reliable analysis, samples were extracted from the same time frame of an equilibrated 30 ns section of trajectory each monomer on a protein system prior to drug exit (Figure S4.1 and Tables S4.5.1 – S4.5.2).

4.3.3.1 Mutation-induced asymmetric shift in the active site inter-residual interaction per monomer revealed using contact heatmaps

Subsequently, the ligand-protein interactions were analyzed at the atomic level in order to identify immediate 5-FU contacts and residues at position 201, 259, 342, 432, 492, 886 and 993 in drug bound systems. These structures were extracted at the lowest energy wells using the MD-TASK-web comparative ED automated conformation extraction feature (499). Recently, heat

maps and residue-residue contact maps were used to characterize atovaquone drug resistance in the Pf cytochrome b protein (345), dimeric Mycobacterium tuberculosis KatG (285) and falcipain 2 in mutations including those linked to artemisinin resistance (287) and mutations associated with clinically relevant data in the DPD protein. Figure 4.7 portrays the interaction of 5-FU with the active site environments of the protein systems over an equilibrated simulated period utilizing weighted residue contact map tools (284, 499). The atomic contact frequency serves to score the 5-FU contact residues. Then after, the script assigns a score of 0 to 1 to each 5-FU atomic connection according to the frequency of interaction. The 5-FU-DPD interaction frequencies were then standardized across all systems. According to Figure 4.7, both WT proteins-maintained high contact frequency with the following residues:

- Monomer A: Thr575, Val583, Arg589, Ile590, Phe607, Leu608, Ile610, Glu611, Leu612, Ile613, Ser640, Ile641, Met642, Glu666, Leu667, Asn668, Leu669, Ser670, Cys671, Pro672, Ala683, Cys684, Gly685, Lys709, Leu710, Thr711, Pro712, Thr735, Thr735, Asn736, Thr737, Val712, Thr735, Asn736, Thr737, Val738, Gly764, Val765 and Ser766 (39).
- Monomer B: Thr575, Val583, Asn585, Val586, Arg589, Ile591, Phe607, Ile608, Asn609, Ile610, Glu611, Leu612, Ile613, Ser40, Ile641, Met642, Cys643, Tyr645, Glu666, Leu667, Asn668, Leu669, Ser670, Cys671, Pro672, His673, Ala683, Cys684, Ala683, Gly685, Gln686, Lys709, Leu710, Thr711, Pro712, Asn713, Thr735, Asn736, Val783, Gly764, Val765 and Ser766. (42)

Similarly, a majority of the mutants showed reduced/lost interaction with residues which predominantly coordinate 5-FU in the WT as follows:

- Monomer A: Leu612, Met642, Cys671, Ala683 and Gly685 .
- Monomer B: Cys643, Ser644, Thr645, Cys671, Lys709 and Leu710.

Generally, the main catalytic residue Cys671, situated in the loop (residues 670 - 683) area of the drug free DPD, has been observed in open conformation (129, 130). When activated (drug bound), the loop adopts a closed conformation, allowing Cys671 to be optimally placed for proton transfer. Two residues anchor Cys671 at opposite ends of the catalytic loop. The original anchor point is founded by hydrogen bonds and FMN's backbone oxygen atom binding Lys709 and side chain linking residue N736, respectively, at the location of residue Ser670. The second

anchor point is on residue His673, which is linked to Glu611, Met675, and Glu677 via a hydrogen bond (130, 131).

Considering catalytic residue C671, in Monomer A, the native protein had high frequency contact with Met642, Ser670, Pro672 and Ala683. Mutant D432N, S492L and R886 lost contact with Met642 and had reduced contact with Ser670, Pro672 and Ala683. Conversely, compensatory gain of inter-residue interactions across all mutant systems were seen in Met642, His673, Gly674, Met675 and Glu677. Similarly, in Monomer B, the native protein had high frequency contact with Ser670, His673, Pro672 and Cys684.

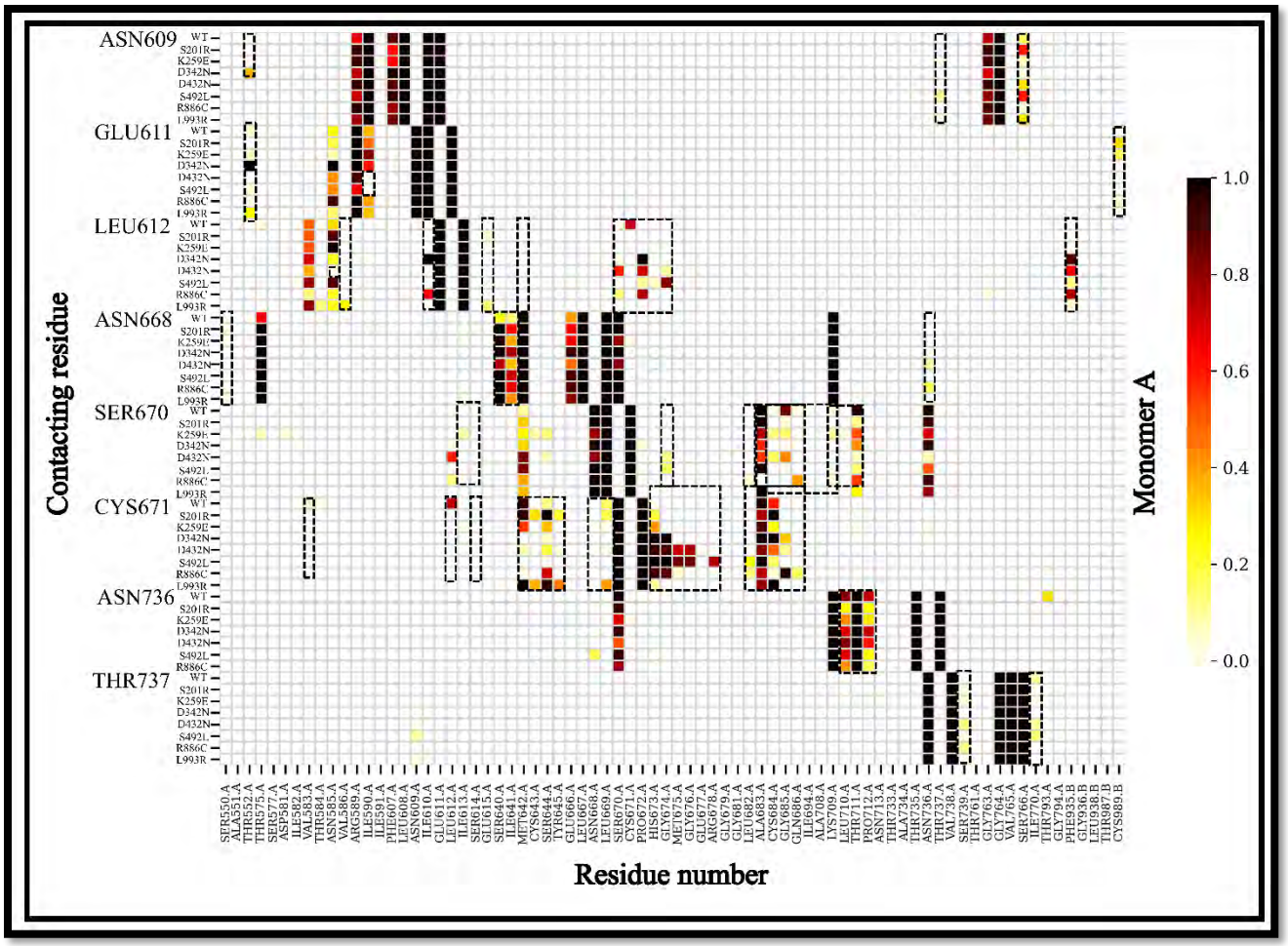


Figure 4.7.1: Drug bound DPD protein weighted residue contacts heatmap demonstrating the inter-residue interaction between 5-FU and the active site environment of A) Monomer A of 5-FU. The shaded boxes in each frame represent significant differences in the frequency of interaction between the 5-FU residues and the surrounding contacting residues in the WT versus the mutations. White (0) to dark-red (1.0) represents no contact (0) to strong contact (1.0).

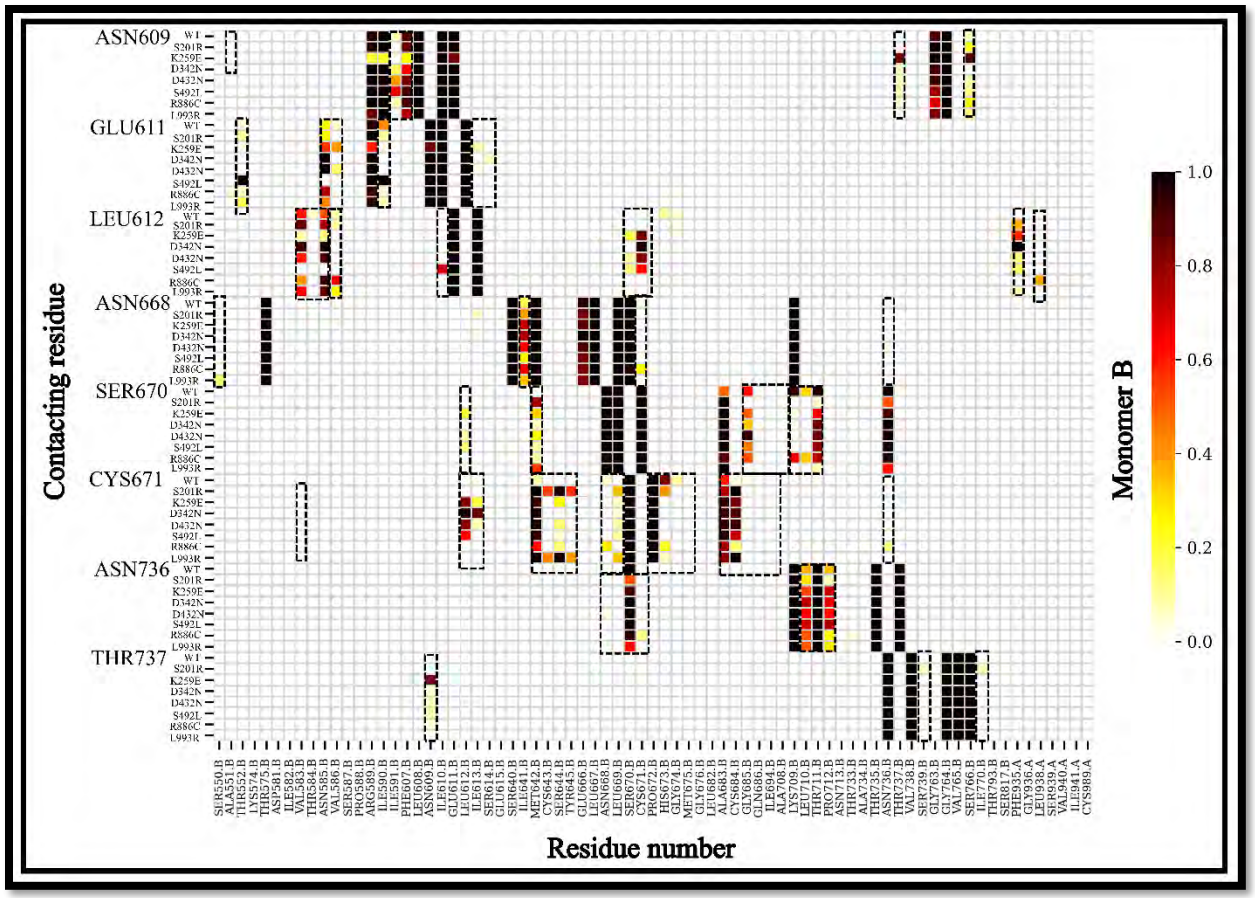


Figure 4.7.2: Drug bound DPD protein weighted residue contacts heatmap demonstrating the inter-residue interaction between 5-FU and the active site environment of A) Monomer B of 5-FU. The shaded boxes in each frame represent significant differences in the frequency of interaction between the 5-FU residues and the surrounding contacting residues in the WT versus the mutations. White (0) to dark-red (1.0) represents no contact (0) to strong contact (1.0).

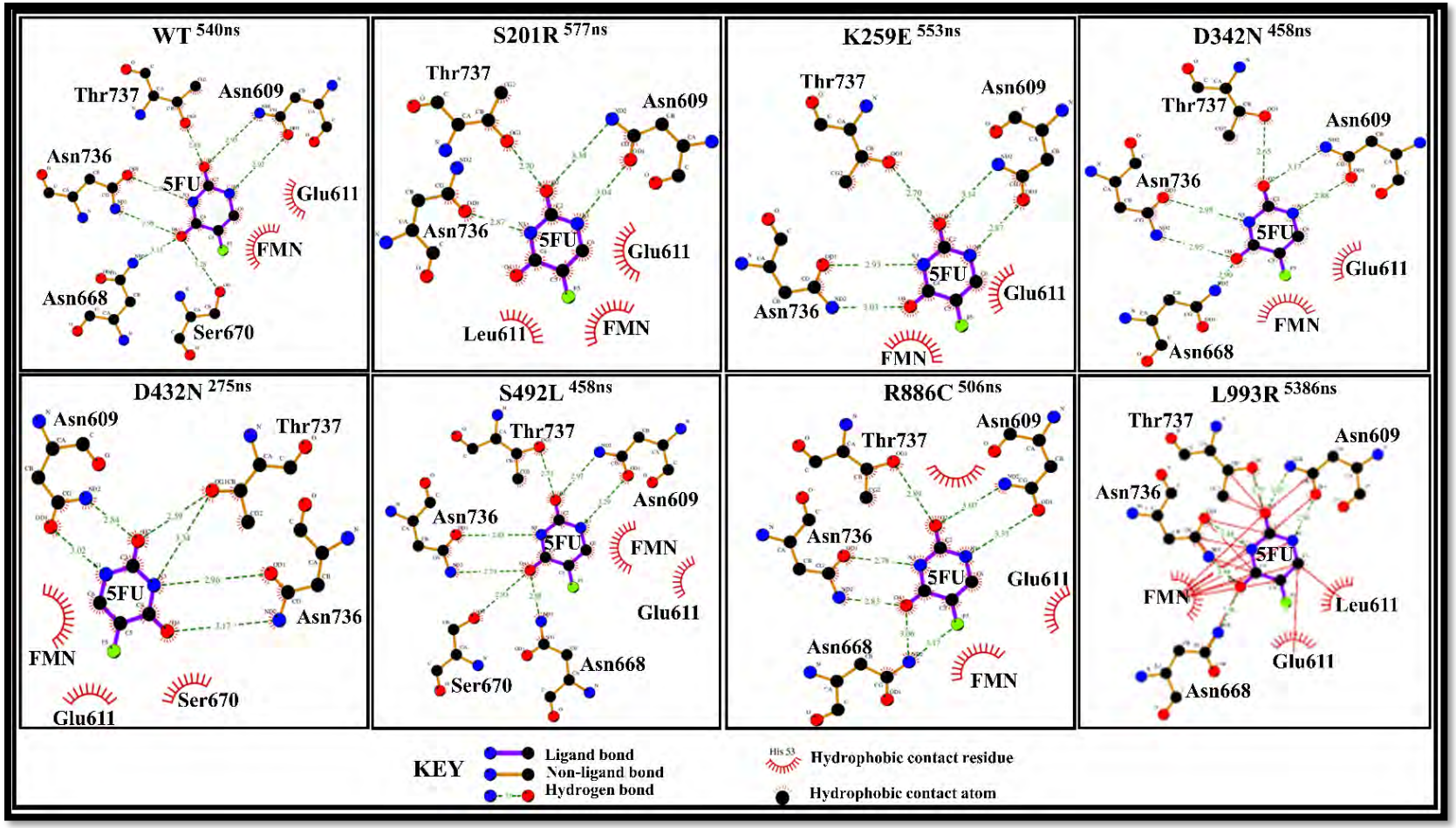


Figure 4.8.1: 5-FU interaction with residues in the active site environment of the WT and the seven mutations of Monomer A are represented in 2D using Ligplot+. The time frame in which the structures were extracted are also highlighted with the mutation names.

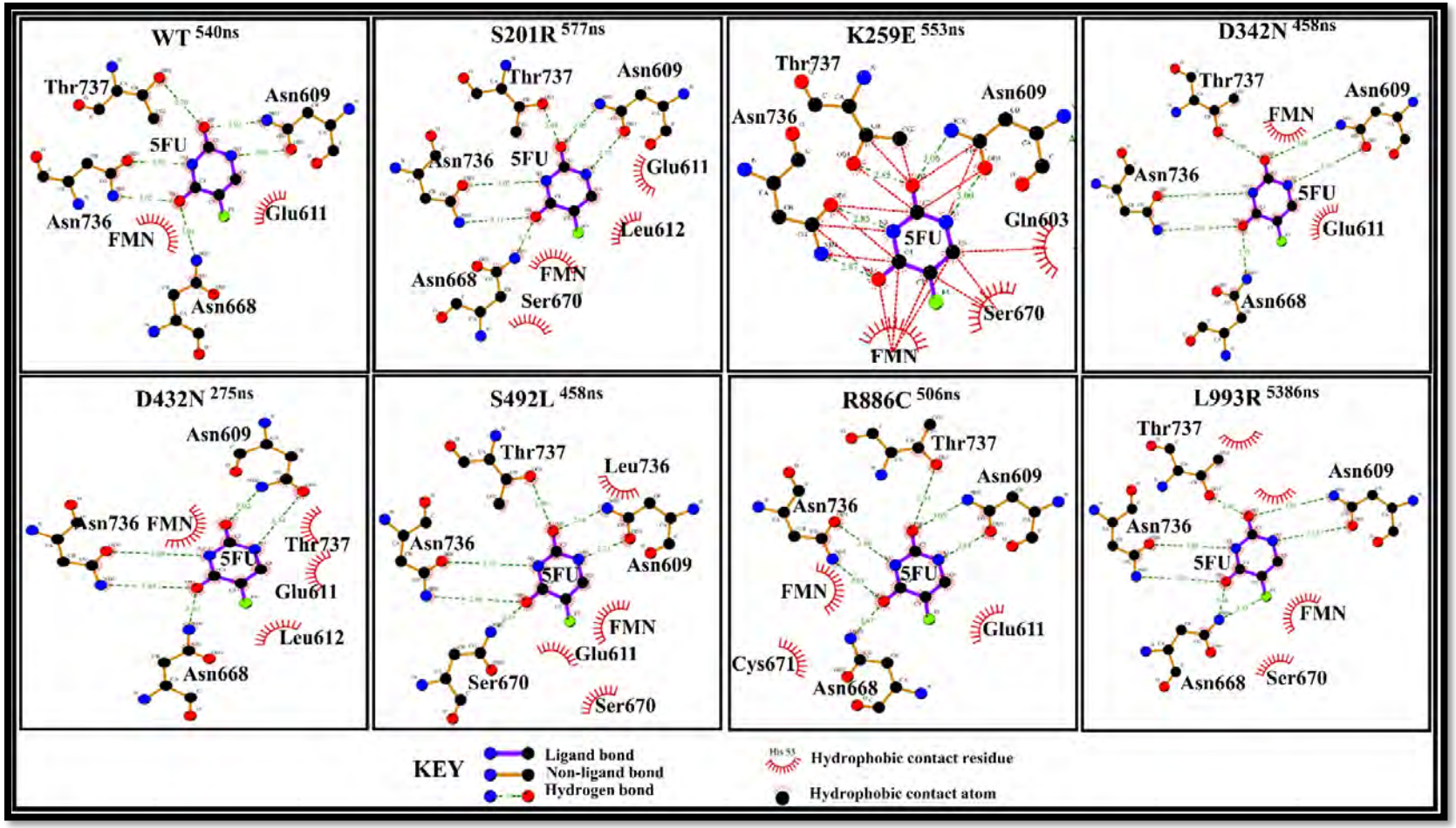


Figure 4.8.2: **5-FU interaction with residues in the active site environment of the WT and the seven mutations of Monomer B are represented in 2D using Ligplot+.** The time frame in which the structures were extracted are also highlighted with the mutation names.

On the other hand, compensatory gain in residue interaction was noted in Ser644 and Cys684. The difference in 5FU-protein interactions on monomer A versus monomer B could be an indication of allosteric signaling. Same tendencies were seen on previous study (285) The clinically relevant remote mutation M166V from our previous study showed retained/enhanced 5-FU interaction in monomer B. Based on these findings, the DPD mutational effect on 5-FU coordination and binding could be one mechanism for lowering drug efficacy and thus drug toxicity. Among the mutations that reduced enzymatic activity (176), K259E, S492L and D432N displayed a fairly asymmetric interaction profile of drug bound monomers. K259E and D432N exhibited increased H-bond instability which corresponded to an increase in COM distances between the COM of active-site to the COM distances of their respective monomers.

We then used 2D schematic diagrams from LigPlot+ v.2.2 to see how allosteric signaling affected the 5FU-DPD residue interaction per monomer (Figure 4.8.1 and 4.8.2) (96). 5FU was found to bind to the protein via two H-bonds with residues N609 and N736, as well as one H-bond with residues S668, N670, and T737 in Monomer A. Monomer B has a comparable 5-FU-protein binding to Monomer A, with the notable exception of N670. Changes in each monomer's 5FU-DPD residue interactions were observed not only in WT, but also in mutations.

In the WT, 5-FU forms hydrogen bonds with Asn609, Asn668, Ser670, Asn736 and THR737 in Monomer A and Asn609, Asn668, Asn736 and Thr737 in Monomer B. These major interactions were reduced and, in certain cases, lost completely in the mutant systems; for example, in Monomer A, S201R, K259E, and D432N lost H-bond interactions with Asn668 and Ser670, whereas D342N, R886C, and L993R lost H-bond interactions with Ser670. In Monomer B, K259E and S492L lost H-bonds with Asn668, while D432N lost H-bonds with Thr737. (Figure 4.7). Although Offer et al. did not observe any change in enzymatic function during their investigation (168), a patient carrying the K259E mutation experienced toxicity following 5-FU administration (180). The alteration in polarity resulting from the K259E mutation is likely responsible for either partial or complete inactivation of the enzyme. This mutation introduces an additional negatively charged residue, leading to a slight shift in the loop backbone away from the hydrophobic β-sheet environment. Consequently, the hydrogen bonds formed with neighboring Leu261 and Ser262 are weakened. This weakening further affects the hydrogen bond between Gly258-N and Lys219-O, which plays a critical role in protecting the adenine moiety of FAD from solvent access, thus affecting cofactor binding. Additionally, the

positioning of Glu218, which interacts with the hydroxyl group (O2*) of FAD, may also be impacted. These changes in FAD binding and orientation within the Glu259 monomer are expected to disrupt the subsequent hydride transfer from the electron-donating to the active site. Moreover, in the case of the Ser492Leu mutation, it is worth noting that serine is positioned within the cofactor FAD's binding site. The distances between the hydroxyl group of serine and the atoms O3* and O5* of the ribityl moiety, as well as the OP2 atom of one of the phosphates of FAD, measure 3.5 Å, 3.5 Å, and 3.8 Å in the WT (Figure 4.1), respectively (171). Although these distances are greater than those typically associated with strong hydrogen bonds, they clearly indicate that the substitution of serine with the larger and hydrophobic side-chain of leucine would hinder the proper binding of FAD. As a result, this disruption would eventually impact the electron transport to the active site.

It is important to note that 5-FU did not leave any of the system's monomers active sites. In a previous clinically relevant analysis, the drug was found to exit from the remote mutation M166V, which was attributed to a loss of H-bonds. In the remote mutation in this chapter, none of the mutant proteins showed drug exit from both monomers.

4.3.3.2 Allosteric signaling pathways were revealed in the presence of mutations using averaged DRN multi-metric analysis

Missense mutations have been found to disrupt protein structure responses at the global (as illustrated by the RMSD, RMSF, Rg DCC, and essential dynamics).and local (as demonstrated by residue-residue interactions and the creation of communication network revealing allosteric signaling network) levels (374, 501, 502). To review what was mentioned previously, in a protein communication network, nodes representing residues interact non-covalently with other nodes forming edges (285, 437). Centrality is defined as a node's prominence in a protein communication network. Changes in the residue chain caused by mutations may result in the reconfiguration of protein network architectures that alter node centrality (283, 285, 287). We investigated the key residues for DPD inter- and intra-monomeric communication in this study using five different averaged centrality metrics: *betweenness centrality (BC), closeness centrality (CC), degree centrality (DC), eigenvector centrality (EC), and katz centrality (KC)*. To guarantee the accuracy of the results, prior to drug exit, samples were extracted from the same time frame of an equilibrated 30 ns section of trajectory for each

monomer on a protein system (Figure S4.1, Table S4.5.1, and Table 4.5.2). The top 5% of all residues in all WT and mutation proteins were then revealed. Given the large size of the protein, we sought to enhance the accuracy and sensitivity of the global top 5% residues. As a result, we calculated the 2, 3, 4, and 5% high significant residues, dubbed the hubs, in all averaged centrality metrics. We then defined the various hubs that could be identified in our studies, such as constitutive hubs: hubs found in either drug free or drug bound monomers; persistent hubs: hubs found in both drug free and drug bound monomers of a metric; and super-persistent hubs: persistent hubs observed in both drug free and drug bound monomers across all averaged DRN metrics.

This section of the study, in particular, demonstrated the significance of unique network data provided by each centrality metric in establishing the overall DPD communication system. As seen in our earlier clinical data mutation results, the network information pattern was mostly concentrated along interface residues and the oxidoreduction communication channel. In this section, the assessment of the remote mutation data focused on the 4% of hubs that were deemed most informative. First, as previously established, it was represented in the form of a heatmap. To better visualize the mutation-induced hub networks, the hubs of each mutation *(*ΔBC*,* ΔCC*,* ΔDC*,* ΔEC *and* ΔKC) in relation to the WT were mapped in DPD 3D structure. The following subsections provide further explanation:

Betweenness centrality (BC) **elucidates the allosteric signaling network intrinsic to the oxidoreduction catalytic channel in mutations versus WT.** In our earlier study (Chapter 3), we highlighted two distinct *BC* communications pathways that are more intrinsic to the oxidoreduction DPD channel (Figure 3.8.1). We went on to show how the presence of mutations (clinically studied mutations) leads to compensatory gain in hubs (residues) that form allosteric communication pathways to help maintain protein function. *BC* defines the number of shortest paths that pass through a node in a specific residue interacting network, thereby quantifying the rate of node utilization. Previous research has shown that *BC* is one of the most useful metrics for calculating not only main allosteric residues (420) but also substrate and mutation-induced residue network modifications in a variety of proteins (283, 287, 374, 422, 439). Figure 4.9.1 shows heatmaps of the top 4% global *BC* variation of inter-system, inter-monomer, and intra-monomer mutations key hubs in reference to the WT. The asymmetric monomer behavior of dimeric proteins has previously been discussed (266, 285). Similar patterns were observed in

clinically relevant DPD mutation data in this chapter. The asymmetric behavior of monomers (A and B) was observed in both the drug free and drug bound states but had no implication on the persistent hubs (L155). Prior studies have also demonstrated that monomers in a dimeric protein behave asymmetrically (283). Persistent hub Leu155, positioned at the α-helix of domain I, was identified as being linked to Gln156, which is implicated in ligation (11) or forms a unique bond with Fe^{2+} (from the first $Fe^{2+}{}_4S^{2-}{}_4$ clusters & residue 27-172) via an oxygen atom. As a result, L155 may be critical in maintaining stability within the $Fe^{2+}{}_4S^{2-}{}_4$ clusters environment. We found 50 hubs in Monomer A of both activated and inactivated WT, whereas Monomer B had 42 hubs in the inactivated state and 45 in the activated state. Moreover, 20 constitutive hubs (Asp61, Leu80, Ile101, Thr102, Asn108, Ile116, Phe117, Phe121, Thr132, Ser133, Leu142, Leu155, Arg372, Val373, Gly597, Phe598, Met599, Thr768, Leu775, and Thr799) and 14 constitutive hubs (Ile62, Ile101, Thr102, Gly127, Phe131, Ser133, Val137, Leu142, Ser149, Leu155, Ile370, Ala37) were identified in drug free and drug bound ensembles respectively. In relation to the WT, these hubs were either retained, lost, or new ones were generated in each of the mutation systems (Table S4.6.1). The following hubs were identified in at least six of the studied systems in either monomer of drug bound: Asn108, Gly138, Leu142, Ala372 and Val373. Residues Ala372 andV373, which is located on the interface of the NADPH binding domain (residues 287 - 441) (130) facilitates interactions with the FAD binding domain.

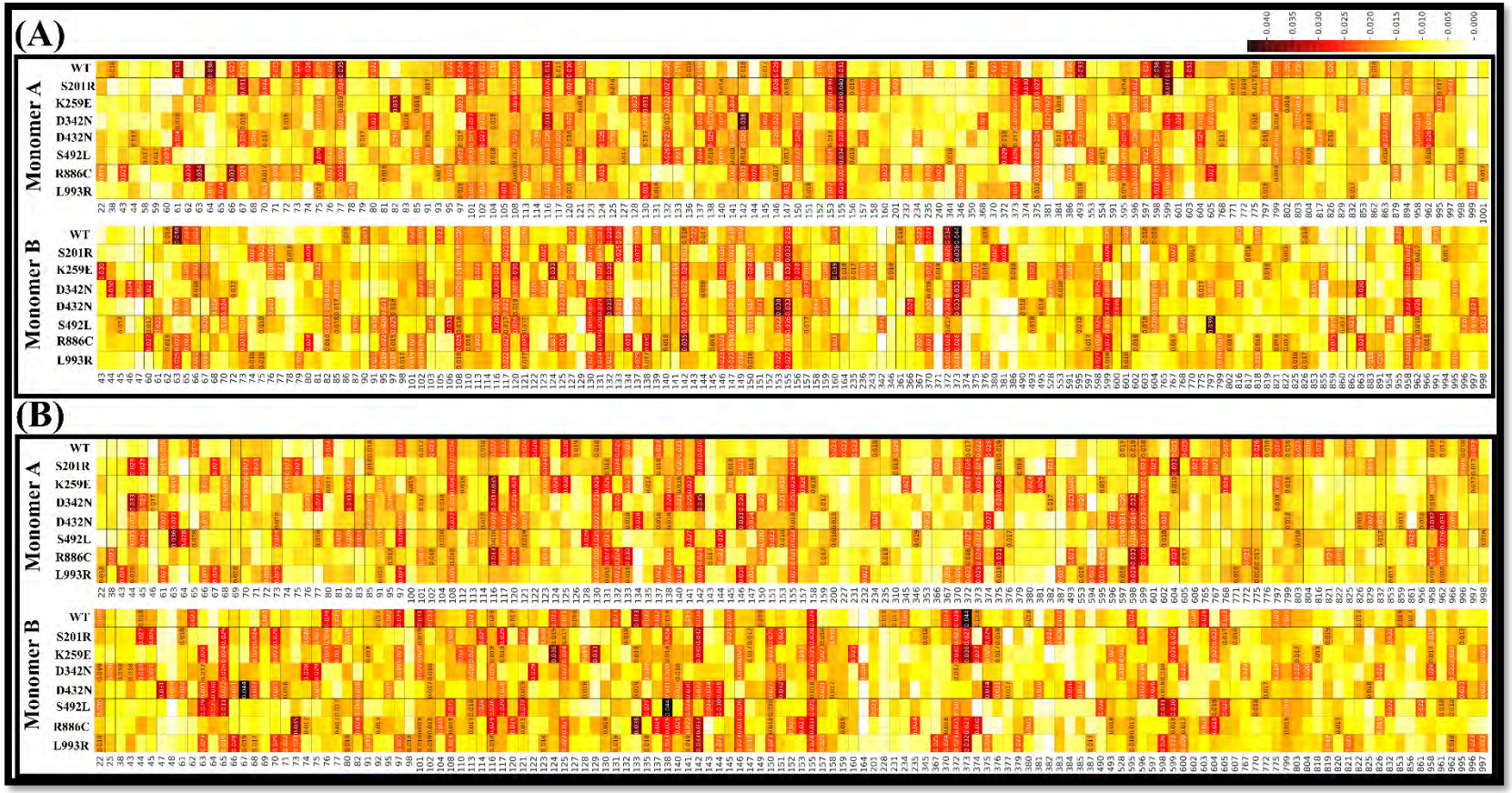


Figure 4.9.1: Heatmap depicting global top 4% key hubs for averaged *betweenness centrality (BC)*. Heatmap of the (A) drug free (inactivated) and (B) drug bound (activated) ensembles. The x-axis represents protein residues, while the y-axis represents WT and the 7 mutation (remote) proteins. Centrality values of hubs are annotated while the values of low to high centrality are colored white to black, respectively.

This hub was discovered in most of the systems, with the exception of R886C (drug free Monomer A); D324N and D432N (drug bound Monomer B); and D342N, D432N, and S492L (drug bound Monomer B). Residue 142 interacts with domain I $Fe^{2+}{}_4S^{2-}{}_4$ clusters binding residue Cys140, whereas Asn108 was discovered to connect three domains: domain I (first set of $Fe^{2+}{}_4S^{2-}{}_4$ clusters and residue 27-172); active site domain IV (FMN, 5-FU, and residues 525-847); and domain V (second set of $Fe^{2+}{}_4S^{2-}{}_4$ clusters and residues 848-1025) forming a dimerization structure (130). Interestingly, in drug bound system some residues (Monomer A: Asn44) and (Monomer B: Thr65) were exclusive to the mutant systems. For instance short-range contact analysis using contact maps showed that residue Asn44 in Monomer A gains new contact with Asp46, Cys52, Lys54, Se90, Ser133, asp134 and Ala144, whereas, residue Thr65 in Monomer B gained new contacts with Thr132, Val137, Lys681 and Gly862 (Figure S4.2). The observed increase in centrality for these residues might be a compensatory mechanism to maintain mutant's functionality.

Surprisingly, 3D mapping of *BC* hubs to DPD WT revealed two well-defined allosteric signaling routes intrinsic to the oxidoreduction catalytic channel, from NADPH-FAD-$Fe^{2+}{}_4S^{2-}{}_4$ clusters-FMN to 5FU (Figure 4.9.2). However, one monomer dominated the number of hubs over the other. These nodes were mostly found near interface residues, cofactors, and substrates (Figure S8 and Table S11). Previously, we discovered that *BC* network hubs are found at interface residues (285), many of which have been linked to protein surface stability and intra-protein communication (440). In comparison to WT hubs, the mutation systems lost the majority of the hubs and established new hubs around the oxidoreduction catalytic channel components, with the active site being the most impacted, especially in S201R, D342N, S492L and L993R. A similar pattern was noticed in drug free protein systems (Figure S4.2.1). These results are in agreement with the global analyses finding which showed S201R, D342N, S492L, and L993R consistently yielded notable conformational changes that could have an impact on DPD structure and function. Supporting investigations have shown total loss of function in enzymatic activities of S201R (<12.5 %) whereas, reduced residual activities (12.5-25 %) were noted in D342N and S492L, however, L993R did exhibit significant change in loss or gain of DPD enzymatic activity.

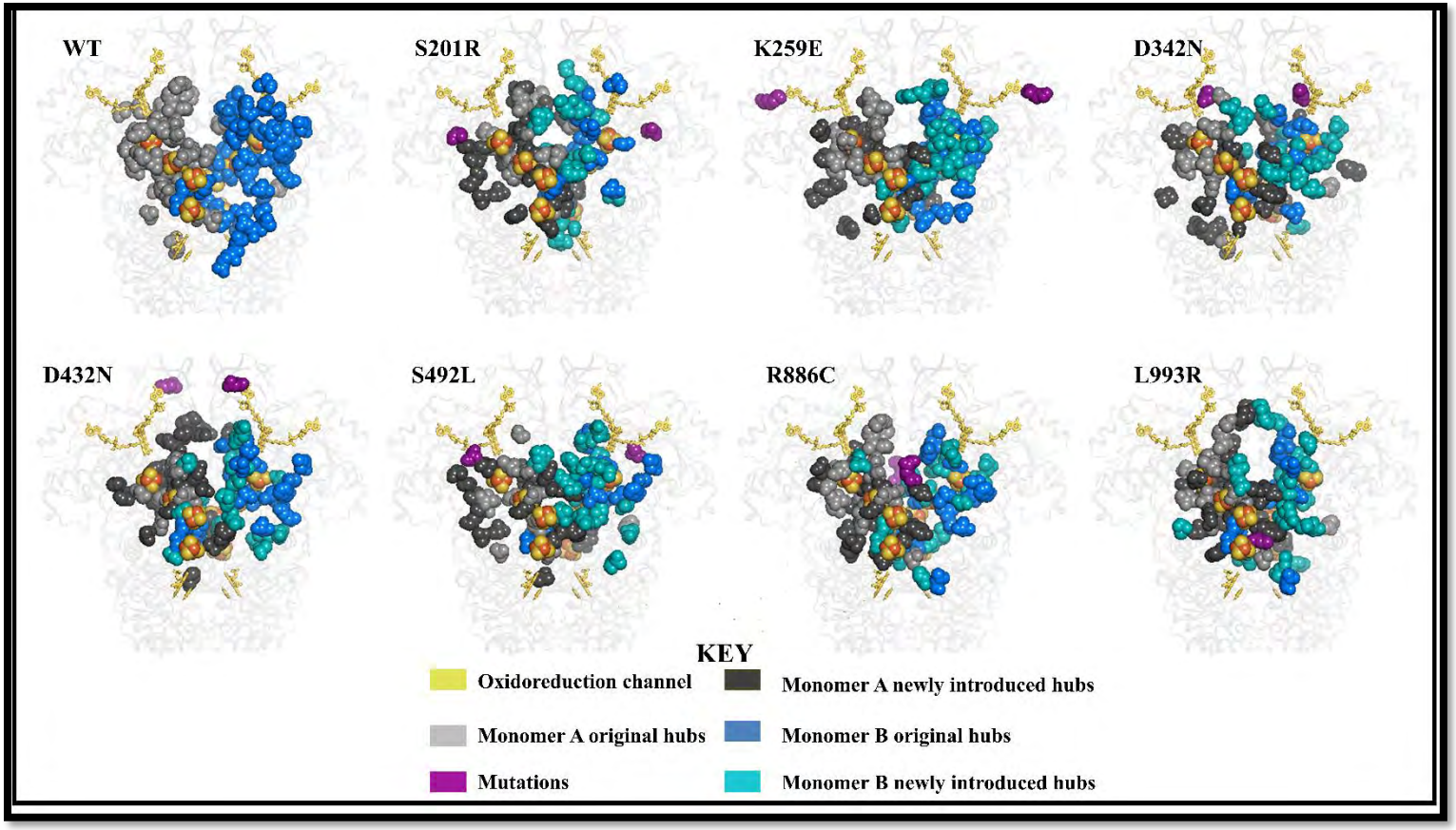


Figure 4.9.2: 3D cartoon representation of the distribution of *betweenness centrality* hubs of the drug bound ensembles along the oxidoreduction catalytic channel. Monomer A WT *BC* hubs are denoted as light-grey, Monomer B WT *BC* hubs as blue. Mutation-induced newly introduced hubs shown in dark-grey in Monomer A and sky-blue Monomer B. Oxidoreduction catalytic channel (NADPH, FAD, $Fe^{2+}{}_4S^{2-}{}_4$ clusters, FMN and 5FU) is shown in yellow. The mutations are represented in indigo.

Some of the hubs forming these ensembles from the active site to NADPH cofactor in drug bound state are as follows:

- Monomer A: Asp61, Thr65, Leu80, Ala85, Ile97, Ile101, Thr102, Asn108, Lys114, Ile116, Phe117, Pro121, Leu122, Gly123, Thr125, Cys126, Cys130, Thr132, Ser133, Cys140, Leu142, Leu155, Ile200, Ser227, Ile231, Phe234, Leu310, Ala372, Val373, Pro374, Glu375, Gly597, Pro598, Met599, Gln604, Ser605, Thr768, Leu775, Arg776, Thr797, Ala799, Leu803, Gln804, Cys816, Thr825, Ser958, Gly962 and 997.
- Monomer B: Thr22, Asn44, Ile62, Gly68, Ala76, Met77, Pro92, Leu95, Ile97, Ile101, Thr102, Ala113, Ile116, Asn120, Leu124, Thr125, Gly127, Pro131, Ser133, Leu135, Val137, Gly138, Asn141, Leu142, Thr145, Ser149, Leu155, Ser228, Ile231, Phe234, Arg370, Ala372, Val373, Pro374, Ala380, Glu383, Thr595, Pro602, Gly603,Ser 605, Leu775, Ala799, Gln822, Ser853 and Thr856.

Residues 367-376 are of particular interest due to their varied interactions with FAD or NADPH binding residues in Monomers A of D432N, K259E, D432N and R886C and B of S201R, D342N and S492L. Particularly, Glu376 (in Monomer A: D432N, K259E, D432N, and R886C) and Pro374 (in Monomer B: S492L and D342N) showed increased interaction compared to the wild type (WT), suggesting a potential allosteric compensatory mechanism that helps maintain protein functionality. (Figure 4.9.1). This finding could explain the absence of significant loss in enzymatic function observed in D432N, R886C, and K259E in previous studies conducted by Offer et al. (168). However, this compensatory mechanism may also disrupt the cofactor binding environment and the oxidoreduction mechanism, as observed in S201R, D342N, and S492L, which could impact the efficiency of the protein. *In vitro* enzymatic experiments have demonstrated a partial loss of functionality in individuals carrying these mutations(168, 445, 503). Similarly, our clinically associated mutation, which is associated with DPD deficiency, exhibits a comparable compensatory allosteric pathway in close proximity to the oxidoreduction channel.

***Degree of centrality (DC)* elucidates the allosteric signaling network extrinsic to the oxidoreduction catalytic channel in mutations versus WT**. Previously, we demonstrated using *BC* and *DC* how clinically associated DPD missense mutations formed two pairs of allosteric pathways to the active site, with some of these mutations found in the catalytic site domain

(S534N and I543V) and others in the remote/distal domains (C29R, M166V, Y186C and D949V). We used heatmaps (Figure 4.10.1) and 3D mapping . (Figure 4.10.2), to investigate the number of adjoining nodes around a single node (*DC*) using data that had not previously been clinically studied (remote mutations) (Figure 4.10.1). Asymmetric behavior of the monomers was demonstrated in both the drug free and drug bound systems of the top 4% *DC*, similar to observations made in the *BC* analysis. With the exception of Leu155, more persistent hubs (Asn108, Val778, Thr779, Ile781, Ile790, and Ser808) were noted in *DC* as opposed to *BC*. L108 is an interface residue interlinking domain I (first set of $Fe^{2+}{}_4S^{2-}{}_4$ clusters and residue 27-172) and active site domain IV (FMN, 5-FU & residues 525-847). Active site domain interface residues V778, T778, I781, I790, and S808 that are involved in maintaining the catalytic area's stability and function (129-131). V788 in the domain α-helix had the most interactions with residues in contact with the substrate (5-FU), indicating a critical role in the active site region. Furthermore, the following hubs were present in at least five of the drug bound systems: Tyr110, Ala113, Thr125, Ala198, Gly282, Val493, Val732, Leu813 and Leu837 in Monomer A, as well as Tyr110, Ala138, Ala198, Ser199, Ser201, Cys202, Ile482, Ala499, Thr554, Leu653, Leu710, Gly740 and Leu837 in Monomer B. They form part of dimeric structure $Fe^{2+}{}_4S^{2-}{}_4$ clusters domain I (residues 27-172), FAD binding domain II (residue 173-286 and 442-524), NADPH binding domain III (residues 287-441) or FMN binding domain IV (525-847). Of interest are residues 437-585 which showed increased, reduced, gain or loss of interaction with residues correlating to either FAD, FMN or 5-FU binding residues in Monomer A and B.

3D mapping of the top 4% *DC* hubs in drug free and drug bound systems were done for better visualization Figures 4.10.2 and S4.2.2. Due to the presence of mutation and 5-FU binding, two distinct cross-monomer clustered ensembles of *DC* hubs allosteric signaling pathways were discovered along each monomer's oxidoreduction catalytic channel. These pathways resembled those observed in *BC,* but they were more extrinsic to the oxidoreduction catalytic channel than the inner pathway observed in *BC* (Figure 4.9.1). The communication network pathway of the activated WT *DC* was formed by 52 hubs in Monomer A and 44 hubs in Monomer B (Table S4.6.2).

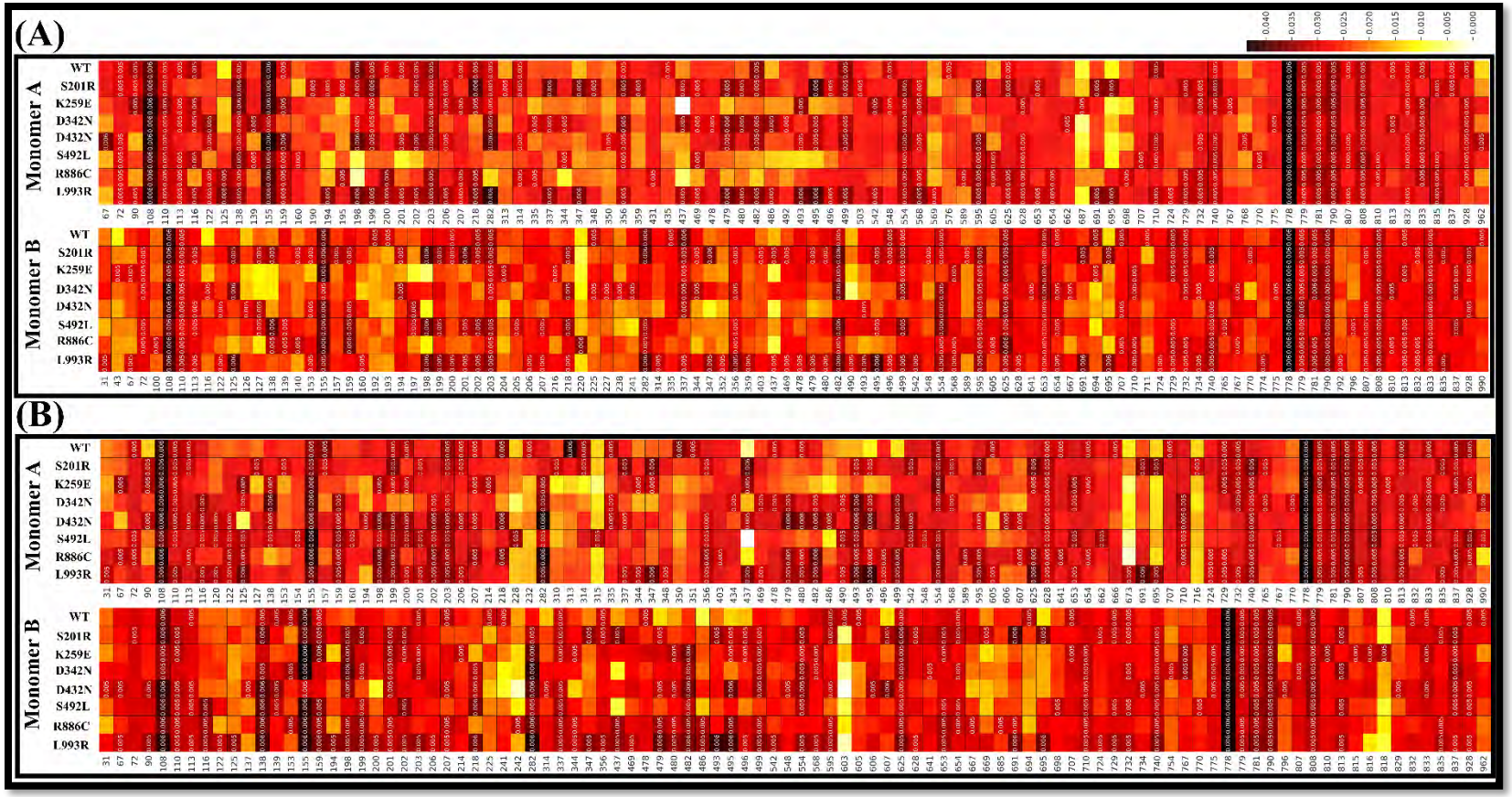


Figure 4.10.1: Heatmap depicting global top 4% key hubs for averaged *degree of centrality (DC)* heatmap of the (A) drug free (inactivated) and (B) drug bound (activated) ensembles. The x-axis represents protein residues, while the y-axis represents WT and the 7 mutation (remote) proteins. Centrality values of hubs are annotated, with the values of low to high centrality colored white to black, respectively.

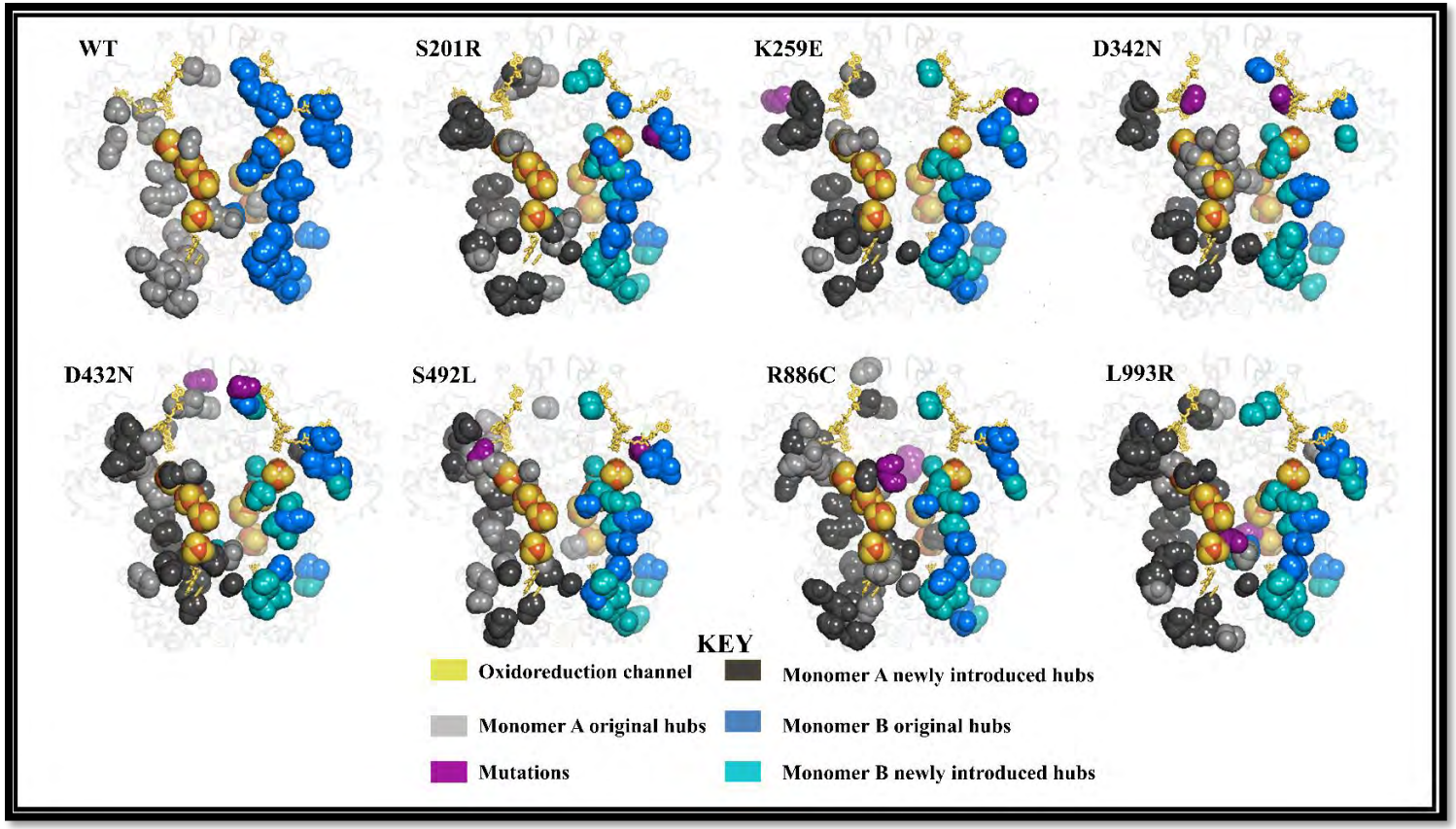


Figure 4.10.2: 3D cartoon representation of the distribution of *degree of centrality* hubs of the drug bound ensembles along the oxidoreduction catalytic channel. Monomer A WT *DC* hubs are denoted as light-grey, Monomer B WT *DC* hubs as blue. Mutation-induced newly introduced hubs shown in dark-grey in Monomer A and sky-blue Monomer B. Oxidoreduction catalytic channel (NADPH, FAD, $Fe^{2+}{}_4S^{2-}{}_4$ clusters, FMN and 5FU) is shown in yellow. The mutations are represented in indigo.

In our previous analysis, one pathway dominated in the number of hubs, similar to current *BC* finding. Due to the presence of mutations or 5FU-binding, both the dominant and sub-dominant pathways were formed from new and/or existing hubs (Figure 4.10.2 and Figure S4.2.2). In comparison to the drug-bound WT, a new NADPH binding residue Ala486 (Monomer A: D432N: Monomer B: R886C & L993R) was formed. This residue forms an extension of residues 485-488 backbone conformation, which is stabilized by two H-bonds formed between the main chain nitrogen Atoms of Asn487 and Thr488 and the carboxyl-group of Glu491 (130). This is thought to be the most important conformational change for proper positioning of the nicotinamide moiety of NADPH involved in Asp342. In a ligand-free system, the FAD binding residue Asp342 forms H-bonds between the backbones of nitrogen atoms Arg371 and Ala372 and the oxygen atom Ala340 (130). New formed interactions were seen in active residue 606 and 607 (Monomer A: R886C & D432N) and (Monomer B: S201R and D432N). This is a good example of probable compensatory allosteric communication signaling around the FAD and NADPH region. The residue Pro493, which exhibited increased/retained FAD interactions in Monomer A (S201R, D342N, D432N, S492L, R886C and L993R), was an example of distinct asymmetric behavior of monomers. Equally, *in vitro* studies indicate weak evidence of changes in enzyme activity as a result of these mutations; thus, these data are deemed insufficient to support the 5-FU dose adjustment guidelines (164, 168, 176, 451, 504, 505).

In line with the clinically significant mutation, a DCC analysis was conducted to investigate the interplay between hub residues based on *BC* and *DC* in both the drug-free and drug-bound states of the DPD enzyme of mutations of uncertain significance. Remarkably, each mutation displayed a positive correlation between *DC* and *BC* residue hubs. This positive correlation underscores the crucial role played by these hub residues in maintaining the functional integrity of the network (Figure S4.2.3 and S4.2.4).

***Closeness Centrality (CC), Eigen Centrality (EC)*, and *Katz Centrality (KC)* analysis revealed a distinct regional communication pattern within DPD.** Previous research has revealed that interface residues and protein core residues (central nodes) that are close to all other residues in the network have high *CC* values (59, 266, 285-287, 427). These were supported by our earlier finding on mutations associated with clinically relevant data (including both catalytic site and remote mutations) as clarified in Chapter three. With this in mind, we

sought to investigate key *CC* hubs first, then *EC* hubs, and finally *KC* hubs solely based on remote mutations. Other studies such as that of Okeke and colleagues (287) found that averaged *CC* was the most restrictive in assigning centrality, followed by *KC* and *DC*, which matched our findings.

The heatmap representation of averaged *CC* (Figure 4.11.1 and Figure S4.2.3) revealed that *CC* had most persistent hubs (Gly71, Ala72, Leu73, Ala74, Glu75, Ala76, Cys79, Leu80, Ala101, Ile104, Leu142, Ile150, Leu155, Pro598, Met599, Pro602, and Ser605) than either *BC, DC, KC*, similar to our previous clinically significant analysis report (Table S4.6.3). These are mainly $Fe^{2+}{}_4S^{2-}{}_4$ clusters and catalytic site interface residues. According to Jayashree etal., interface residues are responsible for maintaining protein stability and function (440). Residues Cys79 is an $Fe^{2+}{}_4S^{2-}{}_4$ clusters binding, whereas Pro602 and Ser605 are catalytic pocket binding residues that are part of the inter-connecting FMN and 5-FU residues (130, 131). Additionally, these hubs were less susceptible to mutations, Krishnamoorthy et al., (468), refer to such regions as mutational cold spots. As a result, we proposed using persistent hub data to discover cold spot residues.

Mapping the hubs to 3D protein structures revealed that all the global top 4% averaged *CC* hubs are mainly located in the protein's center, specifically in the $Fe^{2+}{}_4S^{2-}{}_4$ clusters region of the dimerization domain (Figure 37.2), hence depicting the importance of $Fe^{2+}{}_4S^{2-}{}_4$ cluster domains in the functionality of DPD enzyme in respect to dysfunctional degradation of 5-FU. Collective calculations from *BC*, *CC,* and *DC* were used to determine which interface residues influence communication and enable DPD dimerization in the studied systems. These are new target areas for drug development and innovation in the modulation of 5-FU efficacy and toxicity in relation to DPD.

The correlation between intermonomer distance measurements (Figure 4.6) and the number of *CC* hubs was an intriguing finding (Figure 4.11.2). The results showed that steady intermonomer distances were observed in the WT and the majority of the mutant systems, with the exception of S201R, D432N, and L993R of Monomer A and S201R, K259E, D432N, and R886C of Monomer B, which had a minimal (<20 Å) increase in distance compared to the WT across all simulations. These were distinguished by the loss of *CC* hubs in each system's corresponding monomers. Not only did the study by Barozi et al., revealed similar findings (374), but our previous clinically relevant data did as well. D342N and D432N FAD binding

domain II mutations lead to the loss of centrality/hub status of residues Cys140 and Leu142 in Monomer B and A of these systems, respectively.

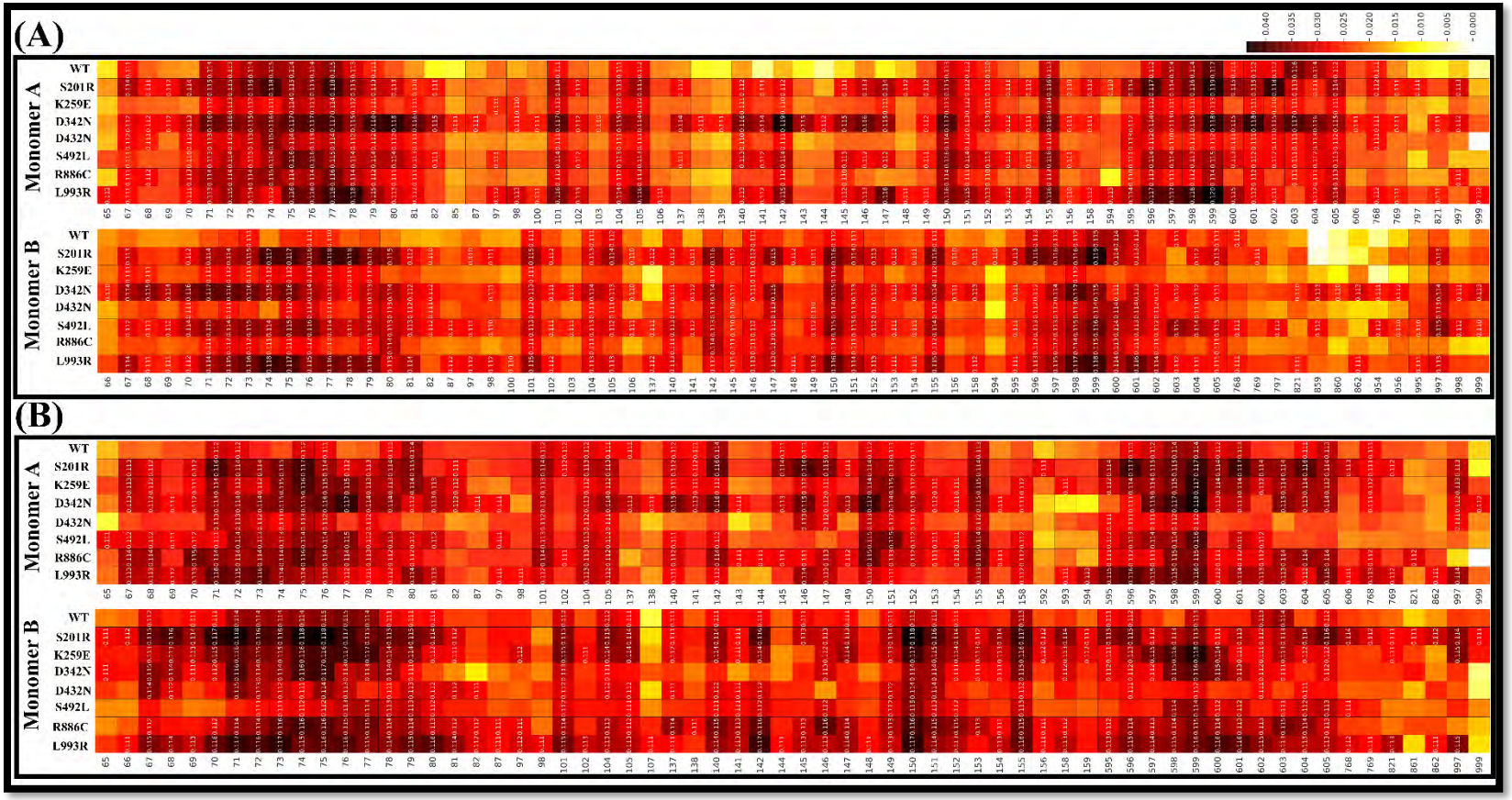


Figure 4.11.1: Heatmap depicting global top 4% key hubs for averaged *closeness centrality (CC)* heatmap of the (A) drug free (inactivated) and (B) drug bound (activated) ensembles. The x-axis represents protein residues, while the y-axis represents WT and the 7 mutation (remote) proteins. Centrality values of hubs are annotated, with the values of low to high centrality colored white to black, respectively.

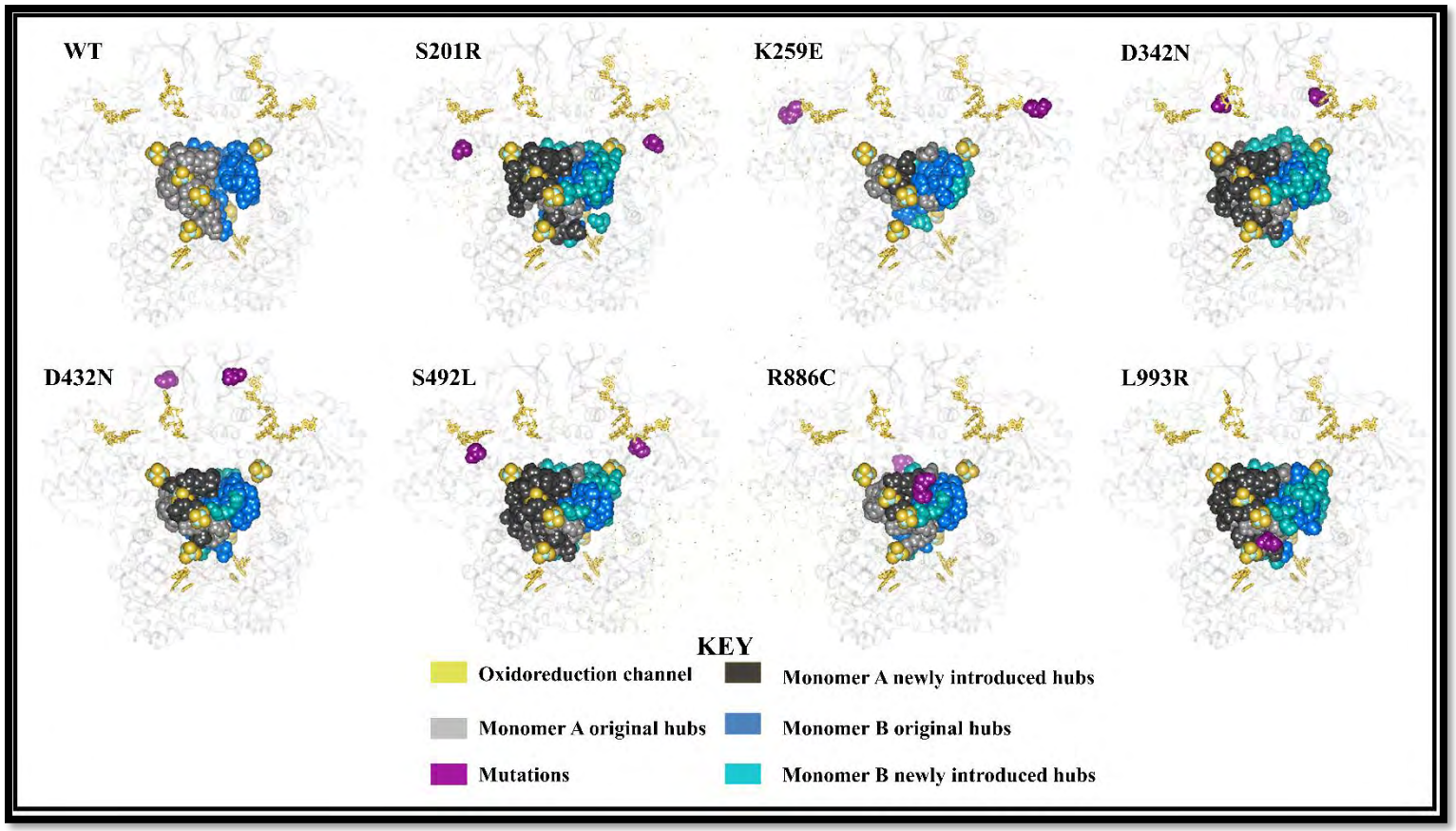


Figure 4.11.2: 3D cartoon representation of the distribution of *closeness centrality* hubs of the drug bound ensembles along the oxidoreduction catalytic channel. Monomer A WT *CC* hubs are denoted as light-grey, Monomer B WT *CC* hubs as blue. Mutation-induced newly introduced hubs shown in dark-grey in Monomer A and sky-blue in Monomer B. Oxidoreduction catalytic channel (NADPH, FAD, $Fe^{2+}{}_4S^{2-}{}_4$ clusters, FMN and 5FU) is shown in yellow. The mutations are represented in indigo.

Cys140 form iron sulfide bond with domain I $Fe^{2+}{}_4S^{2-}{}_4$ clusters (130), whereas Leu142 is an interconnecting residue to Cys140. With the exception of D432N, a compensatory gain in allosteric network hubs was observed in drug-bound system residues Arg70, Met77, Arg78, Phe158, Pro602, and Gly603. Monomer B exhibited compensatory gain in allosteric network hubs in residues Gly68, Glu146, Gly154, Gln156, Phe158, Tyr600, and Gly601, with the exception of D432N (all instances), D342N, and S492L. (some instances). Surprisingly, these hubs are found within the $Fe^{2+}{}_4S^{2-}{}_4$ clusters and active site region. As a result, we propose that they participate in a compensatory allosteric communication network to help maintain protein function. The fact that these residues participate in the allosteric communication network in Monomer A, but not necessarily in Monomer B supports the asymmetric behavior of each system. Collectively, the *CC* analysis demonstrates the importance of proteins core components, primarily all $Fe^{2+}{}_4S^{2-}{}_4$ clusters, in maintaining protein structure. Residues with high *EC* values, according to Fautch et al., form large, well-defined spatial regions. These "patch" regions are most likely the location of vital biological functions (506). In Chapter III (clinically relevant data), these regions were found within the active site of the WT As a compensatory mechanism, the presence of mutations causes the loss of these *EC* in the active site or the gain of new *EC* hubs. Importance of *EC* hubs in the active site was found to decrease in the presence of active site domain mutations S534N and I543V. On the other hand, except for M166V, the majority of distant mutations (C29R, Y186C, S534N, I543V and D949V) were distinguished by allosteric compensatory gain of *EC* hubs in the active site. To further understand the impact of remote mutation on *EC* hubs, we identified the key top 4% (*EC*) hubs that are susceptible to 5-FU and, to a lesser extent, the $Fe^{2+}{}_4S^{2-}{}_4$ clusters binding in the WT (Figure 4.12.1 and S4.2.4). Besides, this metric yielded no persistent hubs (Table 4.6.4).

The 3D mapping of these *EC* hubs in drug bound state revealed the clustering of hub ensembles within the active site (Monomer A & B) and part of $Fe^{2+}{}_4S^{2-}{}_4$ clusters (only in Monomer B). In the WT, the *EC* constituted part of the following hubs (Figure 4.12.2).:

- Monomer A: Phe100, Ser103, Ile104, Asn106, Asn108, Tyr109, Tyr110, Gly111, Ala112, Ala113, Lys114, Met115, Ile116, Ala159, Leu665-Leu667, Ile717, Val718, Ala721, Ala724, Ala729, Val732, Thr733, Ala734, Ala769, Ala771-Ala782 and Ile790-Leu929

- Monomer B: Thr595, Phe707, Ala708, Leu710, Ala721, Val732, Ala734, Gly740-Leu744, Leu775-Thr779, Ile781-Ile790, Phe805 and Ser808-Leu813.

Residue 112 – 159 forms the $Fe^{2+}_4S^{2-}_4$ clusters domain I, which is in charge of electron transport to the active site (130). The active site domain residues 525-847 coordinate inter-residue connections with FMN and 5-FU. Residue 849-928 forms the $Fe^{2+}_4S^{2-}_4$ clusters domain V residue for the stability and electron transfer to the active site. These hubs have either been completely lost, maintained, or formed in the presence of remote mutations. Systems D432N, R886C, and L993R showed a total loss of significance in *EC* hubs at the active site domain (FMN & 5-FU binding environment) when compared to the WT. Similar tendencies were observed with the S534N active site domain mutation (Chapter 3) but not with the others. Offer et al., reported enzymatic activities significantly higher than the native in mutant S534N in two studies (168, 176). Other studies have linked this mutation to decreased DPD activity, but its clinical significance remains unresolved (164). According to current evidence, heterozygous carriers of these variants have a 25% average decrease in DPD activity (164, 505). Based on this Chapter, the mutations S201R and S492L have been identified as potential DPD inhibitors (171). Drastic reduction of 12.5 – 25% in reference to the native were noted in R235W, R592Q, D342N and S492L (168) .Together, irrespective of the asymmetric monomer behavior, the *EC* results show that the $Fe^{2+}_4S^{2-}_4$ clusters domains and the active site domain are essential for DPD protein functionality.

In comparison to *EC,* a more uniform distribution of DRN values was observed for *KC* (Figures 4.13.1), similar findings were reported by Barozi et al (285). Furthermore, 15 key persistent hubs were discovered based on global top 4% calculations (Figure 4.13.1 and Figure S S4.2.4): Asn108, Thr110, Gly111, Leu155, Leu775, Asn777, Val778, Thr779, Ile781, Ala782, Ile790, Phe805, Ser808, Ala810, and Gln833 (Table S4.6.5). *KC* hubs, as reported earlier in the clinically significant mutated data, were an extension of *EC*, emphasizing the importance of not only active site *EC* values but also cofactors. They also demonstrated the asymmetric behaviour of each monomer, with a dominant number of hubs and a subdominant number of hubs monomer, as seen in all of our metrics. The 3D visualization of the key *KC* hubs communication network (Figure 4.13.2) identified three important central locations of these hubs (cofactors, iron clusters; and the 5-FU binding site).

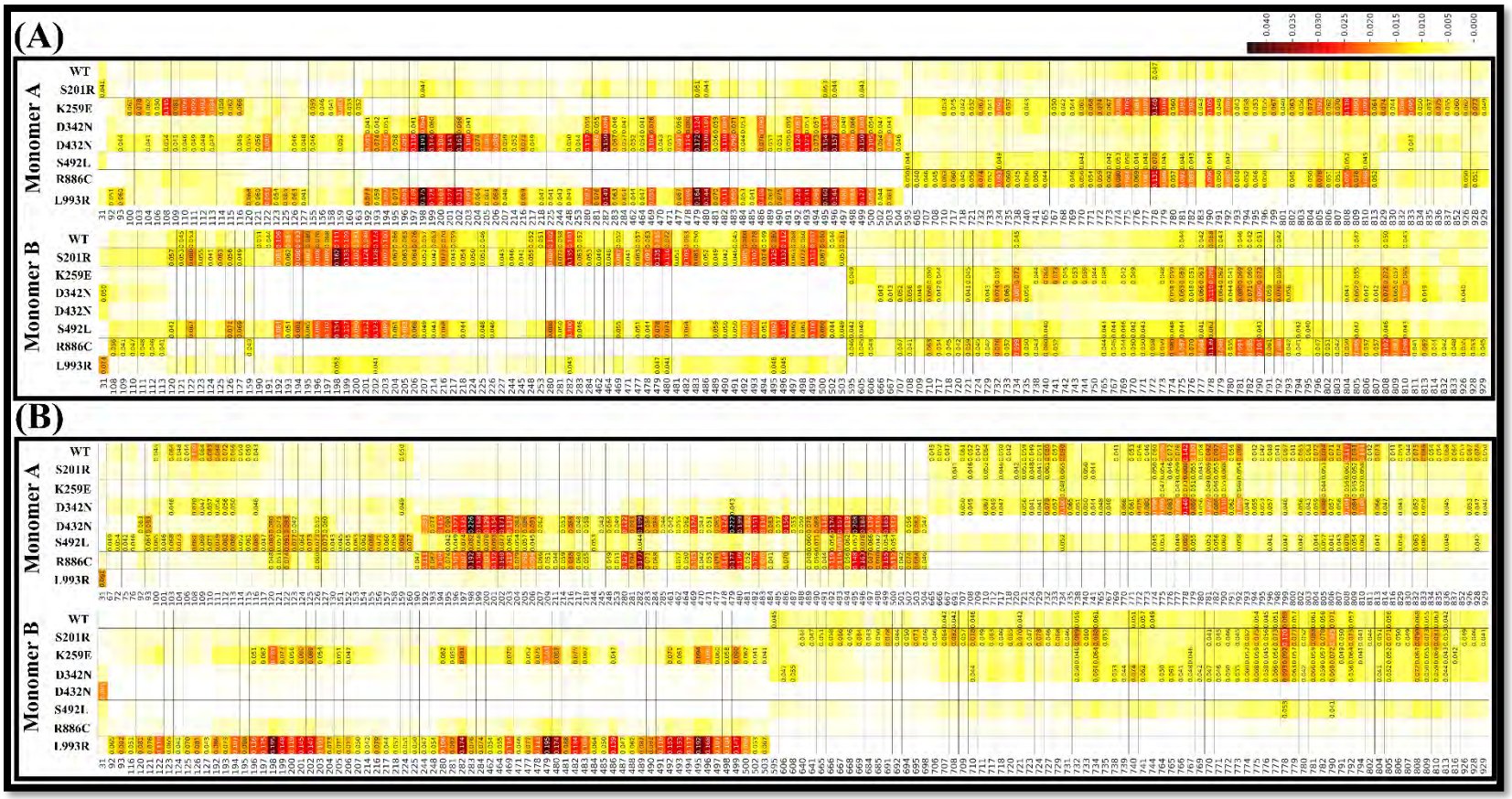


Figure 4.12.1: Heatmap depicting global top 4% key hubs for averaged *eigen centrality (EC)*. Heatmap of the (A) drug free (inactivated) and (B) drug bound (activated) ensembles. The x-axis represents protein residues, while the y-axis represents WT and the 7 mutation (remote) proteins. Centrality values of hubs are annotated with the value, with low to high centrality colored white to black, respectively. Barozi Victor derived the single heatmaps. The author, however, did the entire figure combination.

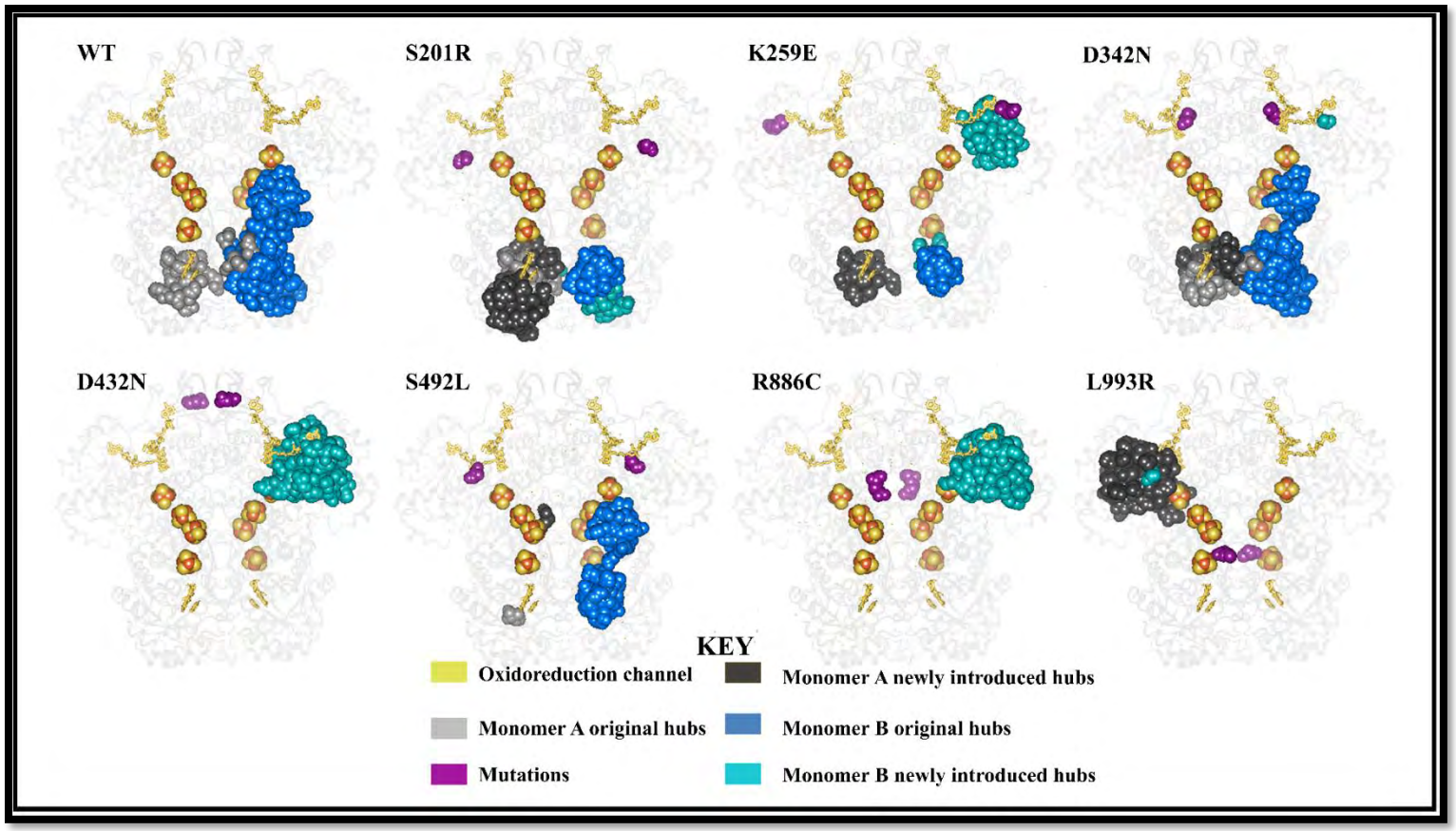


Figure 4.12.2: 3D cartoon representation of the distribution of *eigen centrality* hubs of the drug bound ensembles along the oxidoreduction catalytic channel. Monomer A WT *EC* hubs are denoted as light-grey, Monomer B WT *EC* hubs as blue. Mutation-induced newly introduced hubs shown in dark-grey in Monomer A and sky-blue in Monomer B. Oxidoreduction catalytic channel (NADPH, FAD, $Fe^{2+}{}_4S^{2-}{}_4$ clusters, FMN and 5FU) is shown in yellow. The mutations are represented in indigo.

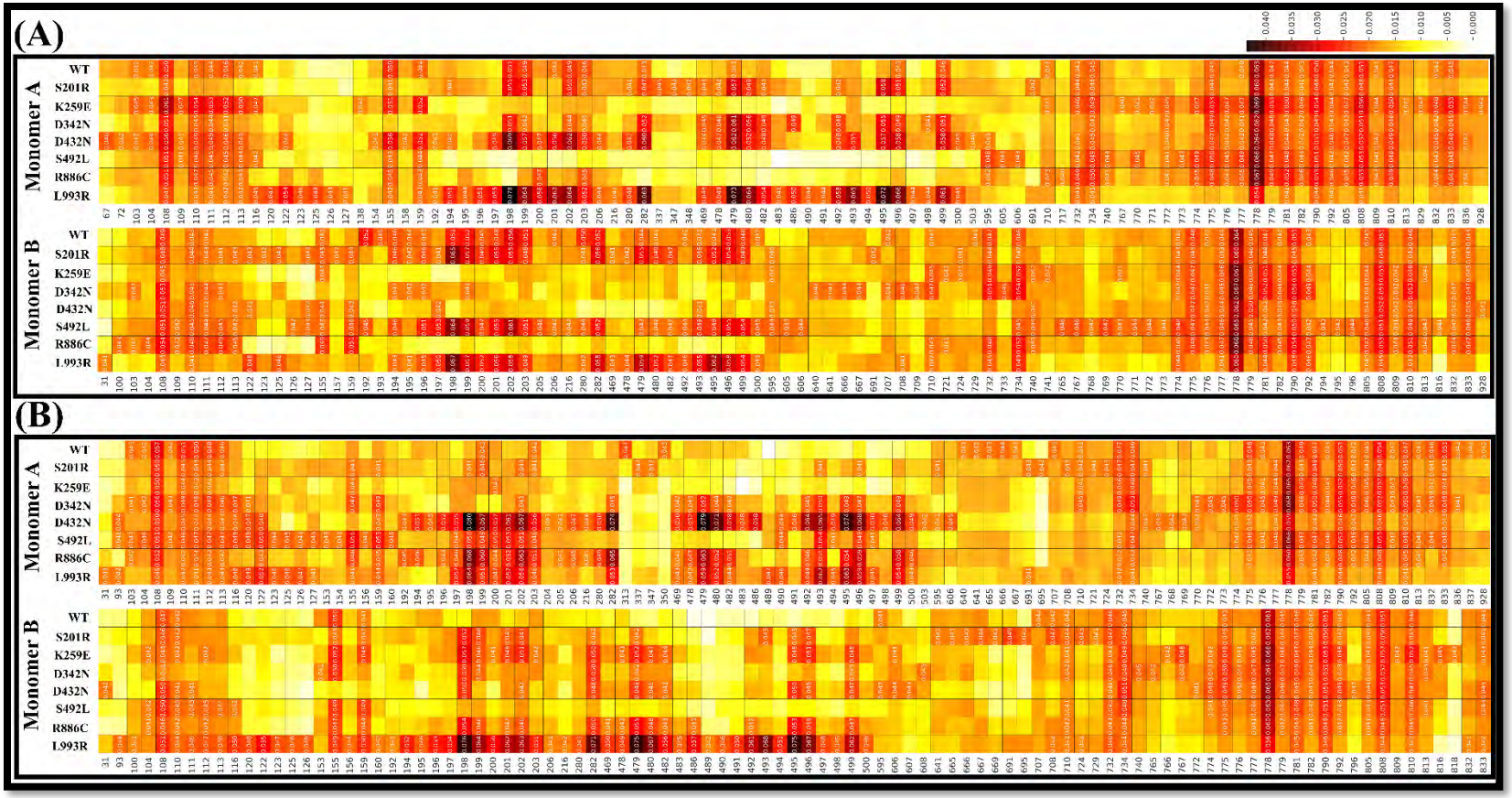


Figure 4.13.1: Heatmap depicting global top 4% key hubs for averaged *katz centrality (KC)*. Heatmap of the (A) drug free (inactivated) and (B) drug bound (activated) ensembles. The x-axis represents protein residues, while the y-axis represents WT and the 7 mutation (remote) proteins. Centrality values of hubs are annotated, with the value of low to high centrality colored white to black, respectively.

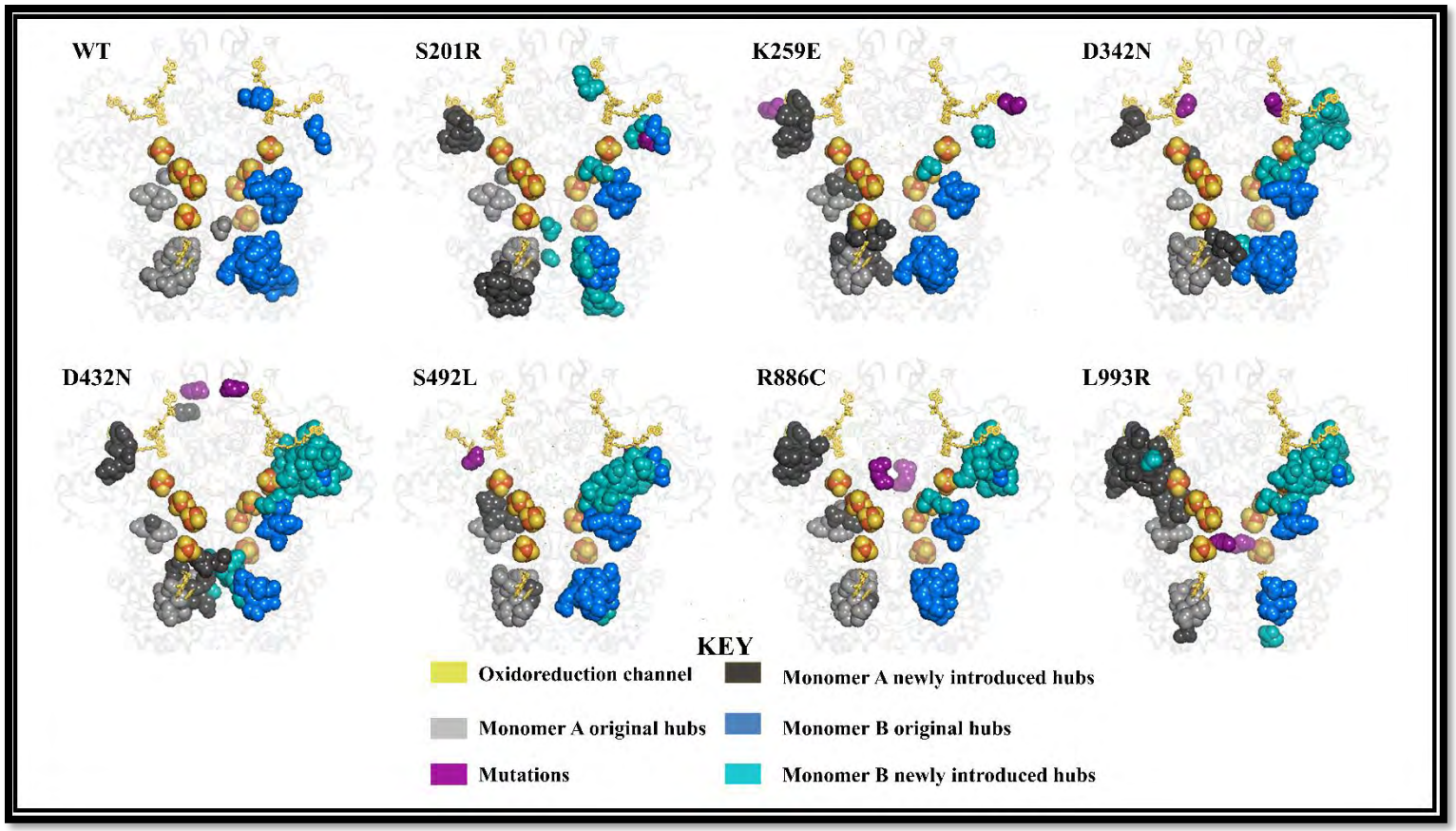


Figure 4.13.2: 3D cartoon representation of the distribution of *katz centrality* hubs of the drug bound ensembles along the oxidoreduction catalytic channel. Monomer A WT *KC* hubs are denoted as light-grey, Monomer B WT *KC* hubs as blue. Mutation-induced newly introduced hubs shown in dark-grey in Monomer A and sky-blue in Monomer B. Oxidoreduction catalytic channel (NADPH, FAD, $Fe^{2+}{}_4S^{2-}{}_4$ clusters, FMN and 5FU) is shown in yellow. The mutations are represented in indigo.

4.3.3.3 Allosteric hotspots and cold spots were determined using persistent hubs

Except for *EC*, the following persistent hubs were identified by the four metrics:

- *BC*: Leu155
- *CC:* Gly71, Ala72, Leu73, Ala74, Glu75, Ala76, Cys79, Leu80, Ala101, Ile104, Leu142, Ile150, Leu155, Pro598, Met599, Pro602, and Ser605
- *DC*: Asn108, Val778, Thr779, Ile781, Ile790, and Ser808
- *KC*: Asn108, Thr110, Gly111, Leu155, Leu775, Asn777, Val778, Thr779, Ile781, Ala782, Ile790, Phe805, Ser808, Ala810, and Gln833

These hubs (persistent) remained unchanged in the presence of mutations, first, clinically relevant (Chapter 3) and then remote (Chapter 4) mutations. The important protein regions that are not susceptible to mutations, according to Krishnamoorthy and colleagues, are referred to as mutational cold spots (468). Furthermore, persistent hub data has been proposed as a method for detecting cold spot residues (266). Figure 3.14 depicts 3D visualization of persistent hubs in the WT in detail. Allosteric cold spots were thought to be better drug design targets because they were less prone to mutation occupancy. In this case, most of the persistent hubs were centrally located within the protein; they involved both the C-terminal and N-terminal $Fe^{2+}{}_4S^{2}{}_{-4}$ clusters domain. This is in line with previous findings. Three major hubs in particular stood out among all the persistent hubs (N108, L155, and V778). V778 was found in the majority of the metrics, including *EC*, and had the highest centrality values, whereas L155 was found in four of them (*BC, CC, DC,* and *KC*). N108 was discovered in the vast majority of systems as well (*BC, DC,* and *KC*). Functionality specifics. Details of the functionality of these residues are in Chapter 3.

4.3.3.4 Mutation-induced impact on inter-residue interactions was revealed by weighted contact maps

The degree of residue interactions with immediate neighbors (within 6.7 Å Euclidean distance) was ranked from 0 to 1 in the mutated residues using the MDM-weighted TASK's contact map tools. This method has previously been used to investigate short-range communication in mutants (59, 266, 287, 345, 446). In our prior reports (Chapter 3), we used a similar approach to demonstrate short-range residue-residue communication in the presence of

mutant. Figure S4.3.1 to S4.3.7 revealed minor differences in residue contacts between monomers, which were attributed to local asymmetry in side chain conformations, which is common in complexes.(285, 449).

In comparison to the WT, Monomer A of S201R (Figure S4.3.1) demonstrated a loss of interaction with Leu122 and Phe205 (up to 2-folds), as well as a compensatory gain in H-bond with Asn120. Furthermore, contact with Leu122 was reduced by more than twofold, while interaction with Asn128, Pro197, Phe205, Gly495, and Lys497 increased by twofold. There was also a weak adaptive gain of the hydrogen bond with Ser500.

Similar findings were seen in K259E (Figure S4.3.2), where there was a more than 2-fold loss of interaction (Monomer A: Phe217, Gly218 & Ile256 and Monomer B: Phe217, Glu218, Lys219, Ile256 & Leu261) and adaptive gain of interaction (Monomer A: Phe217, Gly218 & Ile256 and Monomer B: Phe217,Glu218, Lys219, Ile256 & Leu261) in respect to the WT. In this mutation, the polarity of the charged residue which was initially positive (lysine) changes to negative (glutamate). The K259E mutation's polarity change may be responsible for the enzyme's partial (or complete) inactivation: As a result, the additional negatively charged residue is likely to cause a subtle shift of the loop backbone away from the β-sheet's hydrophobic environment (130). In turn, H-bonds from neighboring Leu261 and Ser262 are weakened, as are those from Gly258-N to Lys219-O (3.5 Å), which also plays an important role in cofactor binding by shielding the adenine-moiety of FAD from solvent access. Furthermore, Glu218 (which fixes the hydroxylgroup-O2* of FAD (2,7)) may be pulled away. As a result of the altered FAD-binding and orientation in the Glu259 monomer, the hydride transfer from the electron-donating cosubstrate NADPH, which is located adjacent to the FAD-flavine moiety, should be perturbed (180).

D342N (Figure S4.3.3) had the most changes in interaction at the point of mutation compared to the other systems, which was due to its proximity to NADPH coordinating areas (124). When compared to WT, there was a complete loss of interaction in Phe365, Asp346, Ala372, and Glu378 in Monomer A. As a compensatory measure, an increase in hydrogen bonding was observed in the Gly339, Ala340, Arg371, and Ala437 of both monomers. Differences in hydrogen bond gain (Ala340, Ala372, Gly439, and Asn487) and contact loss (Pro131, Asp134, and Ala372) were also observed in Monomer B between the WT and D342N.

In contrast to the WT, Monomer A of D432N (Figure S4.3.4) demonstrated a moderate increase in contacts (Gly302 and Arg332), whereas Monomer B demonstrated a loss of interactions (Gly333 and Val334) and a gain of interactions (Gly333 and Val334) (Arg332, Arg402, Ile403 and Val404).

When compared to the other systems, S492L (Figure S20.5) and D342N showed several differences in interaction within the point of mutation. This could be because it is close to FAD (Monomer A: 9.3 and Monomer B: 9.5). In several of the interacting residues, there was a general loss of connections (Monomer A: Gly479, Gly480, Asp481 & Thr488 and Monomer B: Gly479, Gly480, Asp481 & Asn494) and a gain of interactions (Monomer A: Gly196 & Gly496 and Monomer B: Val128 & Pro179).

R886C (Figure S4.3.6) showed a loss of interaction (Monomer A: Gly479, Gly480, Asp481 & Thr488 and Monomer B: Gly479, Asp481 & Asn494) and an increase or gain in comparison to WT (Monomer A: Gly196 & Gly496 and Monomer B: Val129 & Pro197). When comparing L993R (Figure S4.3.7) to the WT, there were fewer gains (Monomer A: Met1004 and Monomer B: Asp1000 & Cys1001) and losses (Monomer A: Cys959) of interaction.

The analysis of contact maps and heatmaps for persistent hubs (Tables S3.9.1–S3.9.3) revealed that the high centrality observed in the dimerization $Fe^{2+}{}_4S^{2-}{}_4$ clusters region is due to new/increased interaction between the dimerization domain and the hub and neighboring residues. Okeke and colleagues reported on the this (287). Most importantly, these findings emphasize the effects of mutations on structural networks, which affect protein interactions and communication patterns. Hence, the partial loss of DPD function in the presence of mutations (174, 445, 462).

4.4 Conclusion

Significant progress has been made toward precision oncology care thanks to a collaborative effort to improve our understanding of cancer by combining multiple data sets and algorithms (59, 507). Integrating *in vivo*, *in vitro*, and *in silico* data, as well as models, is critical for overcoming the inherent challenges of data complexity (508). Fundamentally, these approaches aid in the discovery of the molecular and biological mechanisms underlying drug metabolizing enzymes (509, 510). In this paper, we present a comprehensive synergistic

combined approach of all-atom MD simulation and post MD-trajectory analysis to decipher structural changes on the DPD protein caused by missense mutations. Post-MD simulation was conducted on the WT and seven non-clinically investigated remote DPD mutations seen in people of African descent (S201R, K259E, D342N, D432N, S492L, R886C, and L993R). To begin, known global analytic approaches (RMSD, RMSF, Rg, H-bond, and essential dynamics) were used to gain insight into mutation-induced protein dynamics.

The following are the study's specific significant findings (79): the RMSD, which included framewise RMSD and RMSF, revealed that the mutations increased the asymmetry of DPD monomers in comparison to the WT. This was due to the flexibility of the loop areas, which made up approximately 45% of the protein. Significant fluctuations were also observed in the active-loop region (9). RMSF and DCC data showed that there were very few fluctuations around the protein core in both mutants and WT. Similarly, our previous study on the WT revealed increased variation in the loop area but no change in the protein core (147). Notably, we discovered that mutations increased residual fluctuation within the active site but not in the WT. As a result, we conclude that these mutations, regardless of position, influenced DPD responsiveness to 5-FU via active site residue interaction. Similarly, top 4% averaged *EC* hubs 3D mapping uncovered that the majority of these hubs were located in active site regions extending to the $Fe^{2+}{}_4S^{2-}{}_4$ clusters in the WT. *EC* hubs were reduced in the active site area but increased in co-factor regions in mutation. This demonstrated that the mutations were losing relevance in the active site function of DPD (511). Besides that, we identified the frequency of 5-FU atomic contact across MD simulations based on its interaction frequency. Aside from the mutation-induced asymmetric behavior per monomer, the 5-FU contact residue showed significant loss and gain of interaction in relation to the WT. S670 and C671, in particular, revealed differences in interaction between monomers in each mutation, as well as loss of interactions in some cases (150). We presented data from various DRN metrics and how they define the topology of the residue interaction network. This helped researchers understand how mutations in the DPD protein affect the distribution of global mediating centers and allosteric signaling. Apart from previous DRN analyses in which a clearly defined single allosteric communication network pattern to the active site was obtained with either *BC* (287) or *EC* (283), *BC* and *DC* independently suggested allosteric communication signaling along the electron transport system (NADPH, FAD, $Fe^{2+}{}_4S^{2-}{}_4$ clusters, FMN to 5-FU). The two *BC* routes run

parallel to the electron transport pathways more internally, whereas the *DC* pathway runs parallel to the electron transport channels more externally. Both paths were represented by large hub ensembles, which was also observed in one of our previous studies (374). Some of the large hub clusters present in the WT system were lost or replaced in the mutant system, with the most impacted areas being the regions surrounding the cofactors (NADPH, FAD, $Fe^{2+}{}_4S^{2-}{}_4$ clusters, and FMN), as well as the active site region. As a result, allosteric compensatory communication to the active site is proposed. This explains the previous *in vitro* research on these mutations' partial loss of enzyme activity (168, 176). While the WT enables an efficient communication pathway via the optimal two communication pathways per monomer seen in *BC* and *DC*, the mutations may evoke a broader ensemble of suboptimal and less efficient compensatory allosteric routes via *BC* and *DC* as well. This results in DPD's dysfunctional metabolism of 5-FU (133). Our future research will focus on discovering new allosteric drug modulators to restore DPD protein function. To that end, we have identified a cold spot residue that will serve as a potential target area for pharmacological modulators. This will pave the way for individual tailored therapy (precision medicine) based on revealed missense mutations.

CHAPTER 5

Insight into the pharmacogenomics impact of catalytic domain mutation-induced effects on the dihydropyrimidine dehydrogenase protein structure and functionality, which causes the malfunction of 5-fluorouracil metabolism

This section presents non-clinically significant catalytic domain mutations in reference to *in vitro* and *ex vivo* studies. A hybrid of communication patterns depicts the aspect of cold spot protein regions, as well as the change in residual communication pattern within the protein catalytic and binding residues as influenced by missense mutations.

The first draft manuscript for this chapter is complete and has been presented to all parties involved for review before being submitted for publication.

Contribution: This research was performed under the supervision of Özlem Tastan Bishop and Thommas Mutemi Musyoka. **Maureen Bilinga Tendwa** conducted all global and local calculations, including the DRN calculations. However, the script for the identification of global top 4% high significant multi-metric (*BC, CC, DC, EC* and *KC*) DRN residues was generated by Olivier Sheikh Amamuddy (for another study); and DRN hub calculations were performed by Victor Barozi in this study. The paper draft was written by Maureen Bilinga Tendwa.

5.1 Introduction

Drug metabolizing enzymes have evolved over time, where a variety of functional groups work together to influence the active site microenvironments. These finely tuned groups create complex environments that allow for the catalysis of reactions with unprecedented efficiency and specificity (512, 513). They are generally conserved within protein families, suggesting their significance in substrate recognition and catalysis (514, 515). The optimal function of residues in these groups is centered on their invariable flexibility, which permits them to support necessary changes during the catalytic cycle (516). This flexibility usually comes at a cost because of locally introduced strain, which destabilizes the substrate and prearranges the active site for transitional-state stabilization (517-519). Active site residues, in specific, have been found to catalyze and bind drugs, as well as to stabilize their reaction intermediates. This results in the formation of sufficient points of contact for the drugs' strong binding (520, 521). Hence, the catalytic residues are critical in promoting drug binding and enzyme activity (521, 522). However, mutations of residues close to the active site disrupt the residues' drug binding affinity, interfering with drug binding (523). Over the last few decades, enzymology researchers' ability to obtain molecular structures of proteins and perform site-directed mutagenesis has provided valuable insights into how enzymes perform their functions at the atomic level (524-527). These studies frequently concentrated on the general function of catalytic residues, such as acid/base reaction, covalent, and electrostatic contributions to catalysis (528, 529). However, studies which systematically investigate the consequences of all active site residues indirectly involved in bond formation and breaking are uncommon (527, 530, 531). Previously thought to only be essential for substrate binding, these non-catalytic residues have been discovered to be as important for enzyme function as catalytic residues (522, 530, 531). In this study, we looked at the entire active site environment of Dihydropyrimidine dehydrogenase (DPD) and how active site domain mutations related to people of African ancestry affect its functionality. The DPD active site domain, also known as domain IV, is composed of FMN, residues 525 to 847, and 5-FU per monomer (129-131). The FMN is positioned on top of the domain's $\alpha_8\beta_8$-barrel and is close to the opposing monomer's 2 ($Fe^{2+}{}_4S^{2-}{}_4$) redox cofactors. As a result, the ping-pong transmission of two electrons across the dimer interface to 5-FU is enabled. This ligand is bound directly above

the FMN-isoalloxazine ring, parallel to the pyrimidine ring elements (Figure 40) (130). FAD's O_2 and N10 atoms are about 3.8 times farther apart than 5-FU's O_2 and O_4 atoms. There is a similar gap between the reactive centers of C_6 in 5-FU and the reactive centers of N_5 in FAD. The close proximity of the two allows for the final electron transfers that cause 5-FU degradation. Table 5.1 contains additional information on the interactions occurring within the active site (catalytic-site residues and binding-site residues) environment (129, 131, 151). Therefore, identifying an enzyme's functional groups, including active site residues, is a critical step in understanding its biological functions and the underlying catalytic mechanisms. These are the foundations of both fundamental research and drug design.

In light of our previous findings, we propose that missense mutations in the DPD enzyme's active site domain may play a significant role in 5-FU degradation, resulting in varying degrees of toxicity. Commonly described missense mutations in DPD enzymes, for example, have been linked to varying degrees of toxicity in the Caucasian population (162, 164, 166, 366). Consequently, preliminary screening for the DYPD gene, which encodes DPD, has recently begun in regions throughout Europe and the United Kingdom. The goal of screening is to adjust the dose in the presence of clinically significant missense mutations (163, 164, 166). While increased knowledge has resulted in lower toxicity in some subpopulations, this has not been emulated in people of African descent. Besides, it is unclear how missense mutations in the DPD active site domain affect 5-FU interactions within the active site.

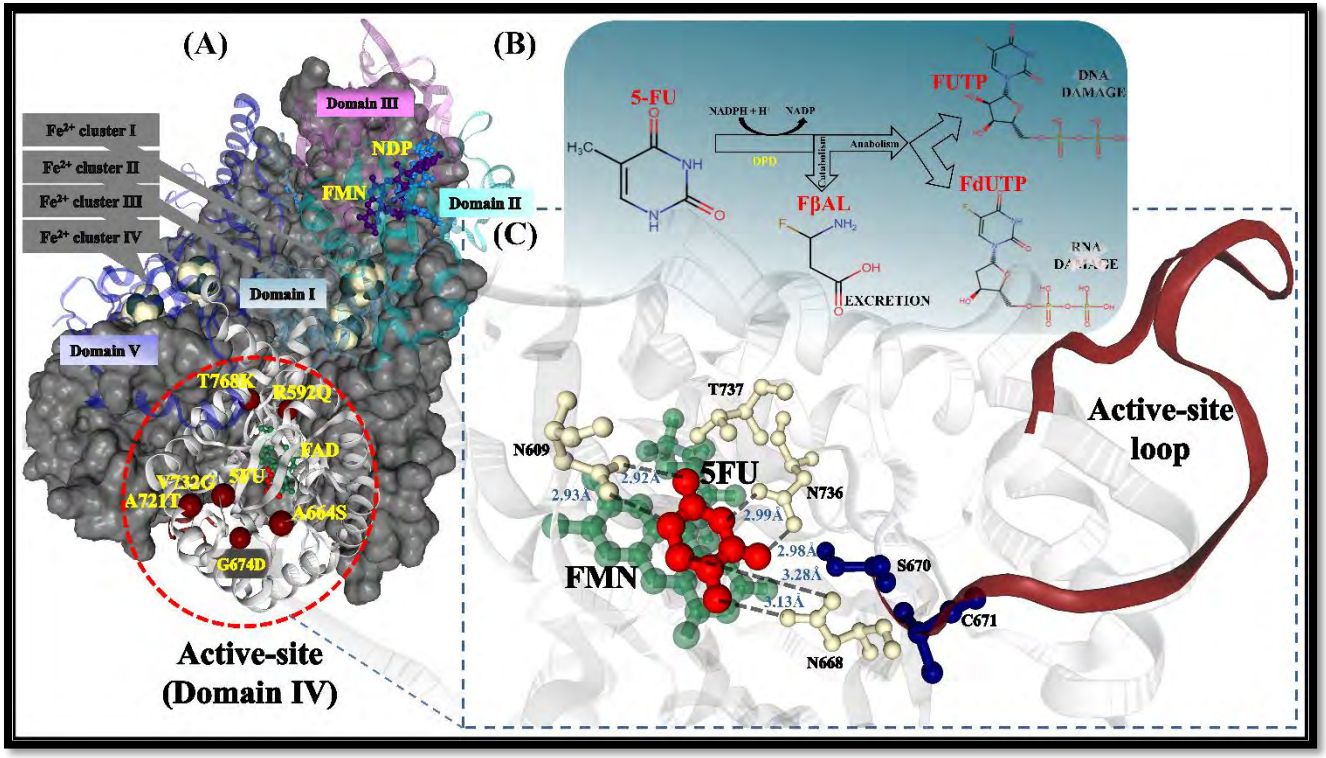


Figure 5.1: The 3D illustration of the dihydropyridine dehydrogenase (DPD) protein. A) Dimeric structure with Monomer A as a grey surface and Monomer B as a cartoon made up of domain I (teal), domain II (cyan), domain III (purple), domain IV (ivory), and domain V (yellow) (royal-blue). The position of six missense mutations shown in maroon, 5-FU medicine (red), FAD (green), and the active site loop define the active site domain (maroon). B) An overview of 5-FU catabolism and anabolism from $NADPH^{+}$ to NADP to 5-FU reduction. Catabolism of 5-FU to 5,6-dihydro-5-fluourouracil (DHFU) results in the synthesis of α-fluoro-β-alanine (FβAL), which is eliminated in urine. Anabolism of 5-FU produces three primary active metabolites: fluorouridine triphosphate (FUTP) which causes RNA damage; deoxythymidine triphosphate (dTTP) which causes DNA damage and integration of fluorodeoxyuridine triphosphate (FdUTP) instead of dTTP which also causes DNA damage. C) The active site residues that interact with 5-FU medicine, as well as the active-loop region (maroon), which adopts a closed conformation upon drug binding (activation) and an open conformation when inactive (drug free).

Table 5.1: An illustration of key residues within the active site of DPD (129-131)

Group	Residues	Function
5-FU binding	N609, N668 and	Interactions of three asparagine hydrogen bonds with 5-FU
5-FU binding	T737 and S670	Interactions between 5-FU side chains and hydrogen bonds S670 bridges residue L709 and N736 which engage in FMN and 5-FU binding respectively. Positioning C671 between 5-FU thiol group and C5 atom at a distance of ~3.3 Å
Active site loop (IVβ4-IVγd)	Residues 675 - 679	5-FU-free exists in open conformation 5-FU-bound exists in closed conformation
	C671 (catalytic	Vital residue catalytically
Active-loop adjacent residues	L669 and A683	Responsible for changes within the pyrimidine binding pocket
FMN-bonding	K709 and N736	Hydroxyl group of S670 form hydrogen bonds to K709 back-born carbonyl and N736 carboxymide nitrogen
5-FU binding	H673	Involved in 5-FU binding by van der Waals interactions
5-FU binding	Q611	Forms hydrogen bonds with H673 back-bone oxygen
Hydrophobic-core	M642, I613, L612 and V583	Form hydrophobic core behind 5-FU from M642, I613, L612 & V583 side chain cluster interaction with M675 & M680 side chain cluster
FMN-binding (loop anchor point)	K709	K790 back-bone oxygen anchors one loop end of S670 through hydrogen bond and 5-FU N736 side chain

Loop anchor points	H673, M675 and E677	H673 forms hydrogen bonds to E611 M675 forms hydrophobic interactions with non-loop residues E677 engages in hydrogen bond interaction with 935 and 936 H673 Stabilize closed active site loop conformation

The trajectories (600 ns) of the WT and six [(WT, R592Q, A664S, G674D, A721T, V732G, and T768K) missense mutations found in African subpopulations were subjected to a combination (global and local) of post-MD analysis. These global [RMSD, Framewise RMSD, RMSF, radius of gyration (Rg), dynamic cross-correlation (DCC), and principal component analysis (PCA)] and local [collective DRN multi-metrics, contact-map, contact heatmaps, center of mass (COM) distances, and Ligplots] investigations revealed mutation-induced changes in (a) DPD stability and flexibility, (b) DPD catalytic environment involving molecular recognition and catalytic mechanism, (c) DPD-drug complexes by studying conformation heterogeneity and dynamic catalysis, (d) the communication network within the active site, and (e) inter-residue interaction at the point of mutation and the active site as a whole. These changes were depicted at each monomer (A and B) as the protein shifted from inactivated closed (drug free) to activated open (drug bound) conformation across all mutation systems in comparison to the WT.

5.2 Material and methods

It is essential to note that the investigative approaches used in this chapter were similar to those used in clinical data. Refer to **Chapter 3.2**. However, in the present case, the following remote mutations were noted: R592Q, A664S, G674D, A721T, V732G, and T768K. To recap, the previously generated AMBER Fe^{2+} force field ion parameters for DPD protein (Chapter 2) were factored in prior to simulation. Each of the modeled mutations in drug free and drug bound were protonated before incorporation of iron parameters. Thereafter, the catalytic domain mutated model structures were subjected to a 600 ns MD simulation. Moreso, the effects of the mutations were investigated at the global and residue levels using post-MD analysis with the aforementioned combined approaches.

5.3 Results and discussion

5.3.1 Non-clinically relevant missense mutations associated with the African subpopulation and found within DPD active site domain were revealed

It should be noted that *In vitro, ex vivo* and *in silico* prediction are vital methods that can be used for estimating DPD phenotypes and functional analysis of identifying missense

mutations (133, 168, 176, 179, 451, 503, 532). Therefore, the three methods were used taken into consideration while identifying the active site domain IV related to non-clinically significant mutations. Using the NHLBI Exome Variant Server and Minor Allele Frequencies (ESP MAF) (74), high coverage African ADME Dataset-HAAD (75), the 1000 Genomes Project (76), Database of Single Nucleotide Polymorphisms (dbSNP) (77), ensemble (78), and UniProt, we identified mutations (1) associated with African subpopulations, and (2) located within the active site domain (33). In addition, data from PharmKGB (30), ClinVar (32), and consensus from *in silico* pathogenicity prediction data from MU, PhD-SNP, Proven, Polyphen-2 were incorporated in Variant Analysis PORtal (VAPOR) (Table S5.1) (29). This approach was similar to our previous chapters' (the clinically relevant and remote non-clinically significant missense mutations) investigations. In this chapter, six active site domain mutations (R592Q, A664S, G674D, A721T, V732G, and T768K) were obtained (Figure 5.1). Overall, these mutations have effects on protein structure that are consistent with clinically relevant effects (Table S1). both the Ensembl Variant Effect Predictor (EVE) and Variant Annotation Integrator (VARITY) programs for sequence-based variant effect prediction consistently identified R592Q, A664S, A721T and V732G as highly pathogenic, whereas G674D and T768K were considered less likely to be pathogenic (Figure S5.1.1 – S5.1.3).

In this study, we found G674D (Monomer A: 16.1 Å and Monomer B: 13.6 Å) to be the closest mutation to the active site compared to the other active site mutations [R592Q (Monomer A: 18.6 Å and Monomer B: 18.3 Å); A664S (Monomer A: 21.6 Å and Monomer B: 21.8 Å); T768K (Monomer A: 20.6 Å and Monomer B: 19.8 Å); A721T (Monomer A: 28.6 Å and Monomer B: 28.3 Å); and V732G (Monomer A: 30.7 Å and Monomer B: 30.5 Å)] (Table S5.2).

PART I

5.3.2 Local analysis

5.3.2.1 Weighted contact-maps reported the effect of mutations on inter-residue interaction at the mutation site

The growing interest in evaluating changes in residue networks to better understand protein function has resulted in the development of a number of tools that provide a reduced matrix representation of the structure (329, 533-535). In MD-TASK, a contact map tool was

implemented to determine residue contact frequencies around a single residue across MD trajectories, which was lacking in previous approaches. Interactions such as van der Waals, hydrogen bonds, and electrostatic bonds are used to infer contacts (329). Residues with loss, gain, and reduced interactions can be identified by assigning values ranging from 0 to 1. Contact maps have previously been used to identify significant interaction changes associated with ligand binding and protein mutations (270, 438, 448). A recent enhancement to MD-contact TASK's map functionality, as implemented in MDM-TASK-web, allows users to calculate contact frequencies for multiple residues of interest (284). Hence, a number of proteins inter residue interaction have been characterized using this approach (266, 285-287, 345, 374). Similarly, MDM-TASK-web (499) was used in previous Chapters (3.3.4 clinically linked mutation (including both remote and active site mutations) and 4.3.3.4 remote mutations) to assess the potential effects of remote mutations on the native inter-residue network. Using the equilibrated trajectory regions of the native and mutant proteins, we calculated the local contact frequencies around the mutated residues and catalytic residues. Samples for equilibrated trajectories were extracted from the same time frame of an equilibrated 30 ns section of trajectory each monomer on a protein system prior to drug exit. Significant changes were observed in both the catalytic (S534N and I543V) and remote mutations (clinically linked: C29R, M166V, Y186C, D949V and non-clinically linked: S201R, K259E, D342N, D432N, S492L, R886C, and L993R), in regard to local perturbation around the mutations.

In this section, we sought to decipher the immediate neighboring active site domain mutation-induced effects in activated DPD before delving into how these effects translate to the 5-FU binding site environment. Prior to conducting these analyses, we used Discovery Studio 19 (394) to create 3D models of each drug bound and drug free mutation from the previously homology modeled human DPD structure (147). To create each ensemble, each specified residue was mutated as needed. The 12 drug free and drug bound modeled structures were found to be native-like after verification with ERRAT (386), ProSA (210), and PROCHECK (209) (Table S5.3). They did, however, contain a number of residues whose protonation state could not be accounted for by classic MD simulation (266). As a result, prior to simulation, all of the modeled structures were protonated at pH 7.5 using the H++ webserver (390).

MD simulation (600 ns) was performed on eight (two WT and six mutations) proteins systems in their drug-free and drug-bound states. The contact map tool in MDM-weighted TASK

was then used to visualize the rate of proximal inter-residue interaction at the point of mutation. (283). This included mutation trajectories for the entire simulation period. Close inter-residue interaction was rated from 0 (absence) to 1 (presence) in 6.7 Euclidean distance (presence). According to Figure S5.1.1 – S5.1.6 visible differences in residue contact between monomers were observed, which is a common feature in complexes. This was believed to be due to side chain conformational asymmetry on a local scale (285).

R592Q has been predicted as potentially damaging by computational predicting tools (Table S5.1), and *in vitro* studies have revealed remarkable reductions this enzyme activity of 12.5-5% compared to the WT (174). Substitution of Arg for Glu in position 592 (R592Q) on a local scale showed loss of interaction (Monomer A: Thr594 and Monomer B: Thr594, Ser598 & Thr600) with a compensatory gain in H-bonds (Monomer A: Gly609, Gln604, Phe607 & Val995 and Monomer B: Gly601, Gly603, Gln604, Phe607 & Val995) (Figure S5.1.1).

Concerning A664S, at least three computational predicting tools postulated that the substitution of Ala for Ser would be likely harmful (Table S5.1). Figure S5.1.2 shows that Monomer A has a slight loss of interaction with Leu540 and no gain of interaction. Monomer B lost interaction by a factor of two around Leu540 and Phe707, while Ala662 and Phe706 gained interaction. There is limited experimental data on the effect of A664S on enzymatic function. However, mutations near the active site have been linked to decreased enzymatic function in prior studies on S534N and I543V (Chapter 3).

Similarly, despite the lack of clinical or experimental evidence on G674D, which is located in the active loop region (174), there was a strong census from computational predicting tools that these mutations were probable damaging (Table S5.1). In comparison to other mutations, G674D had the greatest loss and gain of new interactions, which may be attributed to its relative position in the active-loop area and it being the closest mutation to the 5-FU binding region (Table S5.2). In Monomer A, the substitution of Glu for Asp in position 674 resulted in the greatest loss (Leu612, Cys671, Gly678, Pro672, Ala668 & Phe935) and gain (Gly676, Glu677, Arg678, Gly679, Met680, Gly681, Leu682, Gly993 & Thr934) of inter-residue interaction rate. Monomer B, on the other hand, demonstrated the least loss (Pro672, Gly676, Glu677, Arg678, Leu682 & Thr934) and gain (Phe935 & Gly936) of interactions by a factor of two or more (Figure S5.1.3).

The substitution of Ala for Thr at position 721 (A721T) was predicted to be potentially harmful by computational predicting tools (Table S5.1), despite the limited experimental data. Figure S5.1.4 showed reduced (Monomer A: Leu710 and Lys725 and Monomer B: Ile717) and increased inter-residue contact (Monomer A: none and Monomer B: Leu710, Lys725 & Leu785) by at least 2-fold in mutation A721T .

The substitution of Val for Gly at position 732 was predicted by computational tools to be slightly harmful (Table S5.1). However, there is little experimental evidence for this mutation. In comparison to A721T, V732G showed a slight increase in loss (Monomer A: Leu710, Ala721, Ala724, Ala734 & Ile781) and gain (Monomer A: Lys709, Ala729, Pro789 & Leu791 and Monomer B: Phe707, Lys709, Pro789) of residue-residue contacts per monomer. (See Figure S5.1.5).

Ogura et al., (503), functional analysis of T768K revealed a faster decrease in T768K-related activity than the WT. Furthermore, this variant was also predicted to be potentially harmful (Table S5.1). The substitution of Thr for Lys at position 768 resulted in a loss (Monomer A: Ser596 & Asp797 and Monomer B: Gly795, Asp797, & Ser817) and gain (Monomer A: Ser596, Gly797, Ser605, & Ser817 and Monomer B: Thr594, Thr595, Ser596, Gly597, & Ser60) of inter-residue contact per monomer (Figure S5.1.6).

As we previously discovered (Chapter 3: 3.3.4 and Chapter 4: 4.3.3.1), the difference in inter-residue interactions on monomer A versus monomer B in the presence of mutation could be an indication of allosteric signaling. This is consistent with other studies. (266, 285-287).

5.3.2.2 Mutation-induced asymmetric behavior in 5-FU atomic interaction per monomer was discovered

The clinically relevant mutations found within the active site (S534N and I543V) demonstrated that the presence of mutation interferes with inter-residue interaction at 1) the point of mutation. 2) the active site binding environment (which includes the catalytic loop). The closed catalytic-loop (residues 670-682) conformation is critical for DPD protein catalytic activity.(129, 130). As a result, any change in its conformation will have an effect on protein function. Mandelker D, et al(536), discovered that oncogenic mutations in the catalytic domains of kinases increased activity by changing activation-loop conformations. However, the apparent effect on catalytic site structure varied depending on the mutations. Keeping this in mind, and

employing the MD-TASK-tools described above, the atomic interaction between 5-FU and DPD was investigated. This would demonstrate how mutation-induced changes at the point of mutation affected substrate binding environments (Figure 5.2). 5-FU binding environment include residues (Asn609, Glu611, Leu612, Asn668, Ser670, catalytic residue C671, Asn736 and Thr737) (130). In general, the main catalytic residue Cys671 has been observed in open conformation in the loop (residues 670 - 683) area of the drug free DPD (123, 124). When activated (drug bound), the loop closes, allowing Cys671 to be optimally placed for proton transfer. At opposite ends of the catalytic loop, two residues anchor Cys671. The original anchor point is formed by hydrogen bonds and FMN's backbone oxygen atom binding Lys709 and side chain linking residue N736, respectively, at the location of residue Ser670. The second anchor point is on residue His673, which is hydrogen-bonded to Glu611, Met675, and Glu677 (130, 131). Figure 4.7 in this section depicts the asymmetric behaviour of Monomer A and Monomer B in 5FU-protein interactions. The WT protein maintained high contact frequency with the following residues:

- Monomer A: Arg589, Ile591, Phe607, Leu608, Asn609, Ile610, Glu611, Leu612, Ile613, Met642, Leu667, Asn668, Leu669, Ser670, Cys671, Pro672, Ala683, Gly685, Lys709, Leu710, Thr711, Asn713, Thr735, Asn736, Thr737, Val738, Gly763, Asn737, Val738, Gly764, Val765 and Ser766 (32)
- Monomer B: Thr575, Val583, Arg589, Ile590, Ile591, Phe607, Leu609, Ile610, Glu611, Leu612, Ile613, Ser640, Ile641, Met642, Cys643, Ser644, Glu666, Leu667, Asn668, Leu669, Ser670, Cys671, Pro672, His673, Ala683, Cys684, Gly685, Gln685, Lys709, Leu710, Thr711, Pro712, Thr735, Asn736, Thr737, Val738, Gly764, Val765, Ser766 and Phe935 (40)

Inter-residue interactions differed the most in residues related to Leu612, Ser670, and catalytic residue Cys671. (Figure 5.2). Considering catalytic residue C671, in Monomer A, the native protein had high frequency contact with Met642, Ser670, Pro672 and Ala683. Mutant R592Q, V732G and T768K lost contact with Met642 and had reduced contact with Ser670, Pro672 and Ala683.

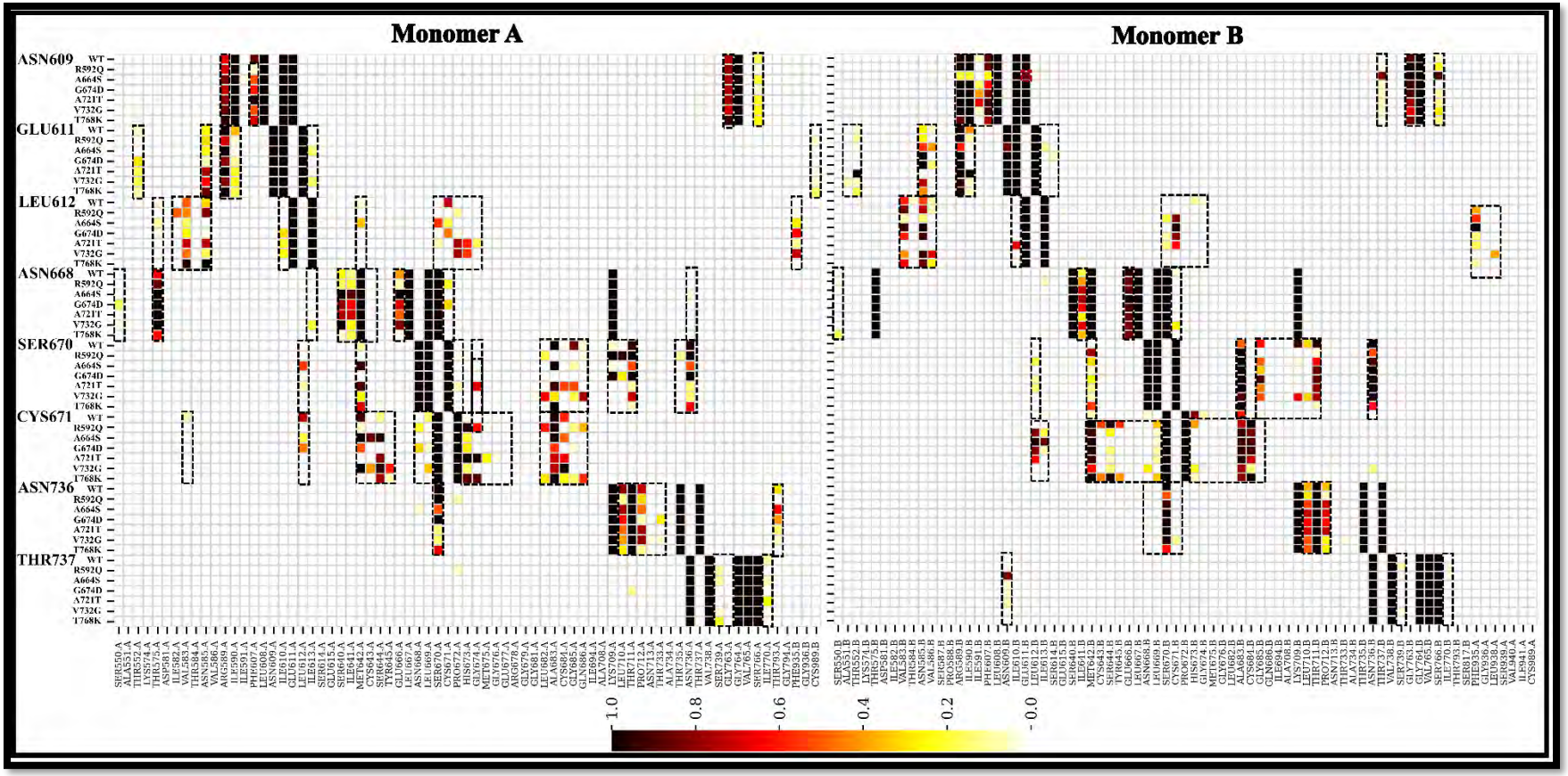


Figure 5.2: Activated DPD protein residue contact heatmap showing the interaction between Monomer A and Monomer B of 5-FU. The shaded boxes in each frame represent significant differences in the frequency of interaction between the cofactor residues and the surrounding contacting residues in the WT versus the mutation. White (0) to dark-red (1.0) represent no-contact (0) to strong-contact (1.0).

Conversely, compensatory gain of inter-residue interactions across all mutant systems were seen in Met642, His673, Gly674, Met675 and Glu677. Similarly, in Monomer B, the native protein had high frequency contact with Ser670, His673, Pro672 and Cys684. On the other hand, compensatory gain in residue interaction was noted in Ser644 and Cys684. The difference in 5FU-protein interactions between Monomer A and Monomer B could indicate allosteric signaling (285). Another example of allosteric signaling was observed at the Ser670 level in Monomer A, where the native had high frequency contact with Asn668, Leu669, Cys671, Ala683, Gly685, Thr711 and Asn736. These contacts were lost with Ala683 (R592Q & G674D) and Gly685 residues (R592Q, A664S, G674D and T768K). There was a compensatory gain of new residue-residue contact with Cys643 (G674D & V732G), Gly674 (A721T), Leu682 (V732G), Lys709 (R592Q & G674D), and Leu710 (R592Q & G674D) (R592Q). Ser670, on the other hand, lost interaction with Gly685 (R592Q, V732G, and T768K) and Lys709 in monomer B. (in all except, A721T). Across all systems, there was a compensatory gain in residue interactions with Met642.

The DPD mutations, according to these findings, may impeded 5-FU catalysis by reducing interaction with catalytic residues Ser670 and Cys671. However, as in our previous study, a compensatory mechanism was noted via contact gain to establish proper catalytic action (285). In a previous study, S534N and I543V mutations in the catalytic site with clinically relevant data showed a loss of H-bond in active site binding residues. This was linked to drug exit from one of their monomers, respectively. As a result, we conducted additional research to better understand the 5-FU binding residues. The drug bound structures were extracted from the lowest energy level of trajectories prior to drug exit and visualized with LigPlot+. (398) (Figure 5.3). With the exception of R592Q, A664S, and G674D Monomer A, the remaining systems retained 5-FU throughout the simulations. Monomer A lost hydrogen bond interactions with Ser670, Asn668, and Thr737 in R592Q, whereas Monomer B maintained all interactions as in the WT. Despite the loss of hydrogen interactions with Asn668, Ser670, Asn736, and Thr737 in Monomer A in A664S, a compensatory gain of H-bonds with Glu611 was noted. However, Ser670 was the only 5-FU binding residue in Monomer B that lost a hydrogen bond. Besides, in Monomer A and B, G674D and A721T lost hydrogen interactions with Ser670. Although V732G lost hydrogen interactions with Ser670 in Monomer A, T768K lost hydrogen interactions with Ser670 and Asn668 in Monomer B but gained compensatory H-bonds with Glu611. The

differences in DPD-5FU binding residues in Monomer A and B of each system suggested not only monomeric asymmetry, but also interference with 5-FU binding and thus dysfunctional metabolism. Furthermore, the partial loss of function could be attributed to the compensatory gain of residue-residue interactions, including H-bonds, in order for the protein to maintain its functionality. Carriers of S534N and I543V have shown partial loss of function, whereas R592Q revealed partial loss of enzymatic activities (176, 451).

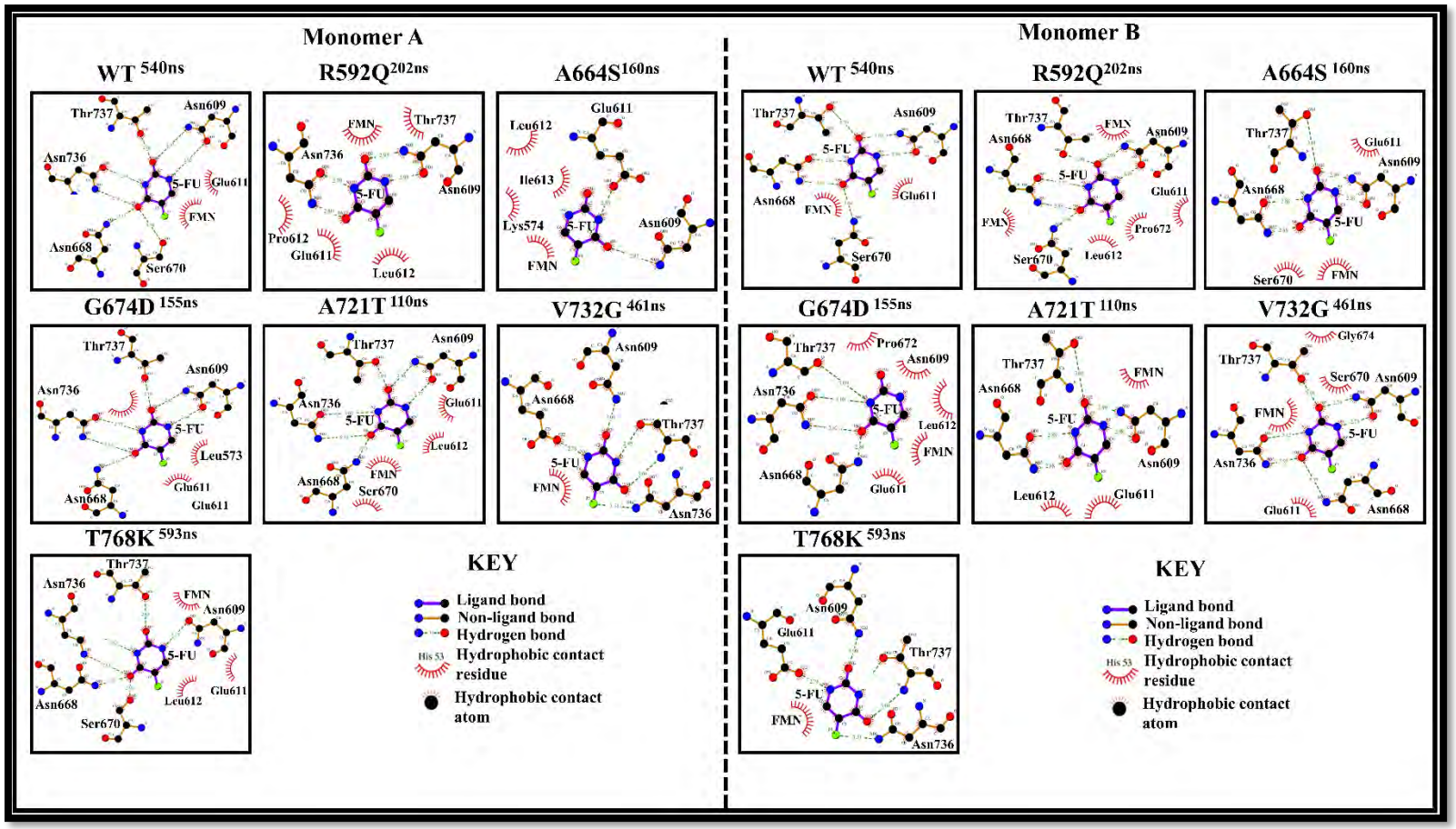


Figure 5.3: A 2D visualization of DPD-5FU binding residues of Monomer A and Monomer B. The structures were obtained from the lowest energy minima of the trajectories as indicated on top of each picture frame.

5.3.2.3 Aspects of protein communication network change in the presence of mutations as represented by the averaged DRN penta-metric approaches

Previously in Chapter 4 (clinically linked mutations) and Chapter 5 ('variants of uncertain significance), we illustrated how DPD remote mutations induced 5-FU hampered metabolism through allosteric pathways indicated by protein communication networks. Inter-residue interactions form the basis of these networks, with residues acting as nodes and interactions within nodes acting as edges (437). As stated earlier, the relevance of a node in a communication network is referred to as its centrality. Mutations cause reorganization of inter-residue network patterns, resulting in changes in node centrality (283, 285, 287). As a result, we sought to decipher the mutationally induced communication network that resulted in the upheaval observed in the active site environment. 30 ns equilibrated trajectories samples were extracted from the same time frame of both monomers of a protein system prior to drug exit for both drug free and drug bound systems (Table S5.5 and S5.6). Each sample produced the global top 5% of five different averaged centrality metrics, namely: *Betweenness centrality (BC), Closeness centrality (CC), Degree centrality (DC), Eigencentrality (EC),* and *Katz centrality (KC)* of the WT and mutations. The combination of the obtained global top 5% highest centrality nodes calculated across the related samples in each of these metrics was referred to as "hubs." Furthermore, due to the large size of the protein, we standardized these calculations by deriving the global top 2, 3, 4, and 5% of hubs for each metric. The 4% cutoff was chosen to include all hub residues while reducing data noise caused by the protein's size. Similar studies adopted the same method to determine communication patterns within proteins (18, 19, 73, 82). These hubs were discovered to correspond to regions around cofactors (NADPH, FAD, $Fe^{2+}{}_4S^{2-}{}_4$ clusters, FMN) or substrates, as well as the protein interface (Table S3.4). as elaborated in the metrics below:

***Betweenness centrality (BC)* reveals a mutationally induced communication pathway within the DPD oxidoreduction channel:** in determining the number of shortest paths across a node in a particular residue interaction network, *BC* denotes the frequency through which a node (residue) has been utilized (284). Thus, it represents communication routes connecting different protein sites in DRN. It has also been useful in studying allosteric phenomena (285), identifying major allosteric residues, as well as mutation and substrate-induced residue network shift in proteins (270, 283, 287, 345, 374).

Intersystem and intermonomer differences were highlighted in a heat map presentation of the top global 4% *BC* values, as shown in Fig 5.4:1. Sheik Amamuddy et al., discussed the typical asymmetric monomer behavior of a dimer, which was observed. (266). Besides, the following "persistent hubs" were identified: Leu142, Leu155, and Val373. These are the hub residues that remained unchanged from the reference WT in the presence of mutations for average BC calculations. Previously, "persistent hubs" were defined as hubs that remained unchanged across a set of comparative systems (266). As such, they are an implication of their functional significance. Leu142 domain I interface residues involved with $Fe^{2+}{}_4S^{2-}{}_4$ clusters interactions (130). Similarly, as explained in the preceding chapters of this study, Leu155 is a domain I helix that serves as a vital hub, connecting to the unique Fe^{2+} center binding residue Gln156. Another point of interest was domain III interface residue Val373 which is involved with NADPH interactions (130). Monomer B WT had the highest Val373 value (0.044), suggesting that after 5FU-binding, this hub assumes a greater role in information transfer throughout the network.

We mapped the global top 4% averaged *BC* hubs to the 3D DPD WT and mutation structures in drug free and drug bound states for better visualization and analysis (Figure 5.4.2 and Figure S5.2.1). Similar to our prior *BC* calculations (Figure 3.9.2 and Figure 4.9.2), a large hub cluster of communication pathways along the interior of oxidoreduction channel components from NADP to the reduction 5-FU was observed in both monomers. However, it was discovered that one path was more dominant in the number of hubs than the other. Mutant systems were gaining importance (form new connections not seen in reference WT) with either of these hub residues Tyr143 (Monomer A: A721T, V732G & T768K and Monomer B: R592Q & A664S); Ala144 (Monomer A: R592Q, A664S, A721T & T768K and Monomer B: A664S, V732G & T768K); Glu146 (Monomer A: All except V732G and Monomer B: All except G674D & T768K); Glu147 (Monomer A: All except A721T and Monomer B: All except V732G). Same tendencies were observed in the clinically relevant (Chapter 3) and non-clinically linked remote mutations (Chapter 4). Nevertheless, these *BC* hubs (Thr143, Ala144, Glu146, Glu147, Ile150 & Gly153) have been introduced around the $Fe^{2+}{}_4S^{2-}{}_4$ clusters in domain I. They have increased interaction with residue connecting Cys140 and Gln156 in comparison to the WT. Cys140 and Gln150 form disulfide bridges with $Fe^{2+}{}_4S^{2-}{}_4$ clusters, which participate in electron transfer to the active site (147).

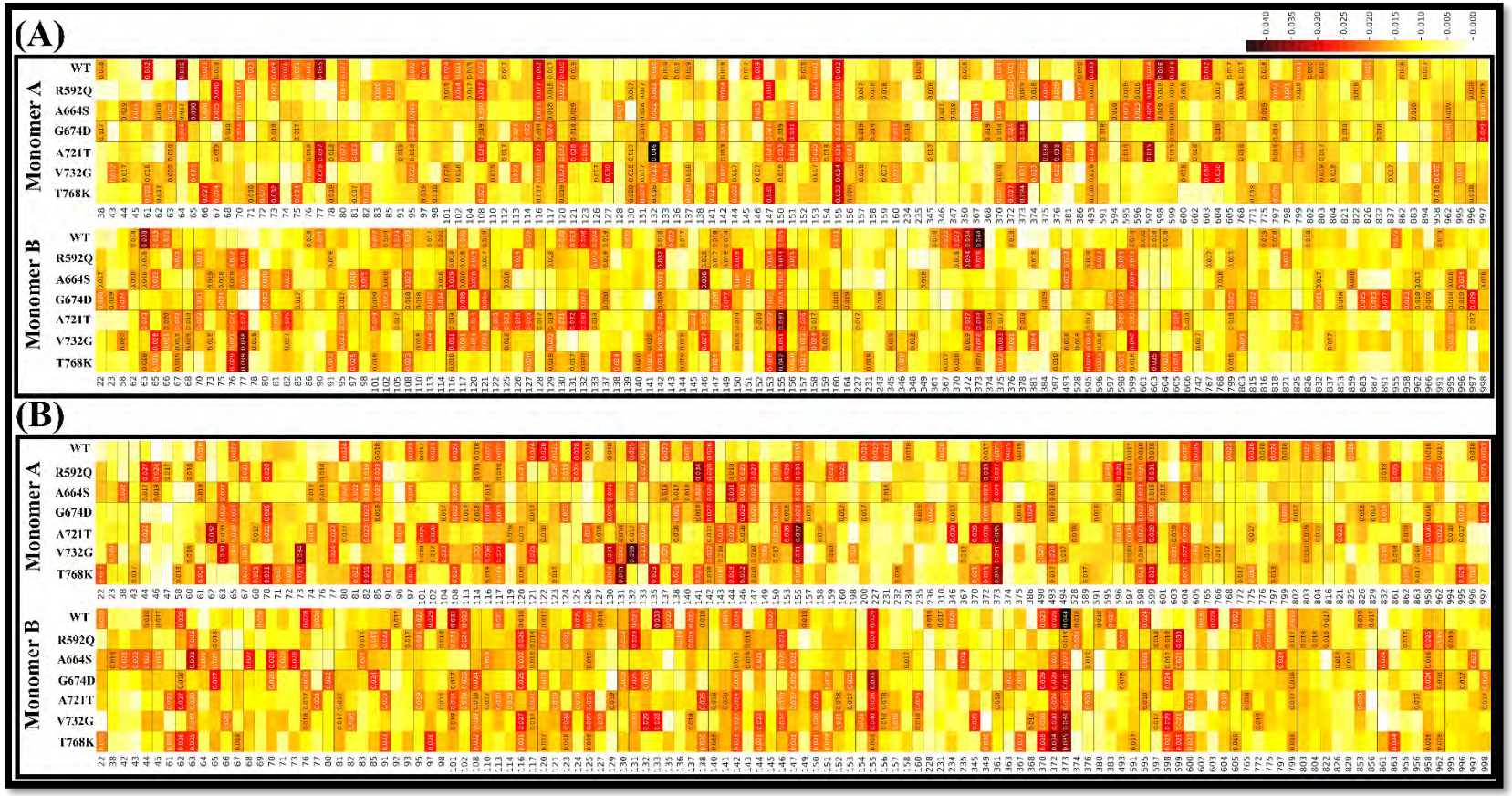


Figure 5.4.1: Heatmaps illustration of significant hubs based on the global top 4% of averaged *BC* metric. The x-axis depicts protein residues while y-axis depicts Monomer A and B WT and mutation proteins. The letters A and B stand for drug free and drug bound ensembles, respectively. The detected hubs have their centrality values annotated, with the values low to high centrality color coded from white to black, respectively.

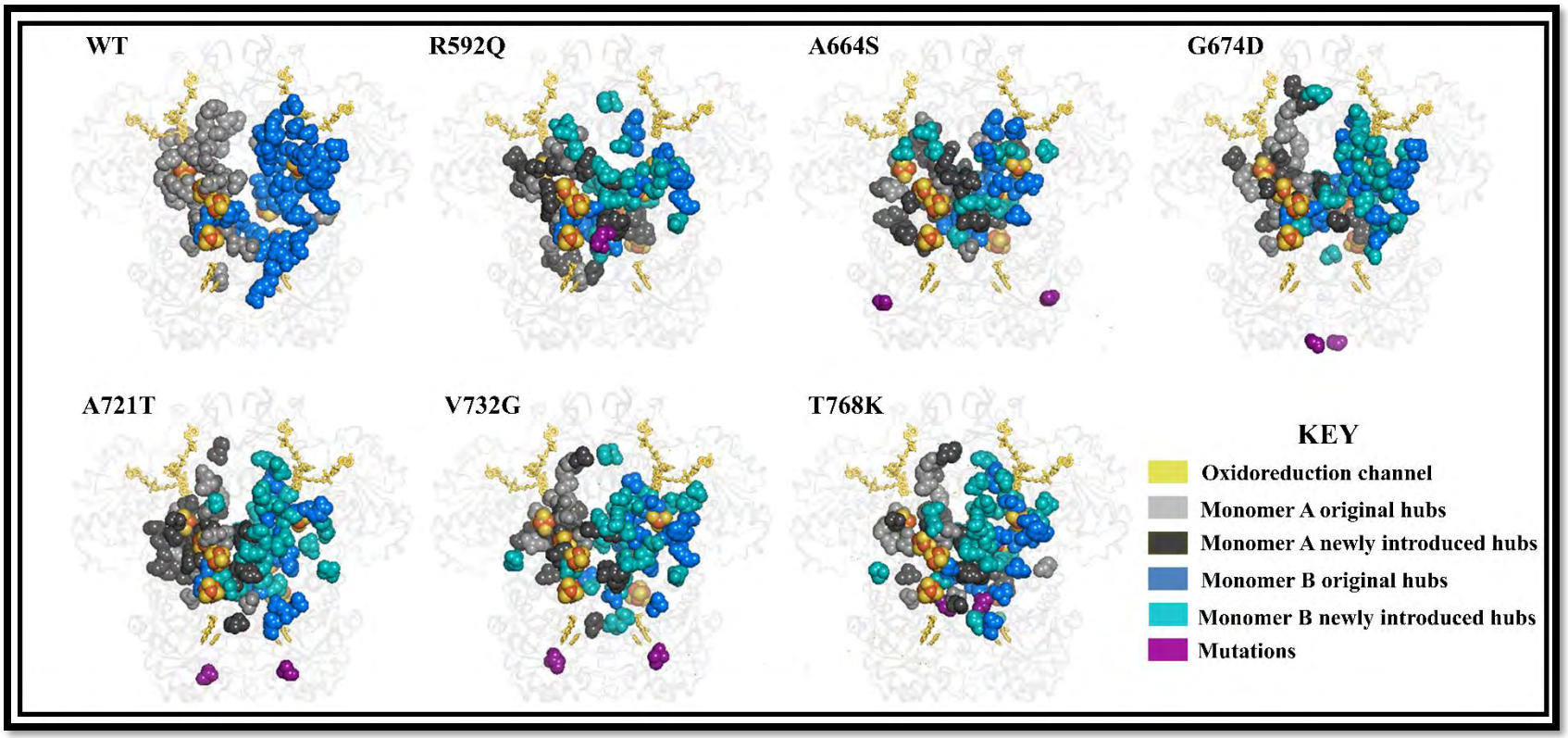


Figure 5.4.2: 3D mapping of the global top 4% for averaged *BC* metric of the WT and mutations in DPD drug bound ensembles. A communication pathway along oxidoreduction (gold) pathway from Monomer A NADPH to FAD to 2 x $Fe^{2+}{}_4S^{2-}{}_4$ clusters, proceeding to Monomer B 2 x $Fe^{2+}{}_4S^{2-}{}_4$ to FMN for the reduction of 5-FU. The hubs in Monomer A are colored light grey (original) and dark grey (newly introduced) and those in Monomer B are colored royal blue (original) and sky blue (newly introduced).

The introduction of these hubs could indicate a compensatory communication path to the active site as a result of mutation that interfere with the redox process. Additionally, most mutation lost contact with active site hub and FMN interacting residue Ser605 (Monomer A: A664S, G674D, A721T, and T768K and Monomer B: R592Q, A664S, G674D, A721T, and V732G), which could also interfere with the redox process.

***Degree of centrality (DC)* revealed a mutation-induced communication pathway outside of the DPD oxidoreduction channel**: unlike *BC, DC* defines the number of neighboring nodes that surround a given node. The heat map presentation of the data revealed that the monomers had slightly asymmetric behavior (Figure 5.5.1). A similar trend was observed not only in the *BC* (5.4.1) data in the current chapter but also in our *DC* (Chapter 3: Figure 3.5.1; Chapter 4: Figure 4.5.1) and *BC* (Chapter 3: Figure 3.5.2; Chapter 4: Figure 4.5.2) analysis from our previous chapter. Intriguingly, more persistent hubs (Asn108, Thr110, Leu155, Val778, Ile790, and Ser808) were discovered in *DC* than in *BC*, with the exception of Leu155, which was discovered in both. As witnessed earlier in our investigations, Asn108 and Thr110 were discovered to connect $Fe^{2+}{}_4S^{2-}{}_4$ clusters domains I and V, as well as FMN-binding domain IV. Val778, conversely, was discovered to be more important in mediating interactions within the active site, with the highest value (0.06) across all systems. Similarly, Ile790 and Ser108 interacted with FMN binding residues, suggesting that they played a significant role in FMN and active site stability in general.

The global 4% averaged *DC* hubs from drug free and drug bound systems were mapped to the 3D DPD WT and mutation structures for better visualization and analysis. In both the drug free and drug bound protein states, a pair of distinct clusters of hubs constituting a communication pathway along to the exterior of the oxidoreduction catalytic channel for each monomer was observed (Figure 5.5.2 and Figure S5.2.2). *BC* hubs displayed a similar pathway but more to the interior of the oxidoreduction channel components from NADP to 5-FU in both monomers. However, in each system it was discovered that one path was more dominant in the number of hubs than the other. The WT *DC* pathway consisted of the following Hubs (Tables 5.7.2):

- Monomer A: Ala72, Asn108, Tyr110, Ala113, Leu155, Gln157, Ser199, Ile200, Ala203, Gly282, Val313, Ala314, Ser335, 350, Ala351, Ala478, Gly480, Thr554, Ser605,

Leu628, Ile641, Leu653, Ser654, Glu666, Phe707, Ala729, Val732, Gly740, Val778, Thr779, Ile781, Ile790, His807, Ser808, Leu813, Gly833, Leu837 and Ala928.= 38

- Monomer B: Ala72, Asn108, Tyr110, Ala113, Gly138, Gly139, Gly153, Leu155, Ala159, Ala203, Ala207, Ile214, Gly225, Val241, Val337, 356, Ala478, Trp568, Thr595, Gly603, Val625, Leu653, Ser654, Ile694, Phe707, Ala729, Val732, Val778, Thr779, Ile781, Ile790, Ser808, Thr832, Gly833 and Ala928. = 35

In respect to the WT, mutant systems lost connection with Pro232 (Monomer A: all systems); Ser350 (Monomer A: all system); Ala351 (Monomer A: all systems), Gly603 (Monomer B: all system) and Ile694 (Monomer B: all systems). However, compensatory gain of hubs was observed around the:

i) FAD in A664S (Monomer A: Ala198, Ile200, Cys202 and Monomer B: Ala198, Ser199, Ala203), G674D (Monomer A: Gly194, Ala198, Cys202, Gly282 and Monomer B: Ala198, Ser199, Ser201, Cys202, Ala203, Ala207, Ile214, Glu218, Gly282) and T768K (Monomer A: Ala198, Ser201, Cys202 and Monomer B: Ala198, Ser199, Ala203, Leu206, Glu218)

ii) FMN in R592Q (Monomer B: Leu775, Leu813 & Cys816), A664S (Monomer A: Ile770 and Monomer B: Cys816), G674D (Monomer A: Leu813 and Monomer B: Leu813) and A721T (Monomer A: Leu775, Val815 and Monomer B:Leu775 & Cys816)

iii) NDP in R592Q (Monomer B: Ala486), A664S (Monomer A: Ile482, Ala486 & Ser493 and Monomer B: Asp495); T768K (Monomer A: Ile482, Ala486, Ser493 & Asp495 and Monomer B: Ile482, Val490, Ser493 and Asp495) and G674D (Monomer A: Ile482, Ala486, Val490 & Asp495 and Monomer B: Ile482, Ala486, Ser493 & Asp495)

iv) 5-FU in R592Q (Monomer A: Thr595, Val625 & Leu710 and Monomer B: Thr595, Leu710 & Gly740), G674D (Monomer A: Thr595, Val625 & Leu710 and Monomer B: Thr595) and T768K (Monomer A: Thr595 & Val625 and Monomer B: Thr595, Leu710, Ala729, Val732 & Gly740).

Overall, we demonstrated that: i) two distinct communication pathways per monomer were observed using both *BC* and *DC*. The *BC* pathway was located anterior to the oxidoreduction channel, while the *DC* pathway was located exterior to the oxidoreduction pathway. ii) The compensatory gain in *DC* hubs along the oxidoreduction pathway to the active site may point to mechanism for maintaining DPD functionality iii) In parallel with the clinically significant

mutation and remote VUS, we conducted a direct cross-correlation analysis to investigate the interplay between hub residues based on their *BC* and *DC* in both the drug-free and drug-bound states of the DPD enzyme. Across all mutations, a positive correlation was observed between *DC* and *BC* residue hubs per system. This correlation underscores the vital role played by these hub residues in maintaining the functional integrity of the network (Figure S5.3.3 and S5.3.4).

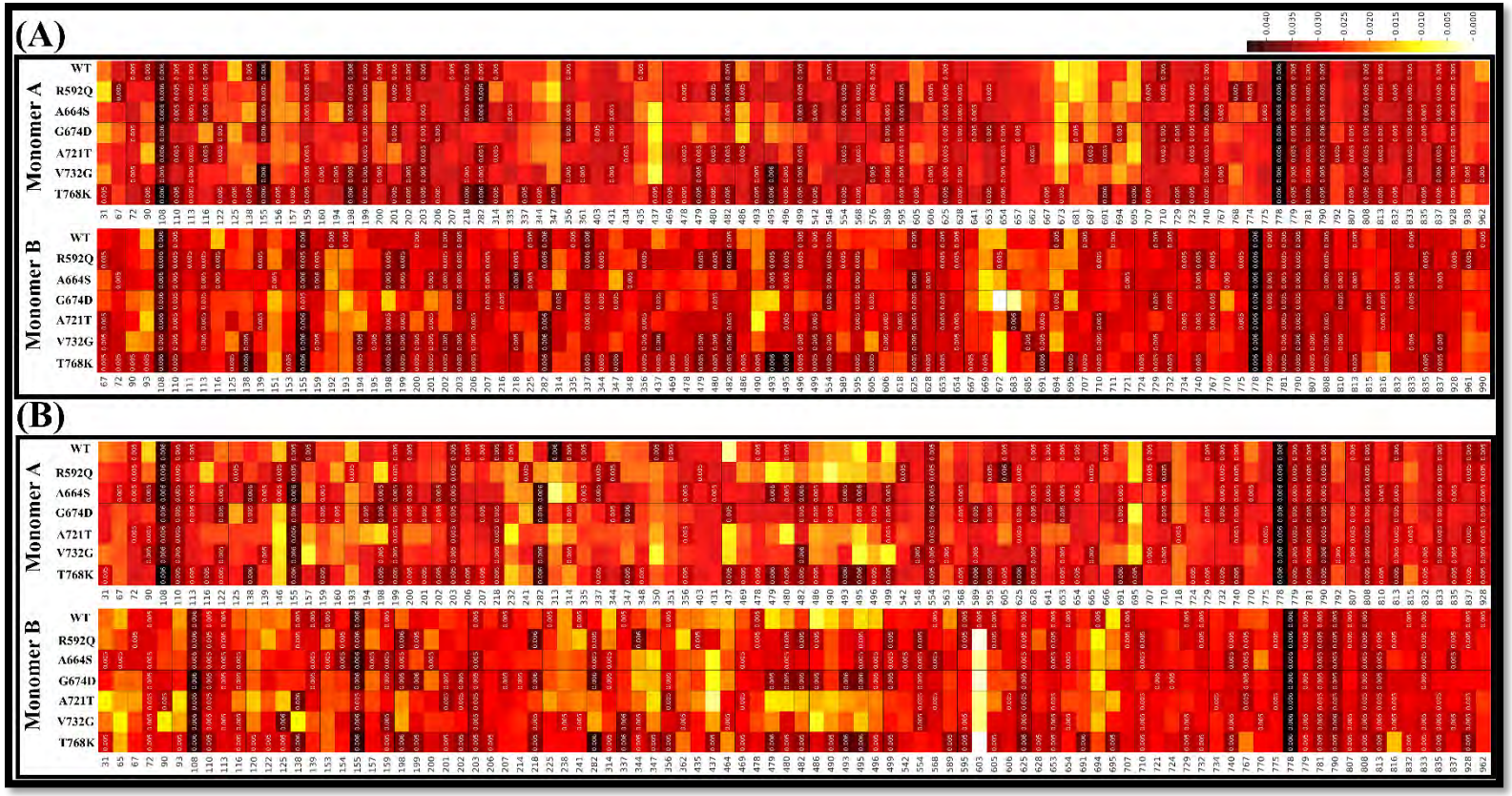


Figure 5.5.1: Heatmaps illustration of significant hubs based on the global top 4% for averaged *DC* metric. The x-axis depicts protein residues while y-axis depicts Monomer A and B WT and mutation proteins. The letters A and B stand for drug free and drug bound ensembles, respectively. The detected hubs have their centrality values annotated with the values low to high centrality color coded from white to black, respectively.

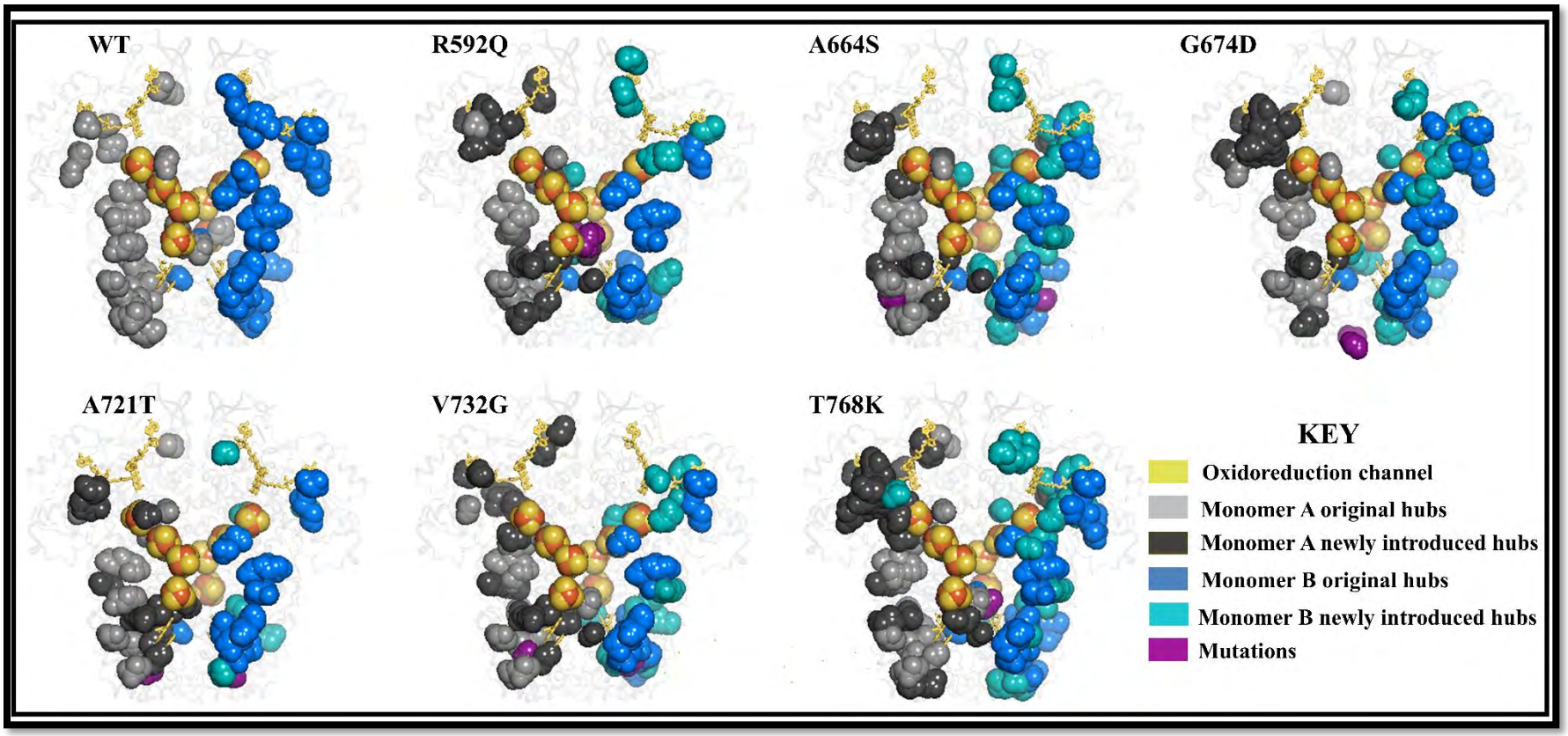


Figure 5.5.2: **3D mapping of the global top 4% for averaged *DC* metric of the WT and mutations in DPD drug bound ensembles.** A communication pathway along oxidoreduction (gold) pathway from Monomer A NADPH to FAD to 2 x $Fe^{2+}{}_4S^{2-}{}_4$ clusters, proceeding to Monomer B 2 x $Fe^{2+}{}_4S^{2-}{}_4$ to FMN for the reduction of 5-FU. The hubs in Monomer A are colored light grey (original) and dark grey (newly introduced) and Monomer B are colored royal blue (original) and sky blue (newly introduced).

An insight into *Closeness centrality (CC), Eigen centrality (EC),* and *Katz centrality (KC)* revealed a distinct regional communication pattern within DPD. Similar to our earlier investigations, the number of central nodes that are closest to other central nodes *(CC)* was first investigated, and thereafter we focused on high connectivity nodes which are surrounded by other high connectivity nodes centered on residue proximity *(EC).* Our subsequent research concentrated on *KC*, which is largely an extension of *EC.* Okeke and colleagues (287) discovered that averaged *CC* was the most restrictive in determining centrality, followed by *KC* and *DC*, which was consistent with our findings.

Also, comparable to our previous clinically significant analysis report, the heatmap depiction of averaged *CC* (Figure 5.6.1 and Table S5.7.3) disclosed that *CC* had the most persistent hubs in comparison to *BC, DC, KC,* and *EC* (Gly71, Ala72, Leu73, Ala74, Glu75, Ala76, Cys79, Leu80, Ala101, Ile104, Leu142, Ile150, Leu155, Pro598, Met599, Pro60). There are no proven mutations in this region, according to the literature. These territories which are not vulnerable to mutations are referred to as mutational cold spots by Krishnamoorthy et al., (442). Mapping the hubs to 3D protein structures demonstrated that all global top 4% averaged *CC* hubs are primarily located in the protein's center, particularly in the dimerization domain's $Fe^{2+}{}_4S^{2-}{}_4$ clusters region (Figure 5.6.2 and Figure S5.2.3). Thus, the relevance of $Fe^{2+}{}_4S^{2-}{}_4$ cluster domains in DPD enzyme functionality in relation to dysfunctional 5-FU degradation was demonstrated. The WT *CC* pathway consisted of the following Hubs (Tables 5.7.3):

- Monomer A: Glu71, Ala72, Glu75, Ala76, Cys79, Leu80, Ile101, Thr102, Ile104, Ser105, Val137, Cys140, Asn141, Leu142, Glu146, Glu147, Ile150, Asn151, Ile152, Leu155, Ser596, Gly597, Pro598, Met599, Tyr600, Gly601, Gln604 & Ser605.
- Monomer B:Leu67, Glu69, Arg70, Gly71, Ala72, Leu73, Arg74, Glu75, Ala76, Met77, Arg78, Leu80, Ile101, Ile104, Ser105, Val137, Cys140, Leu142, Thr145, Glu147, Ser149, Ile150, Asn151, Ile152, Leu155, Thr595, Ser596, Met599, Pro602, Gly603 & Ser605.

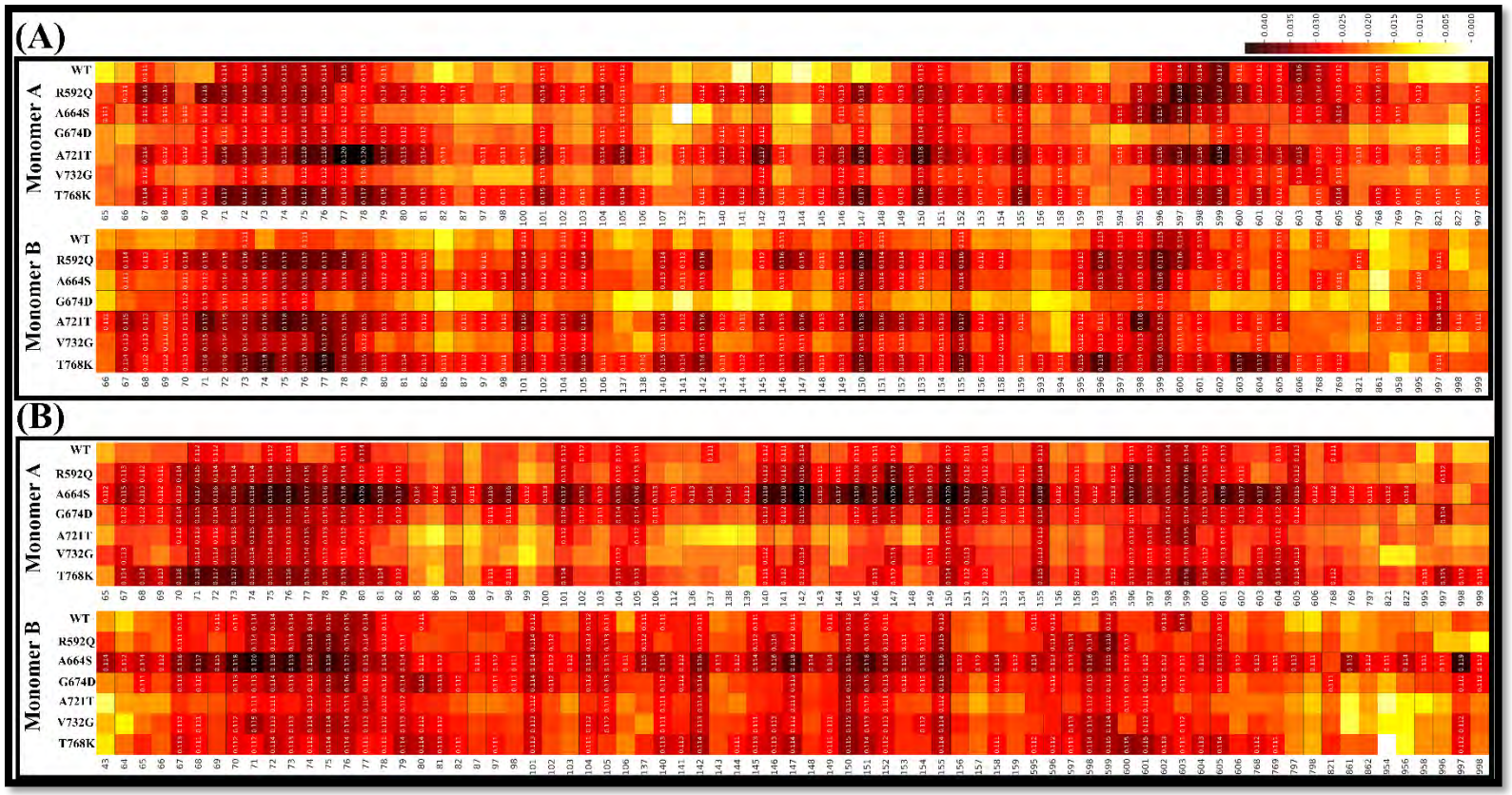


Figure 5.6.1: Heatmaps illustration of significant hubs based on the global top 4% of averaged *CC* metric. The x-axis depicts protein residues while y-axis depicts Monomer A and B WT and mutation proteins. The letters A and B stand for drug free and drug bound ensembles, respectively. The detected hubs have their centrality values annotated with the values low to high centrality color coded from white to black, respectively.

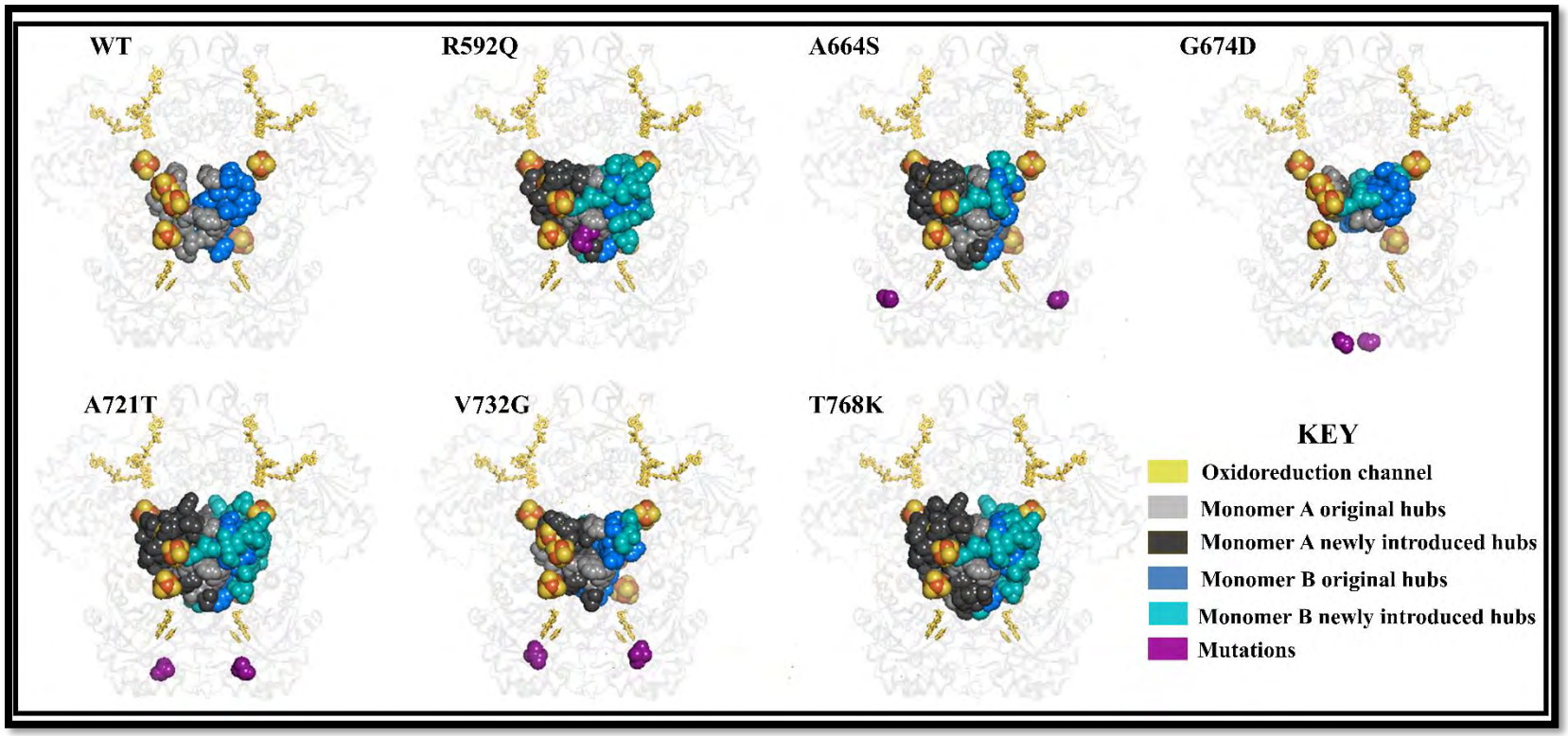


Figure 5.6.2: **3D mapping of the global top 4% for averaged *CC* metric of the WT and mutations in DPD drug bound ensembles.** A communication pathway along oxidoreduction (gold) pathway from Monomer A NADPH to FAD to 2 x $Fe^{2+}{}_4S^{2-}{}_4$ clusters, proceeding to Monomer B 2 x $Fe^{2+}{}_4S^{2-}{}_4$ to FMN for the reduction of 5-FU. The hubs in Monomer A are colored light grey (original) and dark grey (newly introduced) and those in Monomer A are colored royal blue (original) and sky blue (newly introduced).

Residues 70-80 and Val137-Leu155 are inter-connecting residue to $Fe^{2+}_4S^{2-}_4$ cluster in domain I whereby Cys79 and Cys140 form a disulfide bond with the cluster. Residues 595-605 are active site residues interconnecting with FMN and 5-FU (129, 130). Relative to the other ensembles, T768K had the most maintained top 4.0% high *CC* hubs (Monomer A: 28 and Monomer B: 29) with several newly gained high *CC* hubs (Monomer A: 19 and Monomer B: 16). Furthermore, R592Q (Monomer A: 26 and Monomer B: 23), A664S (Monomer A: 19 and Monomer B: 32), G674D (Monomer A: 27 and Monomer B: 26), A721T (Monomer A: 15 and Monomer B: 17), and V732G (Monomer A: 21 and Monomer B: 22) also retained *CC* hubs. Conversely, R592Q (Monomer A: 19 and Monomer B: 8), A664S (Monomer A: 45 and Monomer B: 42), G674D (Monomer A: 21 and Monomer B: 18), A721T (Monomer A: 6 and Monomer B: 4), V732G (Monomer A: 7 and Monomer B: 10), and T768K (Monomer A: 19 and Monomer B: 16) gained new *CC* hubs. The majority of *CC* hubs were gained in A664S and G674D, two systems that exhibited drug exit in Monomer A during simulation. G674S has been computationally predicted to have total loss of function; however, there has been little experimental research on this mutant (168). As a result, additional research is required to back up this prediction. As seen in other metrics, the differences in the clusters of high *CC* hubs found at the protein core of both monomers ($Fe^{2+}_4S^{2-}_4$ clusters) depicted the asymmetric behavior of the monomers. Similar to our earlier findings, *EC* centrality identified the key top 4% *(EC)* hubs that were prone to 5-FU and, to a lesser extent, the $Fe^{2+}_4S^{2-}_4$ clusters binding in the WT (Figure 5.7.1). This validates the findings of Foutch et al., (506), who discovered that sites for vital biological functions in residues could be identified using *EC* hubs. Herein, this metric did not produce persistent hubs. The 3D mapping of these *EC* hubs demonstrated that most mutations are losing hubs communication around the active site indicating a total loss of significance in *EC* hubs at the 5-FU binding domain (Figure 5.7.2 and Table S5.7.4). In reference to the WT, systems D432N, R886C, and L993R showed a total loss of significance in *EC* hubs at the 5-FU binding domain. Hence, *eigen centrality (EC)* showed mutation-induced disruption within the DPD protein active site.

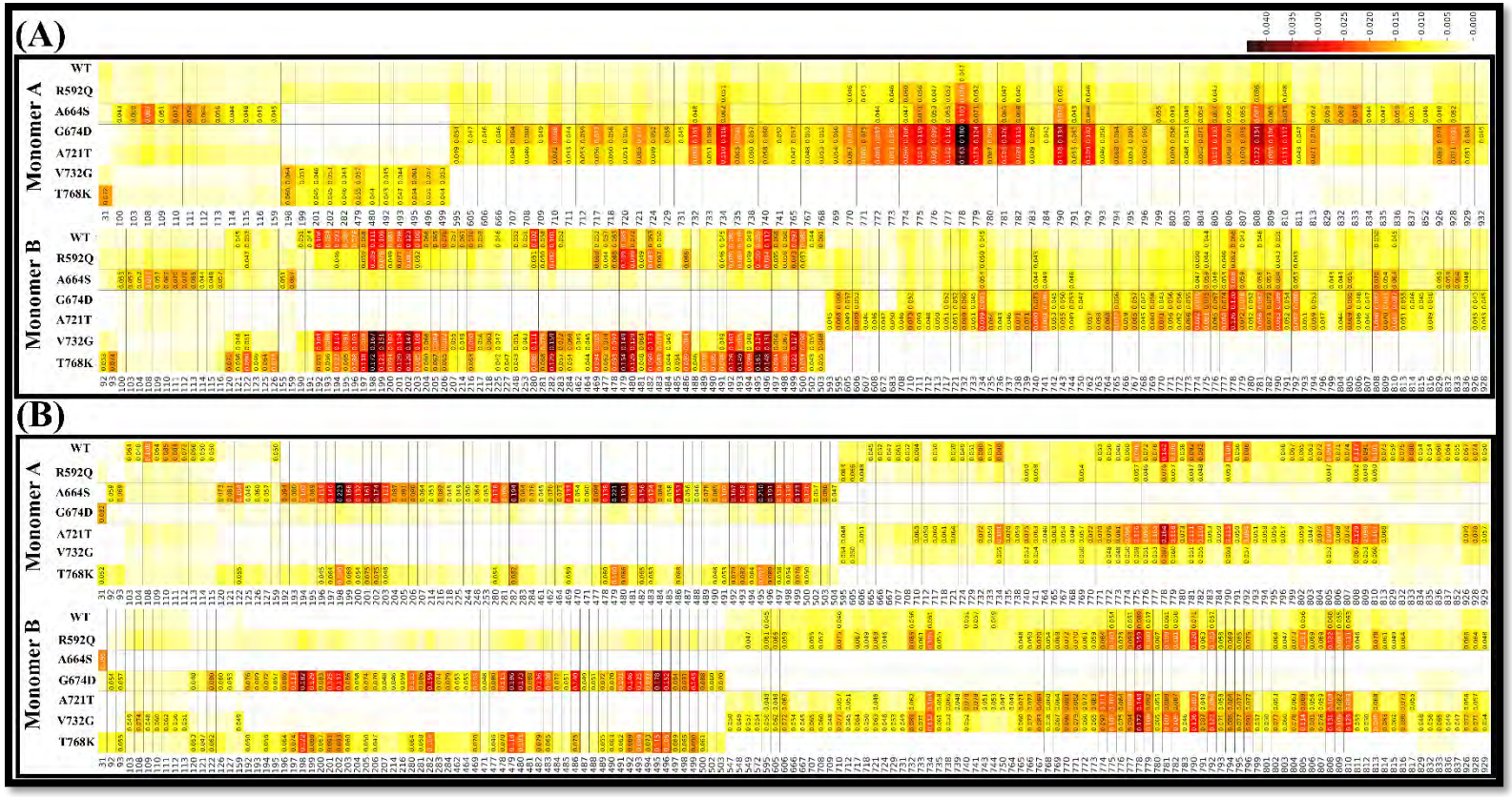


Figure 5.7.1: Heatmaps illustration of significant hubs based on the global top 4% of averaged *EC* metric. The x-axis depicts protein residues while y-axis depicts Monomer A and B WT and mutation proteins. The letters A and B stand for drug free and drug bound ensembles, respectively. The detected hubs have their centrality values annotated with the values low to high centrality color coded from white to black, respectively.

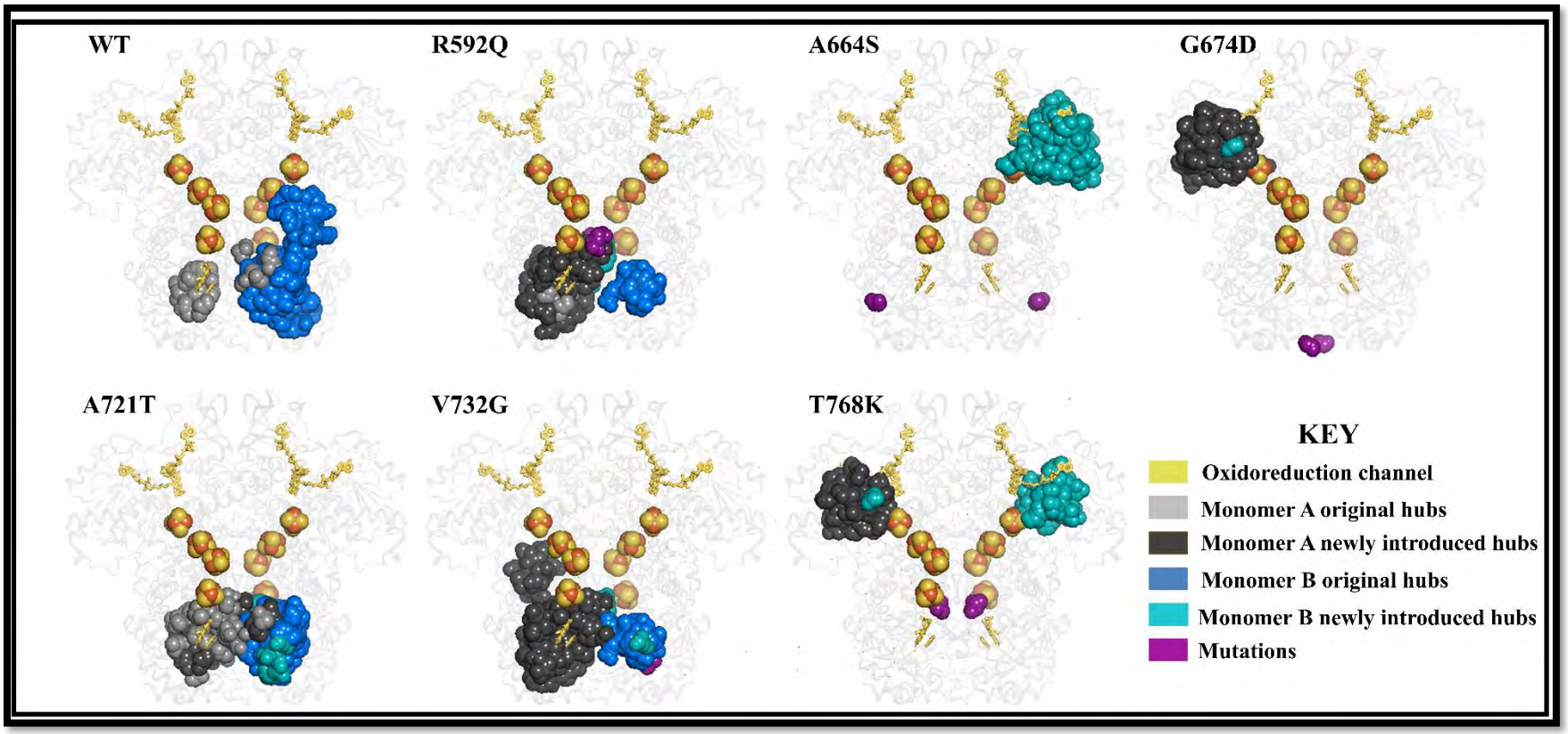


Figure 5.7.2: **3D mapping of the global top 4% for averaged *EC* metric of the WT and mutations in DPD drug bound ensembles.** A communication pathway along oxidoreduction (gold) pathway from Monomer A NADPH to FAD to 2 x $Fe^{2+}{}_4S^{2-}{}_4$ clusters, proceeding to Monomer B 2 x $Fe^{2+}{}_4S^{2-}{}_4$ to FMN for the reduction of 5-FU. The hubs in Monomer A are colored light grey (original) and dark grey (newly introduced) and Monomer A are colored royal blue (original) and sky blue (newly introduced).

Figures 5.8.1 and S5.2.5 displays a heatmap of the top 4% *KC* hubs, with 15 key persistent hubs identified (Asn108, Tyr110, Gly111, Leu155, Leu775, Ala777, Val778, Tyr779, Ile781, Ala782, Ile790, Phe805, Ser808, Ala810, and Gly833) as the protein transitioned from an inactive to an activated state (Table S5.7.5). The clinically significant mutated proteins had similar outcomes. In comparison to the WT, drug bound T768K Monomer A and Monomer B had more newly formed hubs than the other mutation systems. However, Monomer A and B of drug bound A664S had some of the fewest newly acquired *KC* hubs (7 hubs), as did Monomer B of V732G (6 hubs). Newly formed hubs have been implicated in protein compensatory mechanisms in mutations, which differ from the original residue-residue communication network observed in the WT. *KC* hubs were an extension of *EC*, emphasizing not only the importance of active site-like *EC* values but also cofactors, as discovered earlier in the clinically significant mutated data. Besides, they demonstrated each monomer's asymmetric behavior, with a dominant and subdominant monomer, as seen in all metrics. The 3D visualization of the key *KC* hubs network system (Figure 5.7.2 and Figure S5.2.5) revealed three critical central locations for these hubs (iron clusters; FMN and the 5-FU binding site). All the persistent hubs were centered on the iron clusters and active site domains. Thus, *Katz centrality (KC)* revealed a mutation-induced disruption within the DPD protein active site and cofactors.

In both drug-bound systems of clinically significant variants and variants of unknown significance (remote and active-site), an intriguing observation was made regarding the presence of centrality hubs. These hubs include averaged betweenness centrality (BC) hubs in Monomer A (Ser227, Ile231, Leu310, Pro374, Thr768, Cys816, and Thr825) and Monomer B (Glu69, Leu135, Thr228, Ile231, Phe234, Glu383, and Thr856). Additionally, averaged degree centrality (DC) hubs were detected in Monomer A (Pro232, Val313, Ser350, Ala351, and Ile641) and Monomer B (Val241 and Gly603) exclusively in the wild-type (WT) systems, but not in the mutant systems. Notably, there were no newly formed hubs in relation to averaged BC; however, a DC-specific hub, Gly282, emerged in the mutants but not in the WT. Similarly, Leu67, Ser149, and Gly597 were identified as mutant-specific averaged closeness centrality (CC) hubs. Some of these hubs, such as Leu135, Ser227, Ile231, Pro232, Gly228, Phe234, Val241, Leu310, and Val313, formed a communication network surrounding the flavin adenine dinucleotide (FAD), while Pro374, Glu383, Ser350, and Ala351 constituted a communication network around the nicotinamide adenine dinucleotide phosphate (NDP). These components play a critical role in

initiating electron transfer to the active-site through the oxidoreduction channel. Additionally, the hub Glu603 plays a crucial role in facilitating communication among the catalytic residues. The mutant-specific hub Gly282 is an integral part of the FMN-protein connecting hubs, while Gly597 is instrumental in enabling communication around the catalytic site. Therefore, the enhanced communication involving these hubs, which is absent in the wild-type (WT) system, may suggest the presence of an allosteric compensatory mechanism.

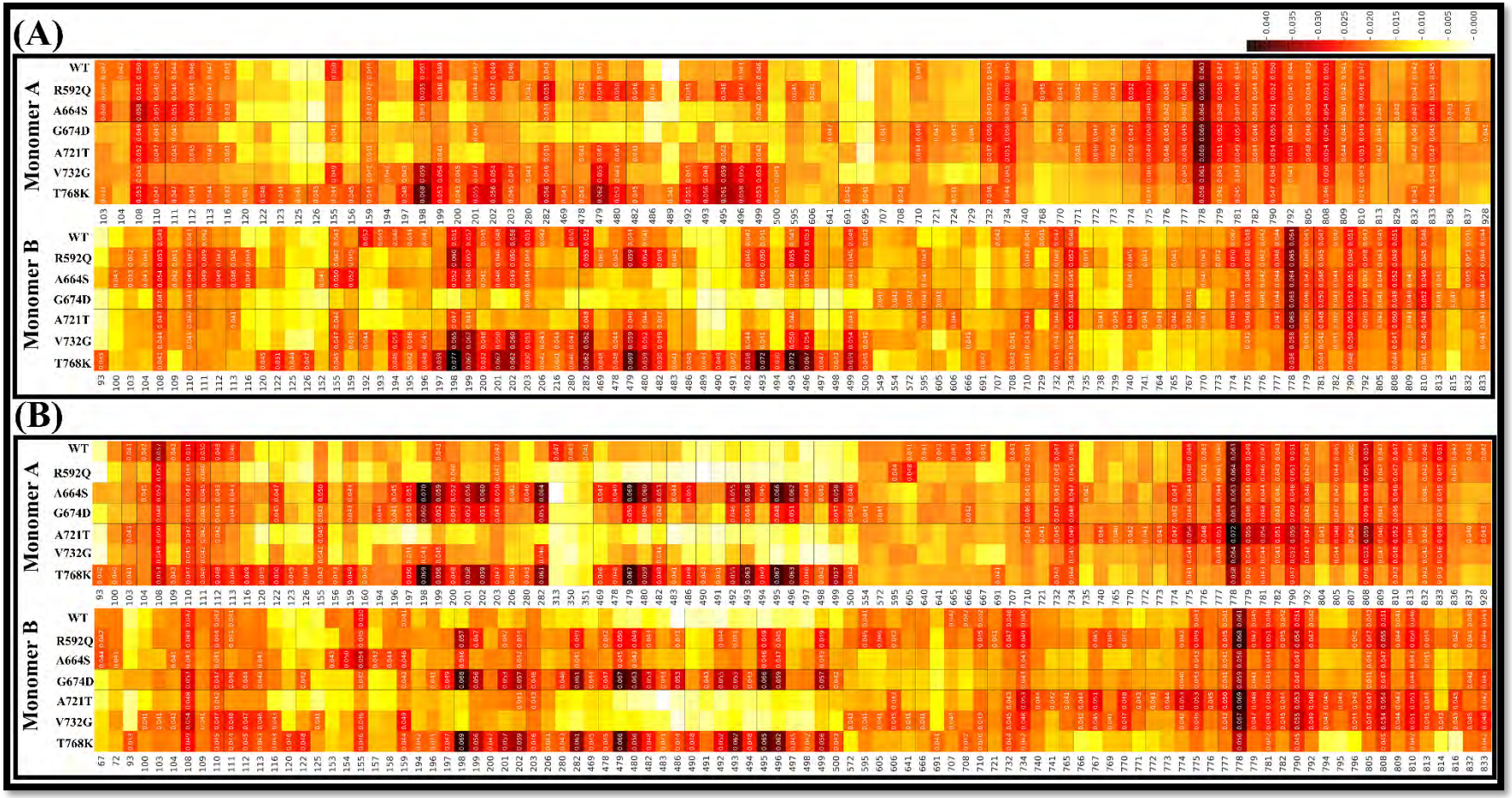


Figure 5.8.1: Heatmaps illustration of significant hubs based on the global top 4% for averaged *KC* metric. The x-axis depicts protein residues while y-axis depicts Monomer A and B WT and mutation proteins. The letters A and B stand for drug free and drug bound ensembles, respectively. The detected hubs have their centrality values annotated with the values low to high centrality color coded from white to black, respectively.

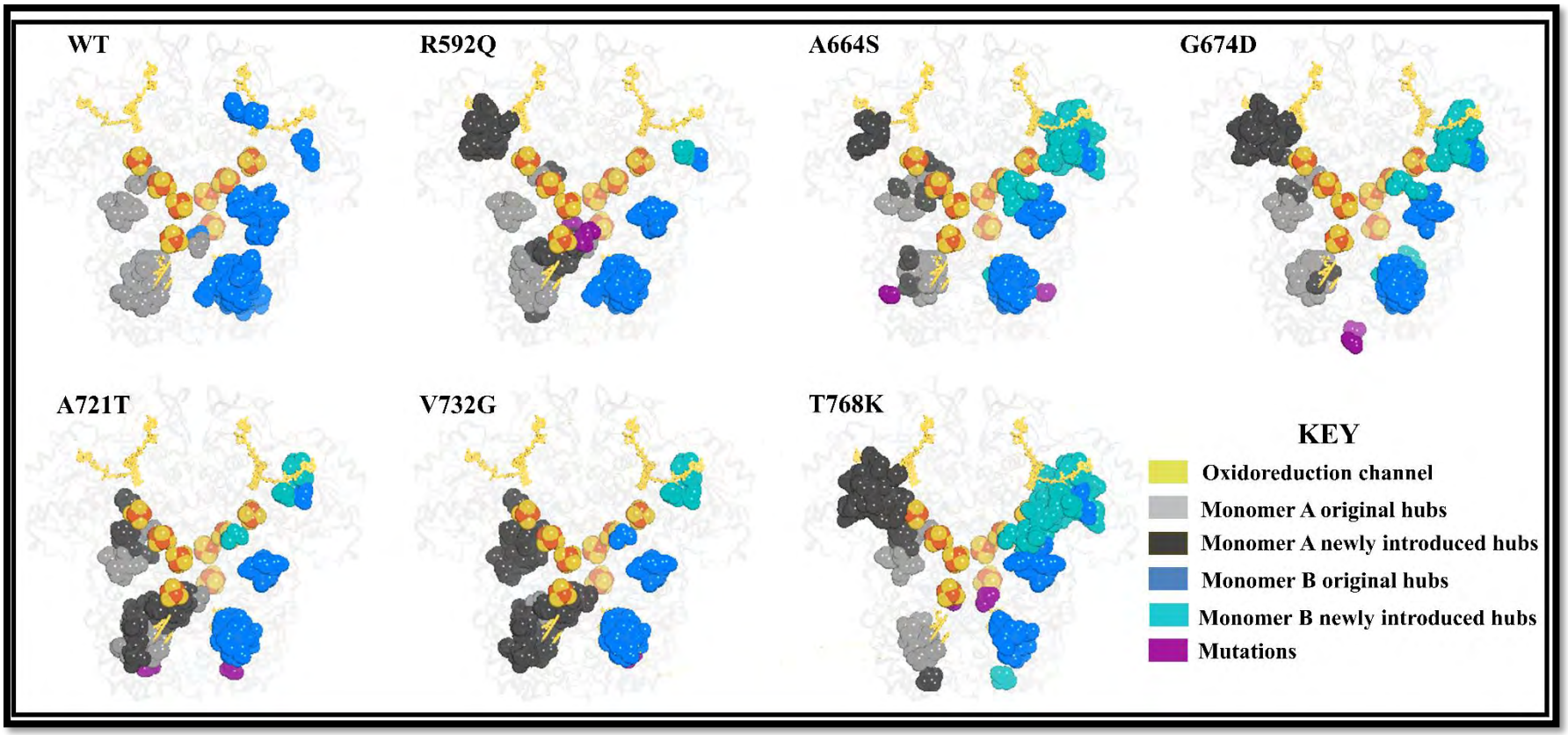


Figure 5.8.2: **3D mapping of the global top 4% for averaged *KC* metric of the WT and mutations in DPD drug bound ensembles.** A communication pathway along oxidoreduction (gold) pathway from Monomer A NADPH to FAD to 2 x $Fe^{2+}{}_4S^{2-}{}_4$ clusters, proceeding to Monomer B 2 x $Fe^{2+}{}_4S^{2-}{}_4$ to FMN for the reduction of 5-FU. The hubs in Monomer A are colored light grey (original) and dark grey (newly introduced) and Monomer A are colored royal blue (original) and sky blue (newly introduced).

5.3.3 Global analysis

An overview of global computational methods such as essential dynamics (ED) featuring principal component analysis (PCA), hydrogen bonds (H-bonds), root mean square of fluctuation (RMSF), framewise (RMSD), and root mean square of deviation (RMSD) have been used to illustrate DPD conformational dynamics introduced by active site mutation in this section.

5.3.3.1 Dynamic global conformation changes were observed in DPD protein in presence of mutations.

The local analysis uncovered a number of monomeric asymmetric inter-residue changes near the active site domain mutation and the catalytic-binding environment. In this section, we used one of our newly developed tools from MDM-TASK-web comparative ED to investigate mutationally induced conformational changes in the active site binding environment per monomer (Figure 5.9) (284). This was similar to what we did in our previous analysis. Sample trajectories from the drug free and drug bound active sites of both monomers were aligned to create a comparable trajectory-specific covariance matrix. Following that, a single covariance matrix was computed. DPD's catalytic site binding environment consists of drug free (open conformation) and drug bound (closed conformation) residues (583, 609, 612, 613, 642, 668, 669, active-loop 670 - 682, 709, 736, and 737) (129-131). Monomer A accounted for 45.1% of the extensively sampled global conformational subspaces in both the drug free and drug bound states (PC1: 29.65% and PC2: 19.45%). This was slightly lower than the values found in Monomer B (Total: 47.36%; PC1: 25.95%; PC2: 21.41%)

.

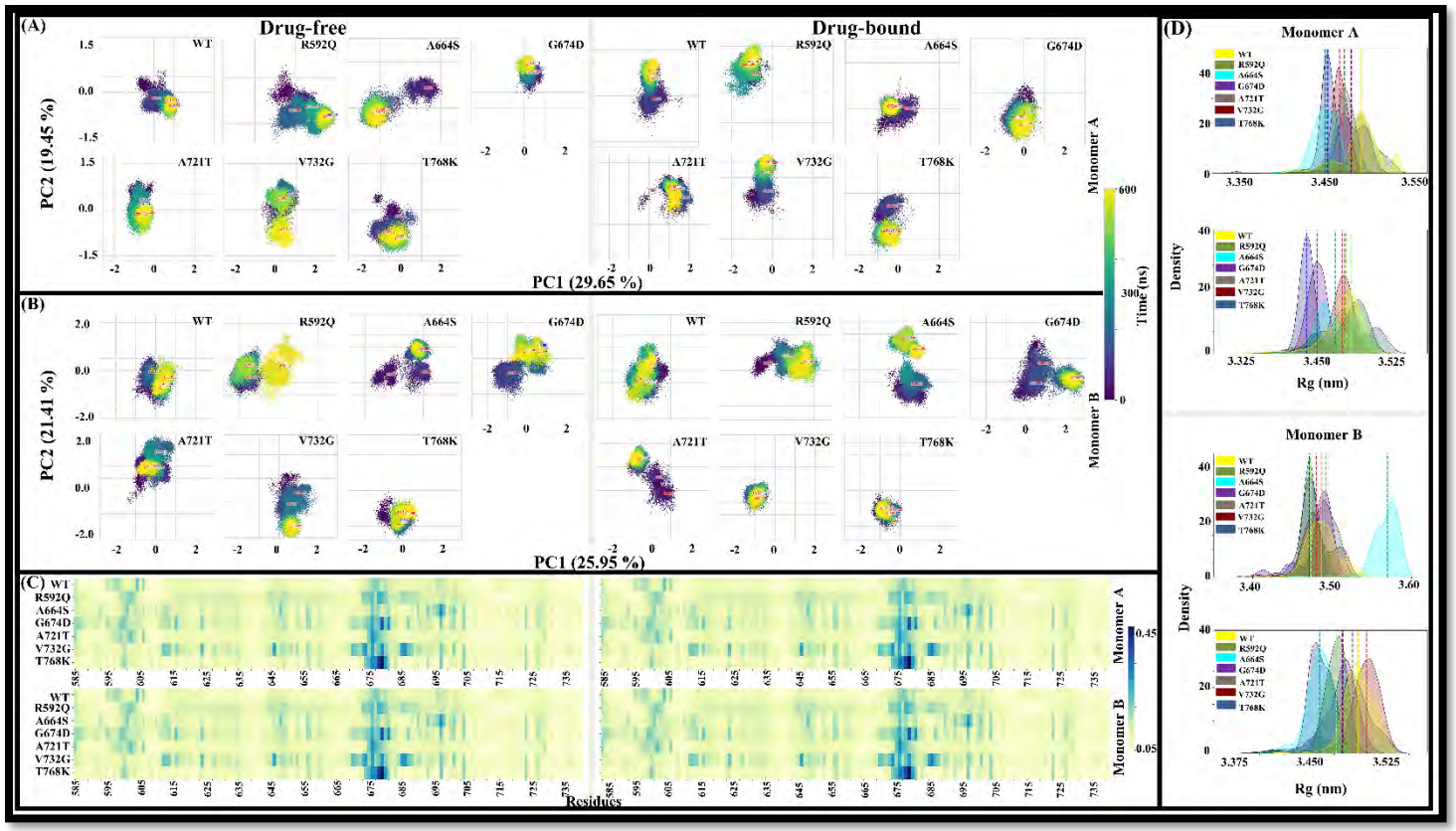


Figure 5.9: An illustration of DPD binding environment. Comparative essential dynamic scatter plots of inactivated and activated Monomer A (A) and Monomer B (B). The x-axis indicates PC1 while the y-axis indicates PC2. (C) RMSF of the active site residues of inactivated and activated systems. The x-axis represents residues while the y-axis represents mutations systems. (D) Density plots of the DPD Rg distribution for Monomer A and B. Yellow (WT), green (A664S), purple (G674D), grey (A721T), maroon (V732G), and blue (T768K) color codes have been used to represent the different systems.

Furthermore, in both the inactive and activated states, Monomer A occupied a more central conformation subspace. Inactivated R592Q, A664S, and V732G conformational subspaces were more dispersed than their activated counterparts, which assumed slightly constricted subspaces. Inactivated G674D, in contrast, had more constrained conformational subspaces that relax slightly when activated. Inactivated and activated R592Q, A664S, G674D, A721T, V732G, and T768K also had a diverse conformation distribution compared to WT. DRN multi-metrics analysis and weighted contact maps both revealed asymmetric behavior. This was also evident in the active site binding environment RMSF (Figure 48.C), where all mutant systems exhibited higher fluctuation, particularly in the area of the active loop region (670 – 682). Similarly, Rg (Figure 5.9.D) exhibited more gyration in the active site binding environment than WT, especially with activated A664S. It should be noted, the activation-loop (residue 670 -682) closed conformation is required for DPD protein catalytic activity (129-131). An earlier study on kinases discovered that oncogenic mutations in their catalytic domains increased activity by changing activation-loop conformations. The apparent effect on catalytic site structure, however, varied depending on the mutation (537). Therefore, the general mutation-induced changes in the catalytic domain and DPD activity might be associated with the dynamic activation-loop conformations.

5.3.3.2 Mutation resulted in a substantial reduction in hydrogen bonds as well as an increase in center of mass distances

We investigated the COM distance between Monomer A and its corresponding active site B based on the dimerization mechanism of the activated DPD protein, and vice versa. Figure 49.A and B depicts the intramolecular H-bond analysis of DPD WT and mutations relative to the 5-FU binding environment over a 600 ns simulation period. Throughout the simulation, an elaborate network of hydrogen bonds was maintained within the active sites of the individual monomers in the activated WT. However, there was a reduction in H-bond contact in Monomer A (A664S, G674D, and A721T) and Monomer B (R592Q, A664S, G674D, A721T, V732G, and T764K) (Figure 5.10.A and B). As a result, these mutants lost non-bonding interactions, protein folding, and protein-ligand recognition, which could have affected enzyme catalysis (500).

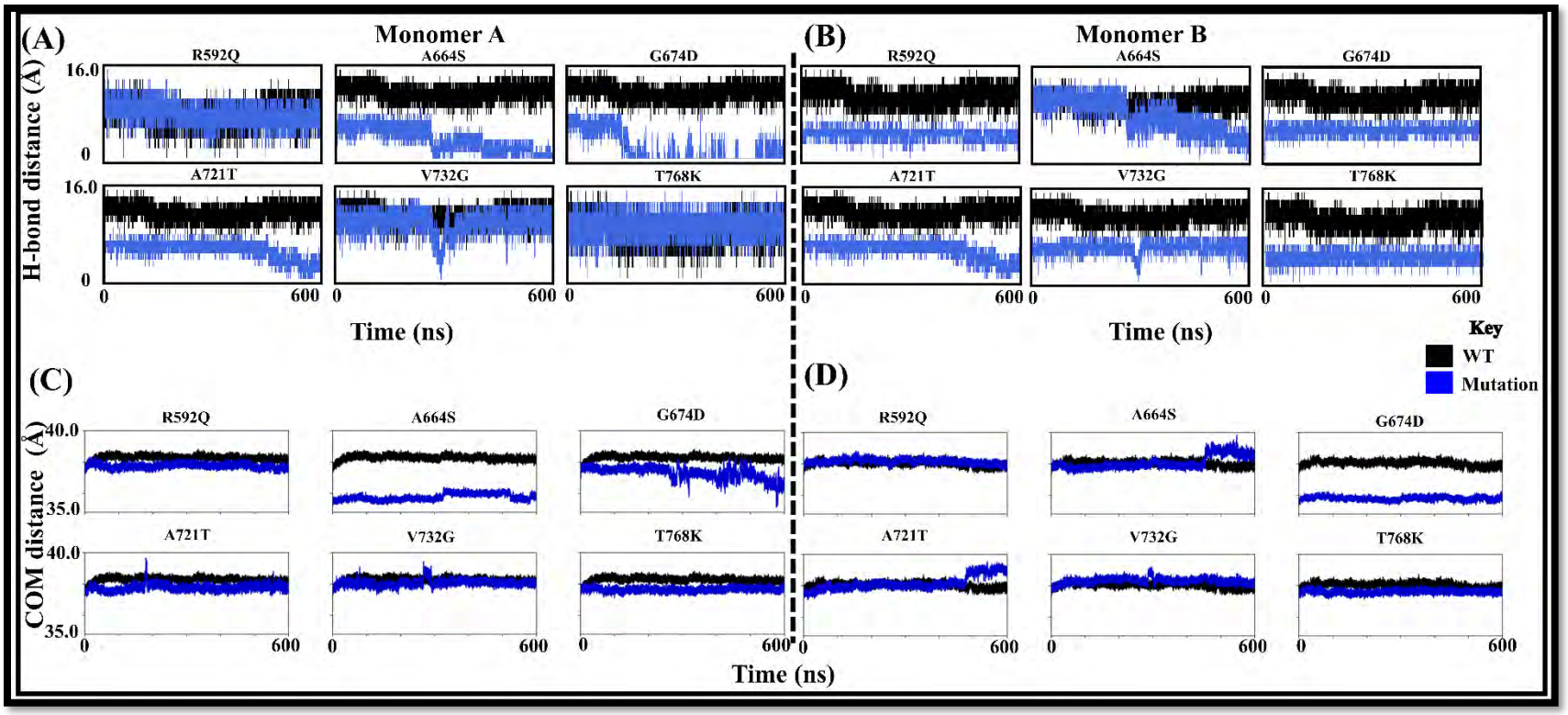


Figure 5.10: Hydrogen bond analysis during the 600 ns simulation period of 5-FU binding pocket of Monomer A (A) and Monomer B (B). Center of Mass (COM) analysis during the 600 ns simulation from (A) 5-FU binding environment in active site domain B to Monomer A and (B) 5-FU binding environment in active site domain A to Monomer B.

The positional stability of 5-FU was investigated by measuring the COM distances from 5-FU to the corresponding monomer (Figure 5.10.C and 5.10.D). These calculations are considered to be more accurate and insightful, since the active site displayed significant variability in different mutation cases (59). Monomer A (A664S and G674D) and Monomer B both showed a marked increase in COM distances (G674D).

5.3.3.3 Comprehensive RMSD and RMSF analyses revealed dynamic asymmetric behavior of the monomers, with emphasized loop area fluctuations in the presence of mutations

To monitor positional divergence of a structure relative to its initial conformation with time, RMSD over 600 ns MD simulation against timeline plots of the WT and mutations were established (Figure 5.11.A). The native and mutant proteins differed from their starting structure in distinct patterns until 600 ns, resulting in backbone RMSD ranging from 1.0 Å to 5.5 Å in both open and closed conformation. After 350 ns, some mutations demonstrated apparent deviation patterns when compared to the WT until the simulation ended. In the open conformation, T768K (2.5 Å), A664S (3.6 Å), and G674D (3.9 Å) of Monomer A differed the most from the WT (3.2 Å), while R592Q (2.7 Å), and G674D (2.8 Å) of Monomer B differed the most from the WT (4.2 Å). T768K (2.4 Å), A664S (5.4 Å), and A721T (5.4 Å) of Monomer A and G674D (2.6 Å) and T768K (2.6 Å) of Monomer B for the activated systems revealed remarkable divergence from their corresponding WT (Monomer A: 3.2 Å and Monomer B: 3.9 Å). With the exception of T768K, which equilibrated at a single conformation (unimodal conformation) in both inactivated and activated Monomer A ensembles, the remaining systems equilibrated at multiple conformations (multimodal conformation) (Figure 5.11.B). The asymmetry of the ensembles is also highlighted by the different distribution patterns of the respective monomers of the same systems in their inactive and activated states. For the inactivated WT, a bimodal distribution (two conformation equilibration) was observed, as previously observed in our DPD remote mutation analysis.

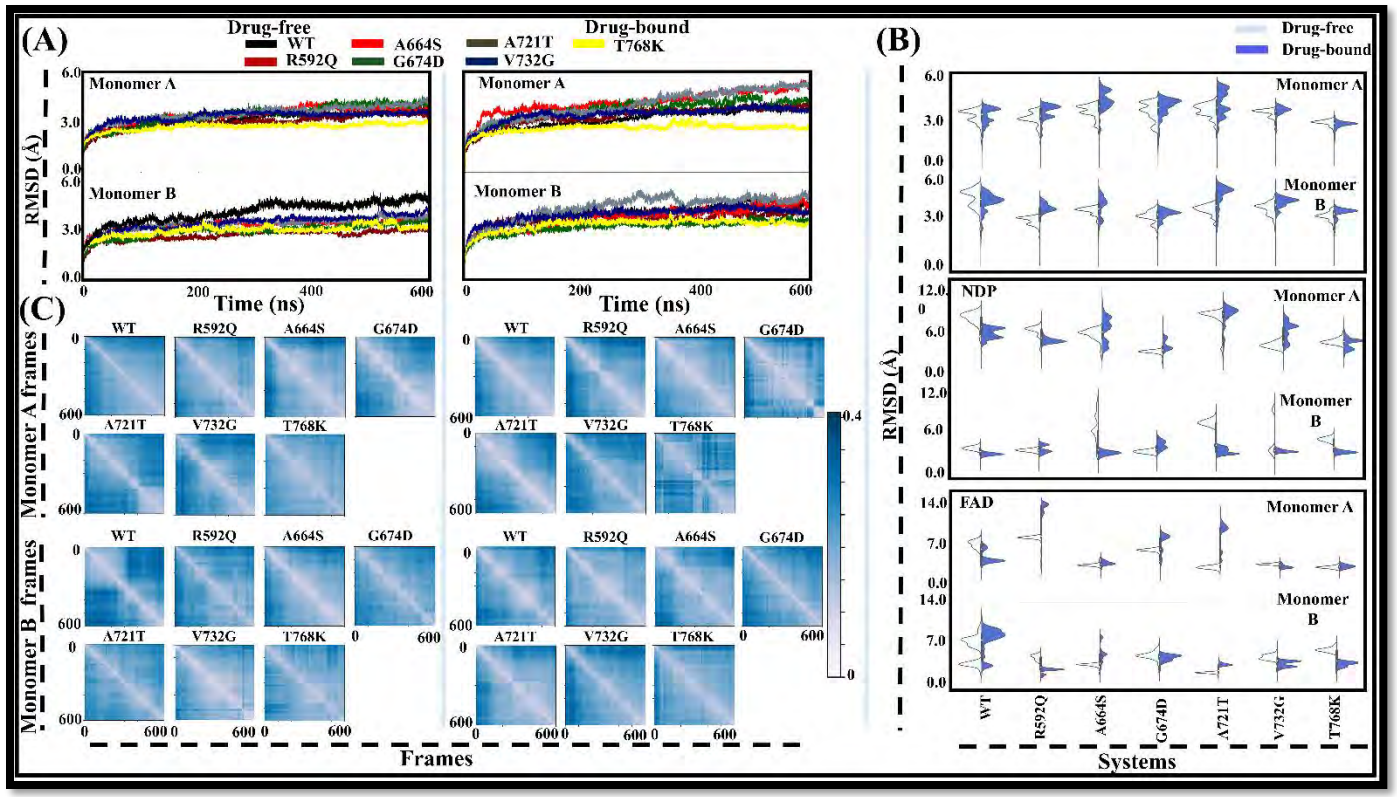


Figure 5.11: The RMSD line graph, violin plots and RMSD framewise representation of the WT and mutations. (A) WT (black), R592Q (maroon), A664S (red), G674D (green), A721T (grey), V732G (blue) and T768K (yellow). The x-axis depicts the time (Å) and y-axis the RMSD (Å). **(B)** The split violin plots RMSD representation for drug free (light blue) and drug bound (royal blue) Monomer A and Monomer B. The x-axis shows the systems, while the y-axis shows the RMSD (Å). **(C)** The RMSD data for all versus all is represented as a heatmap. The x- and y-axes depict the frames at various time frames. **(D)** The split violin plots RMSD representation of cofactors for the drug free (light blue) and drug bound (royal blue) Monomer A and B. The x-axis represents the systems, and the y-axis represents the RMSD (Å).

A multimodal distribution (multiple conformation equilibrium states) was noticed after drug binding. A664S and G674D differ markedly from the WT. G674D (Monomer A: 19.5 and Monomer B: 20.0) was the closest to the active site, with A664S (Monomer A: 19.5 and Monomer B: 20.0) being relatively close to the active site too, in respect to other mutations (Table S4.2). The fact that they were so close to the active site may have contributed to the evident conformation distribution.

Using all versus all frame RMSD heatmaps, additional study compared DPD WT MD frames to themselves and to all other frames (Figure 5.11.C). The drug free WT exhibited two major conformation ensembles for Monomer A (i: 0-50 ns & ii: 50-600 ns) and Monomer B (i: 0-450 ns & ii: 450-600 ns), though the drug bound systems displayed multimodal distribution. Similar patterns were realized in the mutation systems, where different conformation stamps were observed before and after drug binding. Furthermore, asymmetric RMSD tendencies per monomer and the protein's drug free and drug bound states were observed at the cofactors level as well (Figure 5.11.D).

5.3.3.4 Mutation caused asymmetric behaviour with enhanced active-loop fluctuation within the active site, while the protein core remained stable

To depict the residual fluctuation and inter-relationship of the inactivated and activated systems' asymmetric monomeric behavior caused by mutations, RMSF and DCC calculations were used (Figure 5.12). In both the inactive and activated states, each monomer displayed seven regions of significant fluctuation. The most variability was found in the beta-sheets and successive loop regions, which included the following residues: 40 - 52, 316 - 334, 736 - 771, 901 - 922, 964 - 971, N-terminal residues (2 - 12), C-terminal residues (996 - 1025), and active site loop (residues 675 - 679). The active site residues of Monomer A A664S and B T768K demonstrated high fluctuation in their drug free states, whereas Monomer A A664S, G674D, and A721T, as well as Monomer B R592Q and A721T displayed a wide range of fluctuation in their drug bound state. The protein core region ($Fe^{2+}{}_4S^{2-}{}_4$ clusters, FAD, NDP, and FMN) fluctuated the least, suggesting that residues in this environment were relatively more stable throughout the simulation process. This is something that has been observed in past findings (218).

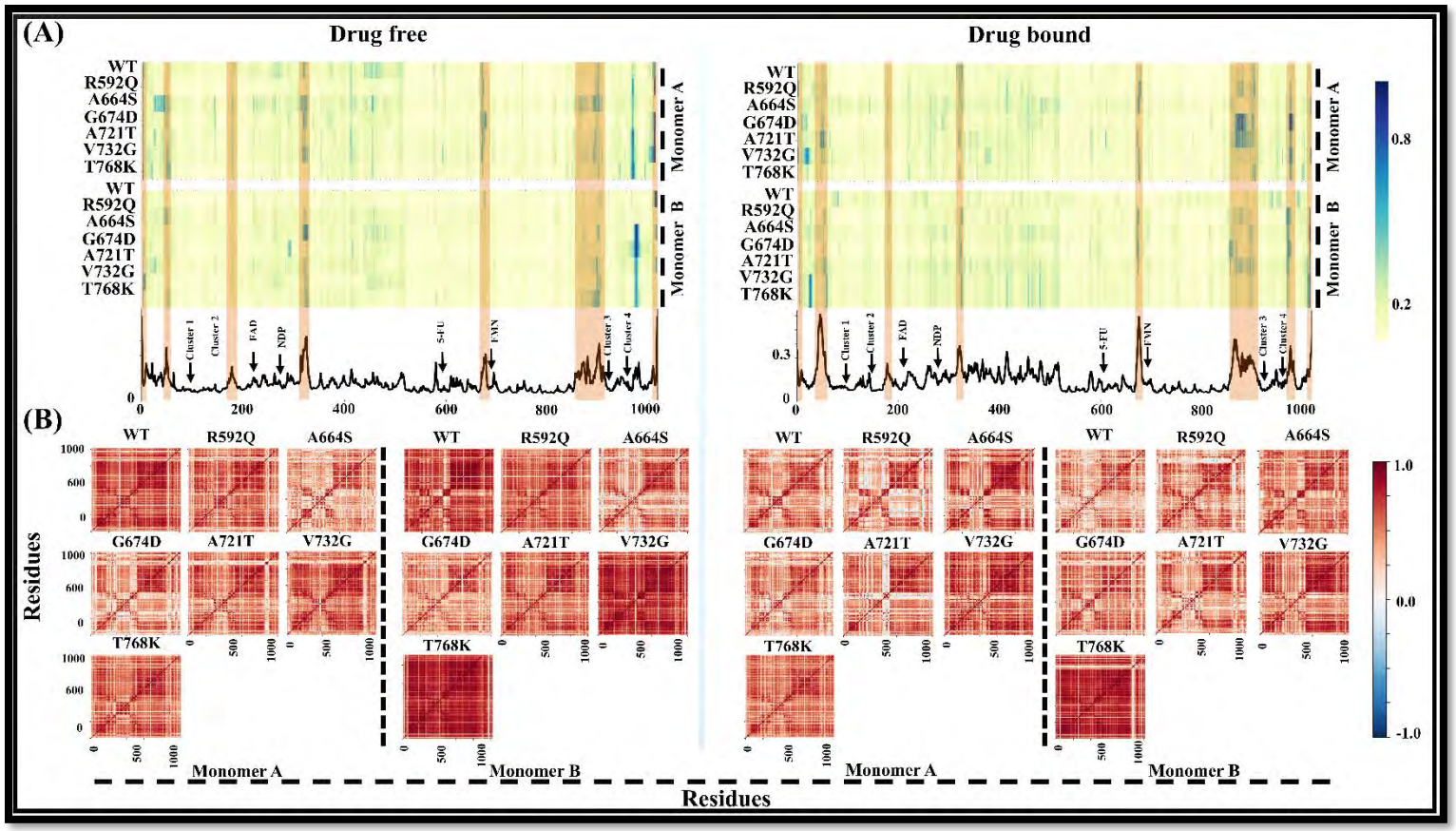


Figure 5.12: DPD drug free and drug bound heatmap showing the root mean square fluctuation (RMSF) and dynamic cross-correlation matrix (DCCM) of the WT and 7 mutants after 600 ns simulation. (A) The RMSF is represented as a fluctuation relative to each atom's mean position, with high fluctuation regions highlighted in light brown box-frames. The x-axis shows the residue numbers and y-axis shows the WT and mutation systems. Cofactor locations are also indicated. (B) DCC represents inter-residue relationship in both the drug free and drug bound systems. The highly correlated residues are colored cream to brown by DCCM, while anti-correlated residues are colored blue to cream.

Following that, DCC was used to depict the correlated conformational motion of residues in order to highlight the possible consequences of mutation on each monomer of the drug free and drug bound DPD (Figure 5.12.B). The matrices show that highly positive regions (colored cream to brown) have strong correlated motions (residue pairs move in the same direction), whereas negatively positive regions (colored blue to cream) have strong anti-cross-correlation motions (residue pairs move in the opposite direction). Although WT residues showed relatively stronger cross-correlation in both monomers' of drug free state, this diminished in upon drug bounding. Furthermore, active site region residues had stronger positive correlation than residues from other domains across all systems. Heightened levels of residue anticorrelation were observed near the mutation site when compared to the WT. Each system demonstrated a distinct inter-relationship when compared to the WT. Notwithstanding, the asymmetric behavior of each monomer per ensemble was confirmed.

In the comprehensive global analysis involving RMSD, RMSD framewise analysis, RMSF, DCC, ED, and H-bond, several mutations, specifically A664S, G674D, R592Q, and T768K, consistently yielded noteworthy results across most of these analyses. A study conducted by Offer et al., (174) revealed that the G674D mutation displayed no residual enzymatic activity (< 12.5%) ($P = 5.7 \times 10^{-9}$). Reduced enzyme activities ranging from 12.5% to 25% were observed in the case of the R592Q mutation ($P = 2.6 \times 10^{-5}$). Surprisingly, despite the diverse global changes observed, the A664S mutation did not exhibit a significant lack of enzymatic activity. Similarly, the A721T and V732G mutations did not demonstrate a significant change in DPD enzymatic activity. Intriguingly, Offer et al.'s study reported no enzymatic activity for the T768K mutation (174). However, Ogura and colleagues, in their investigation of DPD activity in peripheral blood mononuclear cells (PBMC) in Japanese population, reported that the T768K mutant displayed a decreased intrinsic clearance rate (83% of Cl_{int} ratio) compared to the wild type (503).

5.4 Conclusion

Precision oncology medicine will be feasible only if pharmacogenomics are better understood (538). The key characteristics of the discovery of novel drugs that are potent, efficacious, and safe are based on a deeper comprehension of the protein

catalytic mechanism and how mutations affect this process (59, 538). An increasing amount of research indicates that gene mutations in drug metabolizing enzymes and transporters could have a great influence on inter-individual drug reaction and disposition differences (9, 539). Elaborate pharmacogenomic profiles of these enzymes have recently been developed. These studies, regrettably, have ignored the effect of mutations in specific protein regions on protein functionality (478). We used our recently developed dynamic network analysis (DRN) multi-metric approach to show a residue interaction network within the DPD protein, with a special focus on the catalytic domain. To help understand how active site domain DPD mutations (R592Q, A664S, G674D, A721T, V731G, and T768K) cause reconfiguration of residue interactions within the active site environment, different metrics *(BC, CC, DC, EC,* and *KC)* used to characterize the topology of these networks were compared. Additionally, we demonstrated that dimer missense mutations can induce asymmetric paradoxical activation of the active site environment of the monomers through dynamic changes in the active site loop-controlled residue interaction network (residues 675-679), dimer interface sites, and oxidoreduction catalytic path components (NADPH, FAD, 4 X $Fe^{2+}{}_4S^{2-}{}_4$ clusters, FMN, and 5-FU). Two pairs of compensatory communication networks from NADPH to the 5-FU (catalytic site) were observed using both *BC* and *DC*, one within each monomer. The *BC* pathway ran along the inner side of the oxidoreduction catalytic channel, whereas the *DC* pathway ran toward the exterior. In both cases, one pathway was discovered to be more dominant. Correspondingly, the weighted contact-map revealed a mutation-induced compensatory mechanism of active site residues, wherein new residues were obtained to maintain communication. Catalytic residues Cys671 and Ser670 emphasized this mechanism well. Although these residues lost contact with other catalytic key residues, new interactions were formed to reestablish communication. Moreover, *CC* revealed that the most persistent hubs were found within the protein's core, whereas *EC* identified the WT's active site. In addition to identifying the active site, *KC* identified key areas involved in protein communication such as FAD, $Fe^{2+}{}_4S^{2-}{}_4$ clusters, and FMN. These changes were also evidenced in the global analyses. The asymmetric behavior of monomers was highlighted using RMSD, framewise RMSD, RMSF, DCC, PCA, and hydrogen bond analyses. Equally, various conformational changes were detected, which were alluded primarily to increased fluctuation within loop areas, including active-loop areas. G674D and A664S were

the most affected systems across all analyses (local and global). These mutations impair enzymatic function, according to *in vitro* studies (174, 445). Our study's key findings are that mutations mimic the dynamics of the WT communication network, culminating in asymmetric paradox activation of the active site, which may jeopardize proper 5-FU metabolism and lead to toxicity. More importantly, we discovered cold spots that could be used to create novel drug modulators to restore proper DPD enzyme 5-FU metabolism. This will establish a precision oncology strategy focused on the patient's genetic makeup.

CHAPTER 6
OVERALL CONCLUSION AND RECOMMENDATION

Genetic variations in drug metabolizing enzymes have been linked to changes in pharmacokinetics, drug dosage, and other variables that influence therapeutic outcomes (540). Evidently, adverse drug reactions (ADRs) and lack of efficacy represent the most likely causes for drug failure in clinical trials, contributing roughly 50% of all failures (541, 542). Consequently, efficacy and safety have been identified as the two most important factors in drug development (541). Therefore, the holistic goal of this thesis was to use pharmacogenomic knowledge of drug metabolizing enzymes to develop reliable computational approaches for elucidating their molecular mode of action, and to apply this knowledge in drug development.

The Dihydropyrimidine dehydrogenase (DPD) protein plays a crucial role in the metabolism of 5-Fluorouracil (5-FU), a commonly used chemotherapeutic agent. Mutations in the DPYD gene can lead to deficiencies in the DPD protein, resulting in malfunctioning 5-FU metabolism (186). This deficiency can have severe implications on the efficacy and toxicity of 5-FU treatment (179). Using DPD enzyme as a case study, we demonstrated how the presence of mutations in drug metabolizing enzymes influences the dysfunctional degradation of drugs. This may lead to a variable inter-individual response to medications, including an increased risk of ADRs and decreased efficacy. DPD deficiencies can result in severe 5-FU toxicity; even so, significant geographical-ethnic variances exist with regards to DPD missense mutations (158, 543). Our study focused on DPD missense mutations associated with the African descent subpopulation due to a lack of research in this area (146, 186). This information was deemed critical in paving the way for the discovery of allosteric drug modulators from new active ingredients or through drug repurposing/repositioning, as a result addressing inter-individual mutations related to 5-FU efficacy and toxicity. We combined several well-established tools and approaches to create a comprehensive protocol for parametrization of metalloproteins like DPD. Furthermore, we highlighted the dynamics of mutation-induced impact on protein structure and function. These processes entail DPD mutation identification, homology modeling, parametrization, molecular dynamics simulations, and DRN approaches.

In the first instance, Chapter 1, we introduced the different kinds of drug metabolizing enzymes (DMEs), which are classified under Phase I and Phase II. Cytochrome P450s (CYPs) are the most important Phase I enzymes, while UDP-glucuronosyltransferases (UGTs), glutathione-S-transferases (GSTs), and dihydropyrimidine dehydrogenases (DPDs) are the most important Phase II enzymes. We have provided an analytic tool in the burgeoning area of precision medicine by combining pharmacogenomic knowledge of variability in individual drug metabolism due to missense mutations with deep knowledge of DPD case study.

In Chapter 2, detailed descriptions of the methodologies used in our investigations were stipulated. This included extensive clinical, *in vitro*, *ex vivo* and *in silico* identification of DPD missense mutations linked to people of African ancestry; homology modelling and Fe^{2+} AMBER force field parametrization of the WT and mutation proteins, MD simulation and post-MD analysis including DRN combinatory approaches.

In Chapter 3, the generation of force field parameters using the original Seminario method was reported. We went one step further and used the VFFDT Seminario method's collation features to obtain the force field parameters of the same Fe^{2+} ions as a supportive measure. This was accomplished by taking into consideration the human DPD protein's dimeric functionality, which is reliant on well-coordinated inter-chain electron transfer across a four $Fe^{2+}{}_4S^{2-}{}_4$ cluster complex per monomer. Most importantly, the generated novel force field parameters enabled further investigation of DPD protein functionality through MD simulation.

In Chapter 4, we identified six clinically significant pathogenic mutations (C29R, M166V, Y186C, S534N, I543V, and D949V) based on previous studies. Two of these mutations (S534N and I543V) were in the active site domain, while the rest were in the non-active site domain. The pathogenicity of the remote mutation D949V has been well studied in the Caucasian population, with dose adjustment treatment strategies used in carriers of this variant (165, 445, 543). We have demonstrated how the remote clinically significant (C29R, M166V, Y186C and D949V) mutations induce asymmetric dynamic dimerism of the DPD active site through allosteric communication pathways. In addition, we demonstrated the mutation-induced changes in the active site caused by mutations in the active site domain.

In Chapter 5, considering previous *in vitro* and *ex vivo* DPD investigations (168, 176, 450, 451), we went on to identify 13 missense mutations associated with

the African descent subpopulation. Similar to the results of clinically significant mutations, a monomeric asymmetrical induced allosteric communication pathway to the active site due to remote non-clinically significant mutations (PART I: S201R, K259E, D342N, D432N, S492L, R886C and L993R) was observed. On the other hand, the non-clinically significant identified mutations situated at the active site domain (PART II: R592Q, A664S, G674D, A721T, V732G and T768K) not only affected the 5-FU catalytic and binding site residues but also provoked an asymmetric change in the communication pattern along the oxidoreduction channel.

Notably, all systems demonstrated consistent cold spot regions, which were indicated by the presence of persistent hubs (Gly71, Ala72, Leu73, Ala74, Glu75, Ala76, Cys79, Leu80, Ala101, Ile104, Asn108, Tyr110, Gly111, Leu142, Ile150, Leu155, Val373, Pro598, Met599, Pro602, Ser605, Leu775, Ala777, Val778, Tyr779, Ile781, Ala782, Ile790, Phe805, Ser808, Ala810, and Gly833). In next studies, these areas will be targeted for the purpose of discovering novel drug modulators via drug repositioning and repurposing or by searching for novel active ingredients. This would aid in dealing with molecular heterogeneity, particularly mutation-induced variations, resulting in proper DPD functionality restoration.

Conclusively, our findings provide insight on comprehensive protocols that could be used to conduct an integrative framework of pharmacogenomic variability caused by missense mutations in a cancer drug metabolizing enzyme. The identification of hot and cold spots revealed target areas for novel drug modulators that would restore the protein's proper functioning. We believe that the use of computational approaches is one potential avenue for improving drug discovery via new anti-tumour active chemical compounds or repurposing/repositioning analytics. Furthermore, missense mutation protein structure-based pharmacogenomic knowledge in cancer, will pave the way for precision oncology through discovery of drug modulators. It stands to reason that building collaborative multidisciplinary teams to draw on the expertise of molecular oncologists, oncological surgeons, and clinical oncologists, but also pharmacologists and cancer cell biologists, as well as data scientists, bioinformaticians, and computational biologists is critical for further development of these approaches. These teams will be capable of mining databases, conducting clinical data research, conducting high-throughput experiments, validating findings, and enhancing clinical outcomes for the gain of oncological patients.

Study Limitations

One limitation of this study was the computational expense associated with performing molecular dynamic simulations, which resulted in duplicate runs being conducted only for the wild-type (WT) system and not for the mutant system. This limitation may have hindered the ability to conduct exhaustive analyses for the mutant protein. Additionally, the study did not consider mutations that are confidently known to have no effect on DPD activity, which could have served as a benign control. It is important to note that these mutations are typically found in a heterozygous state. Moreover, this study did not model hybrid dimers consisting of one WT and one mutant subunit, which could provide insights into the behavior of mixed protein complexes. Furthermore, the analysis did not specifically focus on how these mutations contribute to the formation of cavities within the protein, despite the known tendency of mutations to induce cavity formation in the interior or on the surface of the protein. Addressing these limitations could enhance the comprehensiveness and depth of the study's findings.

RECOMMENDATION

Computational approaches are essential in predicting the impact of DPD protein mutations. Through the utilization of bioinformatics tools and molecular dynamics simulations, valuable insights into the structural and functional consequences of specific mutations can be obtained. These computational studies aid in the identification of critical residues and regions affected by the mutations, thereby guiding subsequent experimental investigations. To further advance the field of DPD research, several future aspects should be considered. Firstly, modeling a human hybrid protein containing one subunit of wild-type (WT) and mutant protein can investigate whether the enzymatic activity of a WT-mutant hybrid dimer is equal to the average of a WT homodimer and a mutant homodimer. Secondly, the use of a known benign mutant as a control can provide valuable comparisons. Moreover, conducting duplicate molecular dynamics runs for both the WT and mutants can enhance the reliability and robustness of the findings. Furthermore, additional investigations into the effects of mutations on the protein surface and interior are warranted to explore the presence of mutational-induced cavity formation and its potential impact on the interaction network and water residence. Overall, these

findings suggest that DPD could be a target for novel drug modulators or gene therapy, as discussed in preceding chapters.

Experimental approaches, such as site-directed mutagenesis and protein expression studies, provide valuable experimental evidence to validate the computational predictions. These studies help elucidate the biochemical and biophysical properties of the mutant DPD proteins. Additionally, enzymatic assays and cell-based assays can be employed to evaluate the impact of the mutations on DPD activity and 5-FU metabolism. Further characterization of DPD variant enzyme activities could be analyzed according to, Amorosi et al., approach (544). This method, known as deep mutational scanning (DMS), enables the measurement of enzyme activity and steady cellular abundance of missense mutations through a high-throughput functional selection process. The selection enriches variants with high function and depletes those with low function, with high-throughput DNA sequencing used to quantify the frequency change of each variant during the selection. This provides a functional score for each variant in the library (545). Functional assessments in DMS can vary, often relating variant function to cell growth or measuring protein or ligand binding, but rarely measuring enzyme activity directly. For comprehensive functional data, the authors developed click-seq, a multiplexed sequencing-based method for quantifying protein variant activity. This method was employed to measure the activity of missense mutations. Additionally, the abundance of missense mutations in cultured human cells was measured using the previously developed massively parallel sequencing (VAMP-seq) assay by Matreyek et al., (546) which utilizes a fluorescent protein reporter coupled with fluorescence-activated cell sorting (FACS). Comparative analysis of activity and abundance revealed that reduced abundance accounted for at least 50% of the variants causing loss of function. These findings also highlighted key protein regions crucial for function, including residues involved in substrate binding. Clinicians can serve as important sources of evidence when confronted with variants of uncertain significance (VUSs) and can contribute to improving the dosing efficacy of drugs metabolized by an enzyme.

To address the challenges posed by DPD deficiency and 5-FU metabolism malfunction, a comprehensive approach that integrates computational, experimental, and clinical strategies is essential. By combining these approaches, researchers and clinicians can identify and understand specific DPD mutations, predict their

functional consequences, validate them experimentally, and translate this knowledge into clinical practice. The ultimate goal is to develop precision medicine solutions that optimize 5-FU treatment outcomes while minimizing toxicity in patients with DPD deficiency. This multidisciplinary approach holds great promise for improving patient care and advancing personalized cancer therapy.

Appendix

CHAPTER ONE

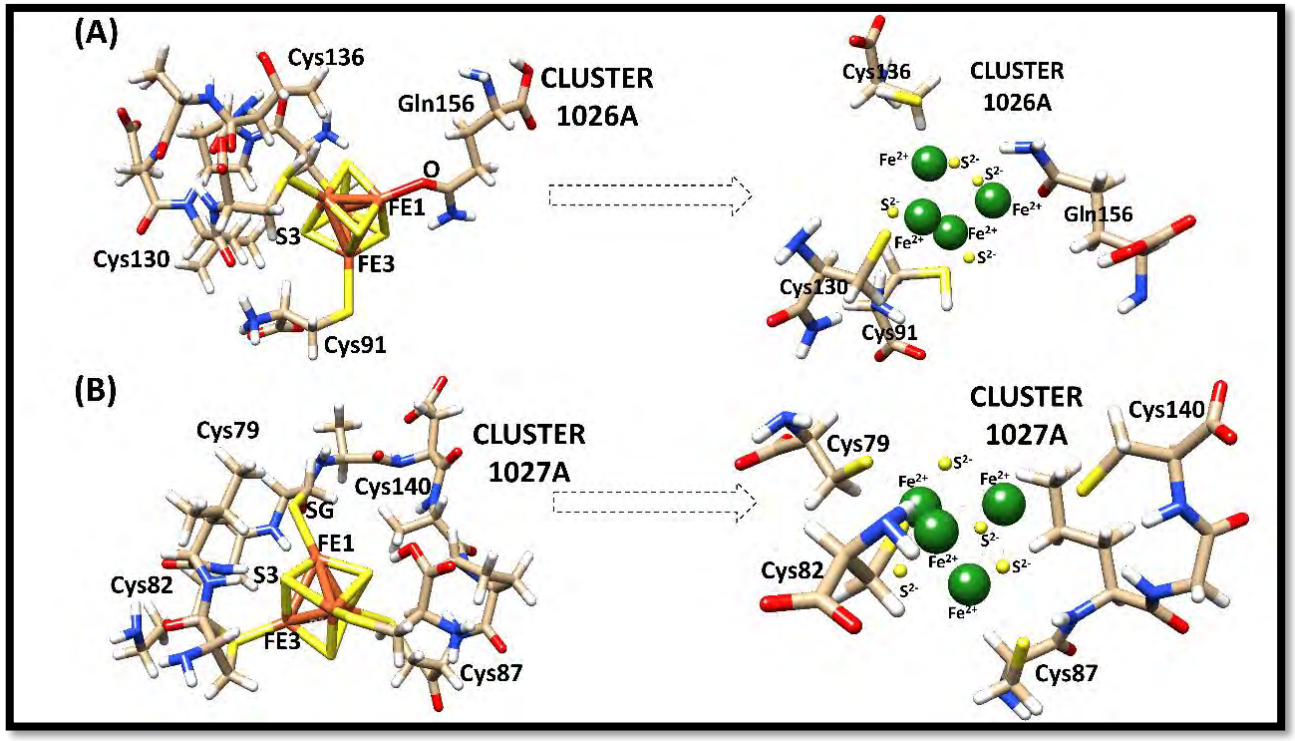


Figure S2.1. A representation of Fe^{2+} center parameterization in the native human DPD Model 2, using the automated (VFFDT) Seminario method. A) A 3D illustration of a coordinating Fe^{2+} sphere for cluster 1026A and an adjacent structure that has been parameterized, **B)** A 3D representation of the coordinating Fe^{2+} sphere for cluster 1027A as well as an adjacent structure that has been parameterized. Figure adapted from Tendwa et al., (147)

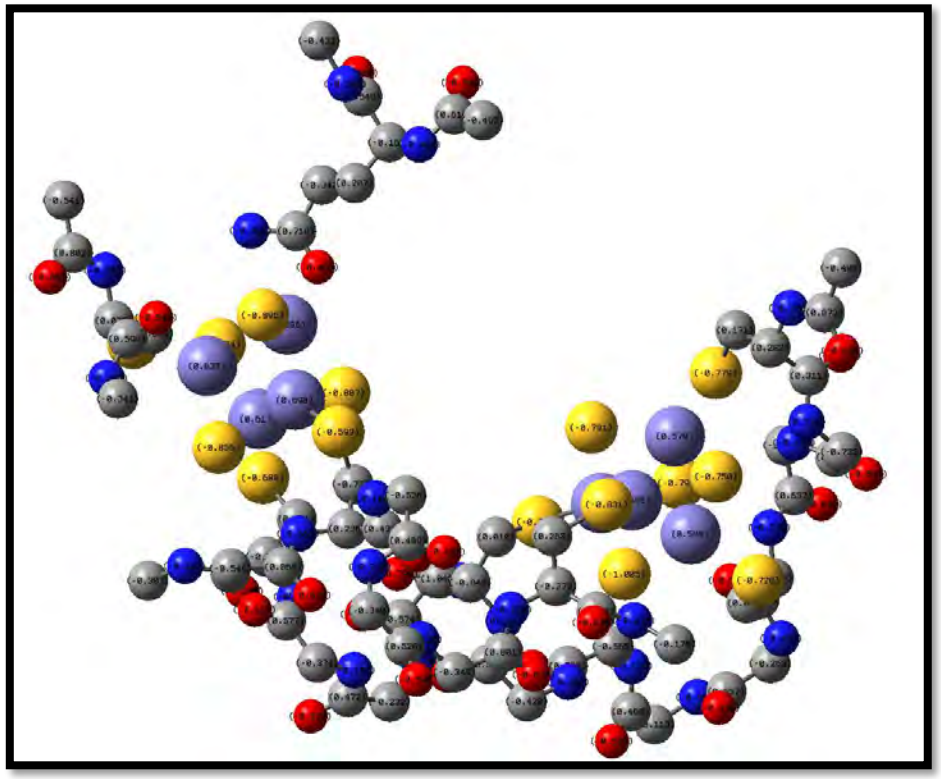


Figure S2.2. Charge allocation to all atoms coordinating with the metal centers of subset clusters 1026A and 1027A in Model 1. Yellow, purple, blue, red, and grey spheres represent sulfur, iron, nitrogen, oxygen and carbon atoms, respectively. Figure adapted from Tendwa et al., (147).

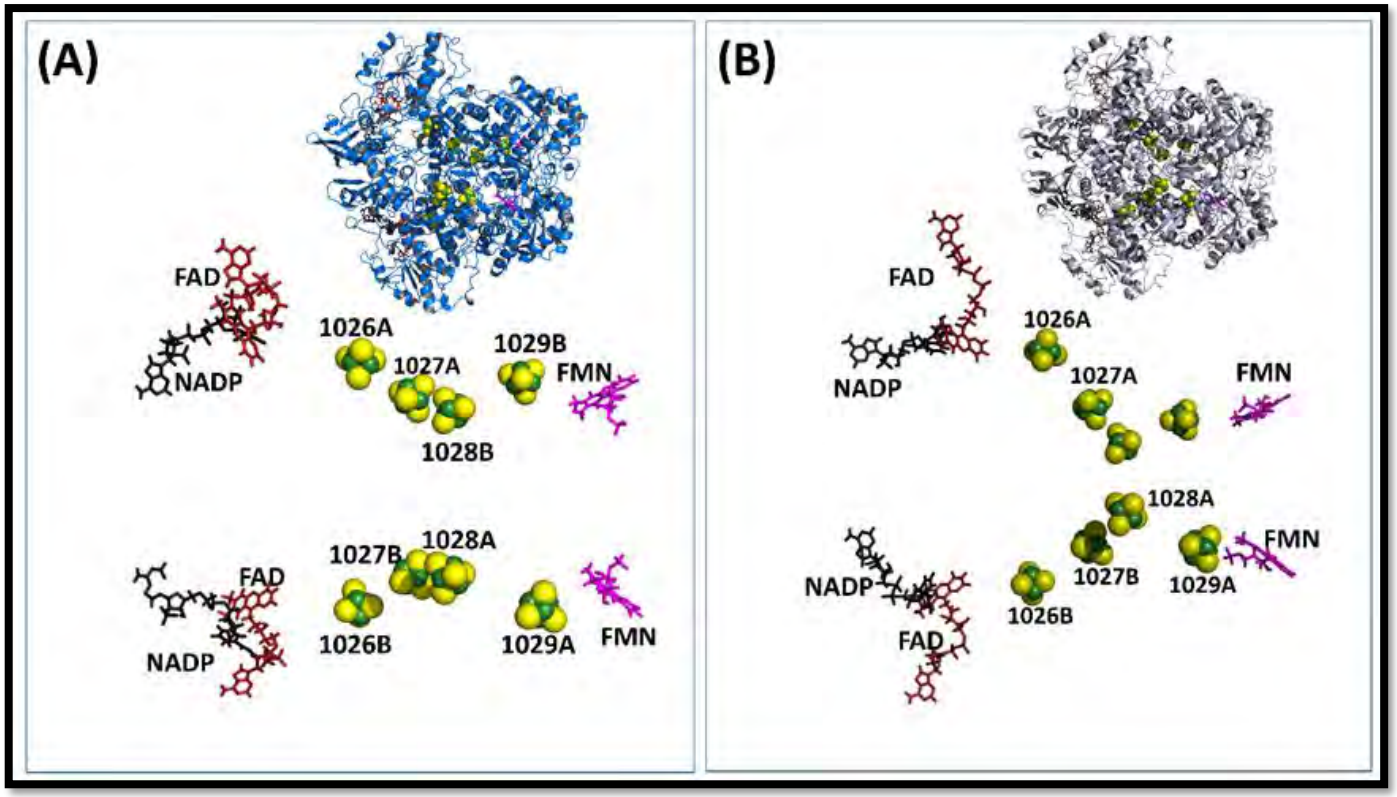


Figure S2.3. 3D structures of Model 2 MD simulation snapshots (timeframes) from regions showing greater conformational changes with atomistic details represented A) at 110.0 ns for drug bound protein and B) at 70.4 ns for holo proteins (without 5-fluorouracil drug). The Fe^{2+} clusters remained intact, and the protein core is preserved throughout the various conformation timelines. Figure adapted from Tendwa et al., (147)

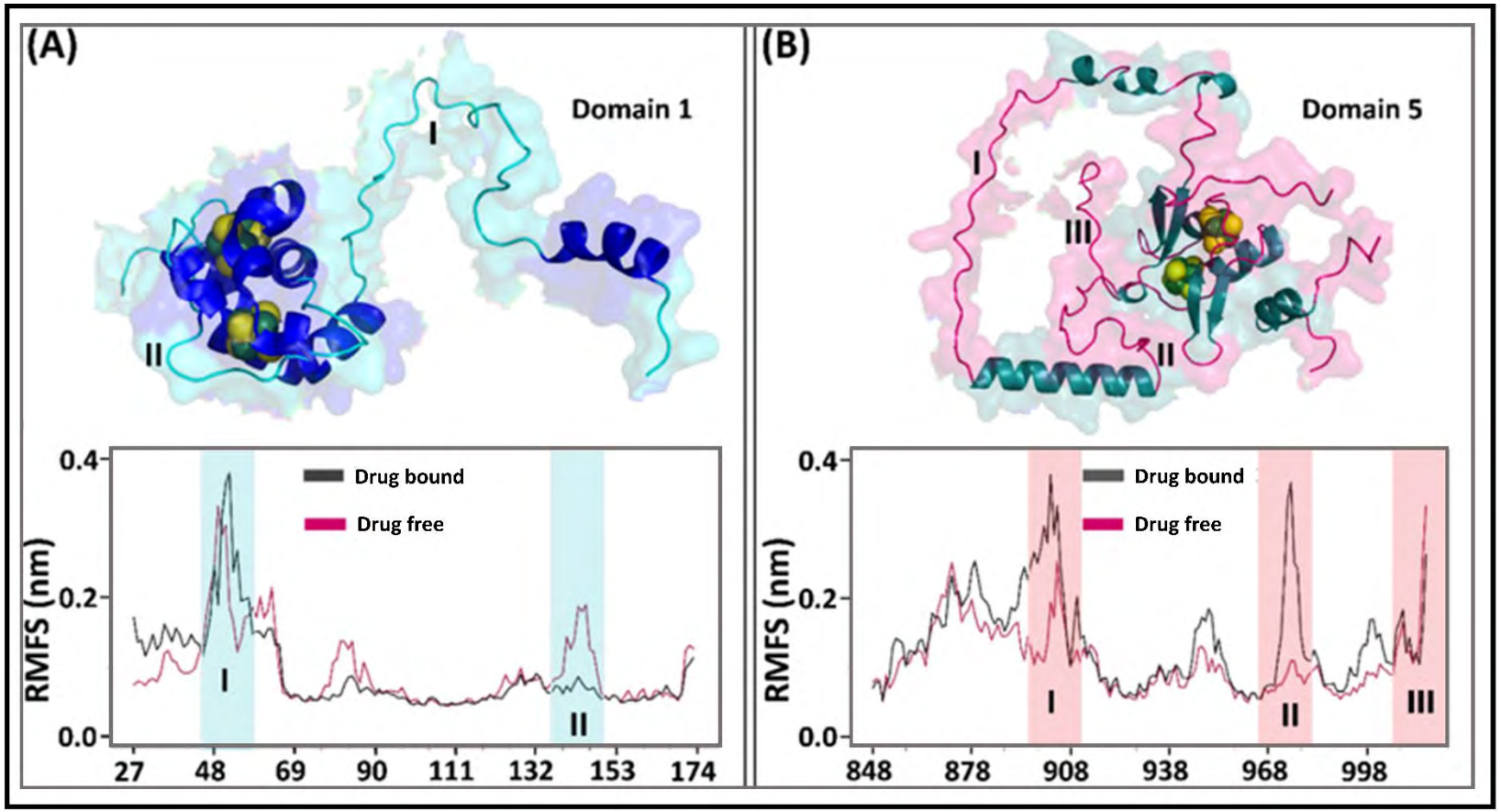


Figure S2.4. An illustration of the root mean square of fluctuation (RMSF) for drug bound and drug free DPD Model 1 during 150 ns simulation. **A)** $Fe^{2+}{}_4S^{2-}{}_4$ clusters in 1026 and 1027 positioned in domain 1. **B)** $Fe^{2+}{}_4S^{2-}{}_4$ clusters in 1028 and 1029 located in domain 5. The area of fluctuation coincides to the protein loop area while the iron cluster remains intact. Figure adapted from Tendwa et al., *(147)*

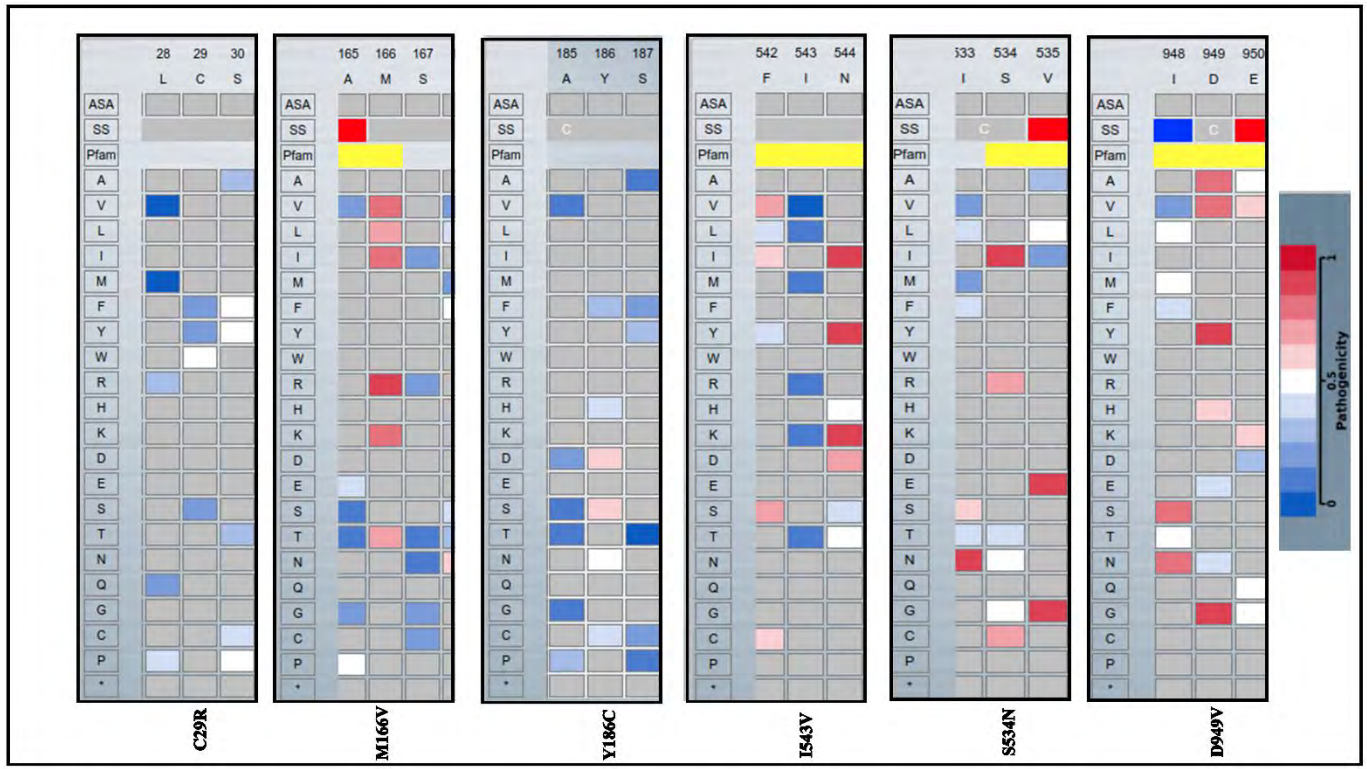


Figure S3.1.1: A heatmap displaying the sequence-based pathogenicity effects probability for clinically relevant mutations (C29R, M166V, Y186C, I543V, S534N, and D949V) generated by the Ensembl Variant Effect Predictor (EVE). The heatmap is color-coded from blue to white, representing a range from 0 to 1, indicating an ascending likelihood of pathogenicity. Mutations assigned a value of 0.5 on the heatmap are considered moderately pathogenic, while those assigned a value of 1 are highly likely to be pathogenic. Conversely, mutations with a value of 0 are less likely to be pathogenic.

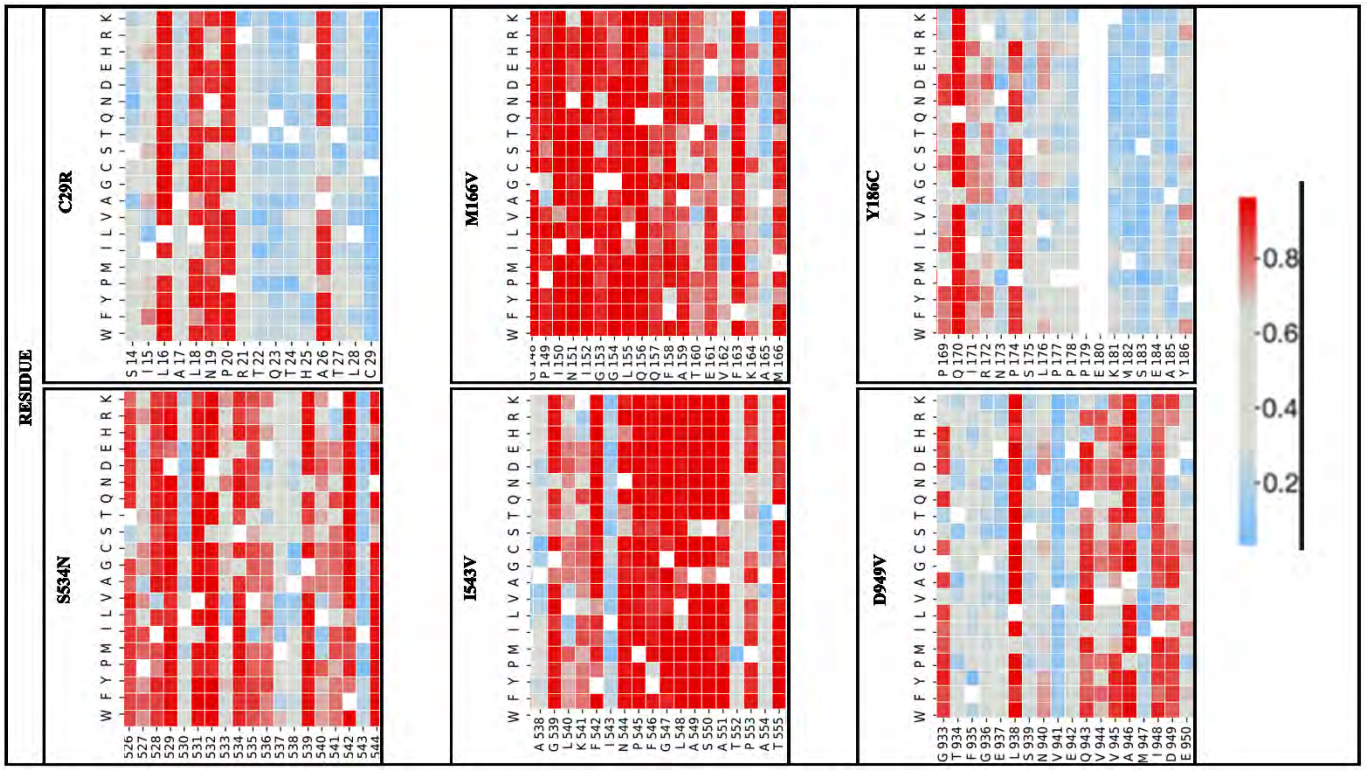


Figure S3.1.2: A heatmap displaying the sequence-based pathogenicity effects probability for clinically relevant mutations (C29R, M166V, Y186C, I543V, S534N, and D949V) generated by the Variant Annotation Integrator (VARITY). The heatmap is color-coded from blue to white, representing a range from 0 to 1, indicating an ascending likelihood of pathogenicity. Mutations assigned a value of 0.5 on the heatmap are considered moderately pathogenic, while those assigned a value of 1 are highly likely to be pathogenic. Conversely, mutations with a value of 0 are less likely to be pathogenic.

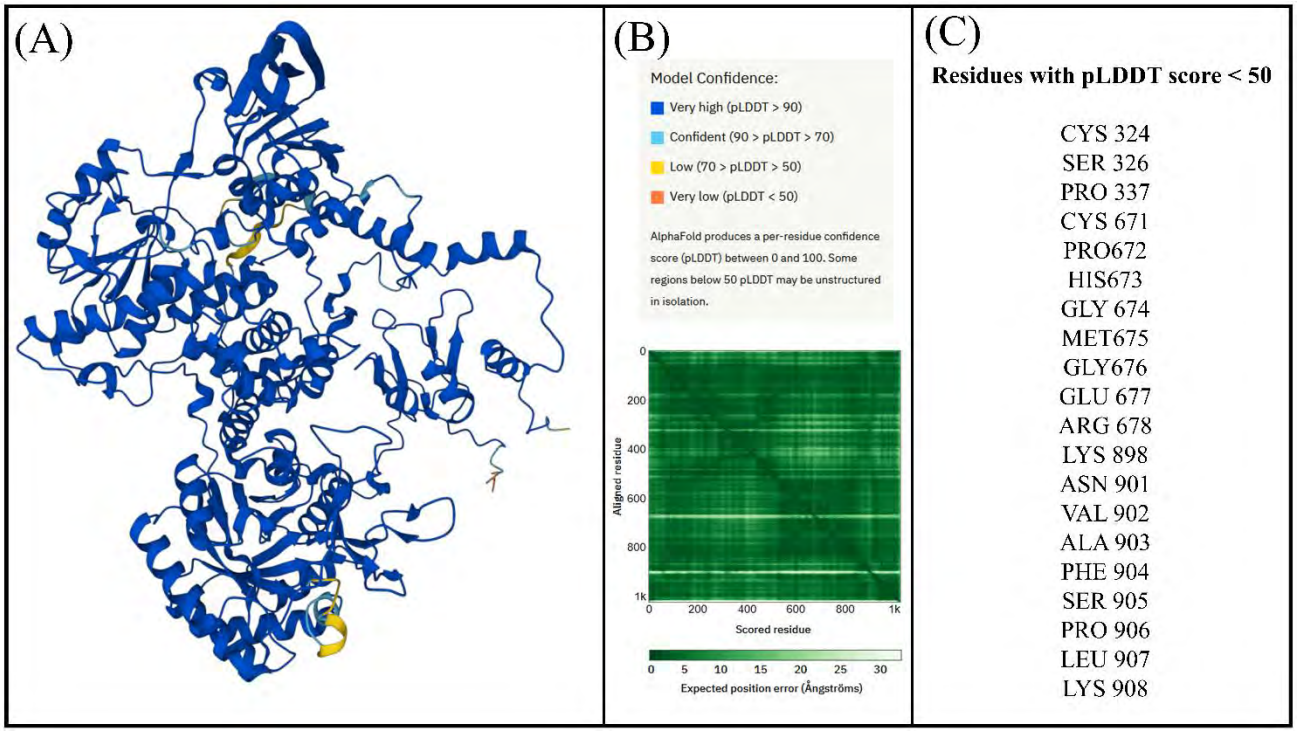


Figure S3.1.3: Represents the prediction of the Dihydropyrimidine dehydrogenase [NADP(+)] structure using AlphaFold. (A) The AlphaFold algorithm generates a confidence score (pLDDT) for each residue, ranging from 0 to 100. Regions with pLDDT below 50 may indicate potential disordered regions or regions that exhibhits higher dynamic. (B) A heatmap visualizes the expected position error per residue. (C) Residues with a pLDDT score below 50 are highlighted.

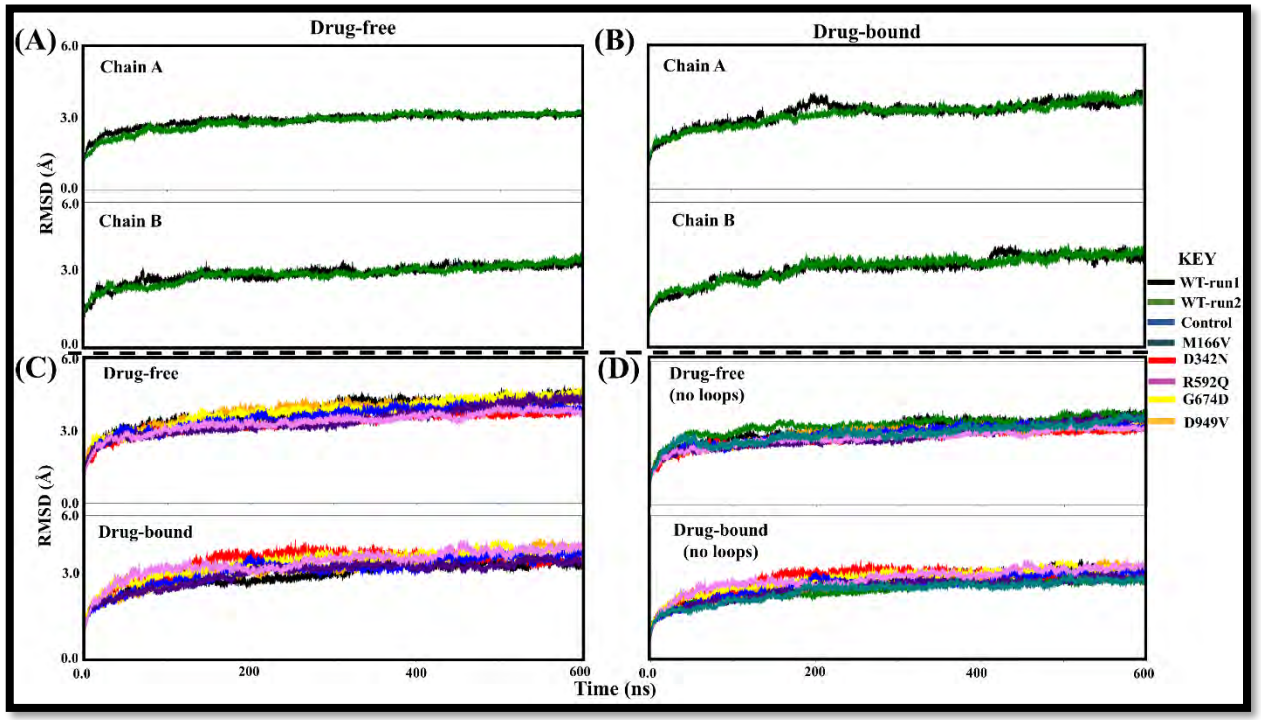


Figure S3.2.1: A line graph illustration of different 600 ns simulations represented as root mean square of deviation (RMSD). The duplicate WT runs are represented by black and green. A) Duplicate RMSD runs for drug free Monomer A (Monomer A) and Monomer B (Monomer B) B) WT systems**. B)** Duplicate RMSD runs of Monomer A (Monomer A) and Monomer B drug bound (Monomer B) systems. **C)** RMSD representation of entire protein (Monomer A and Monomer B) in drug free and drug bound systems **D)** RMSD depicting the entire protein without loop regions for drug free and drug bound systems

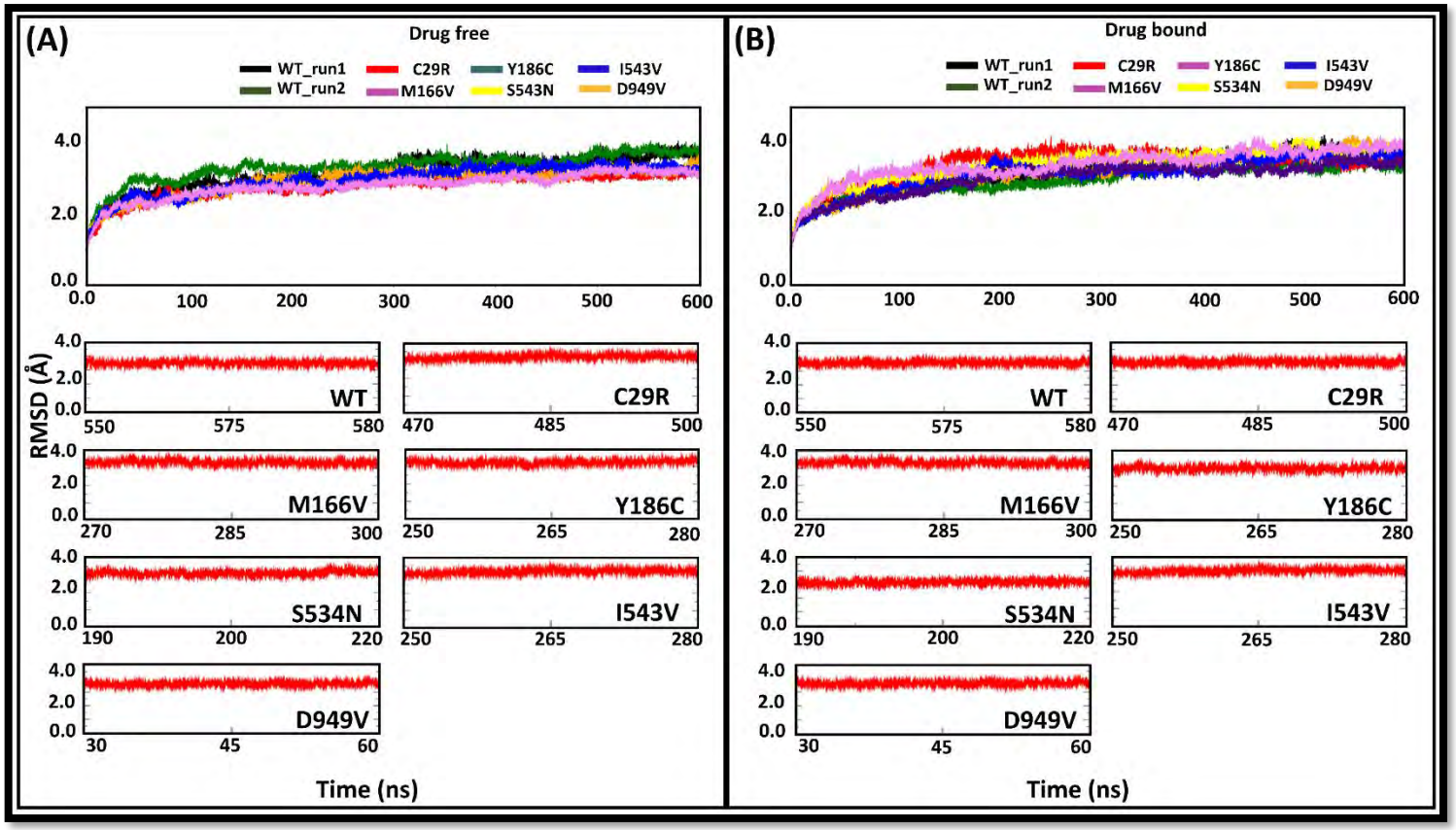


Figure S3.2.2. Line graph depicting 600 ns RMSD simulation for drug free (inactivated) and drug bound (activated) systems. The red lines represent samples of the protein's equilibrated regions for **A)** the drug free and **B)** drug bound ensembles, which were extracted for direct network analysis (DRN) calculation.

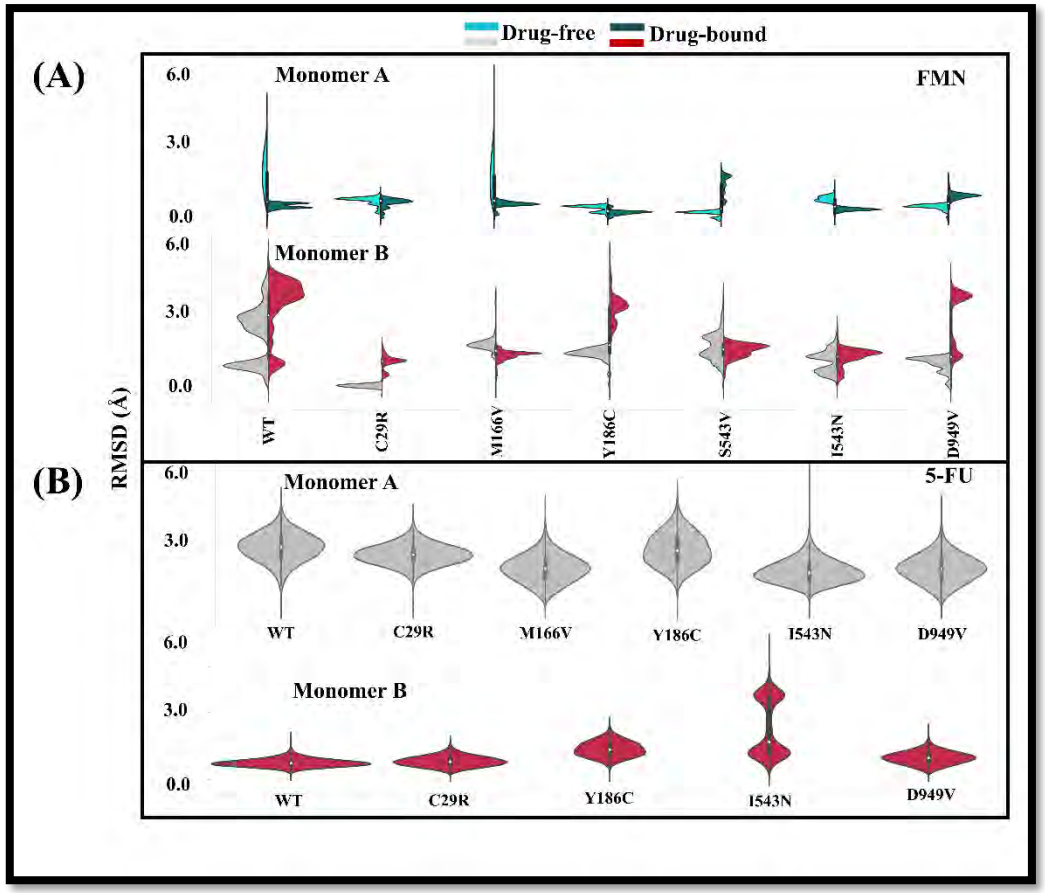


Figure S3.2.3. split Violin plots depicting 600 ns RMSD simulation for drug free and drug bound systems of cofactor (FMN) and ligand (5-FU)

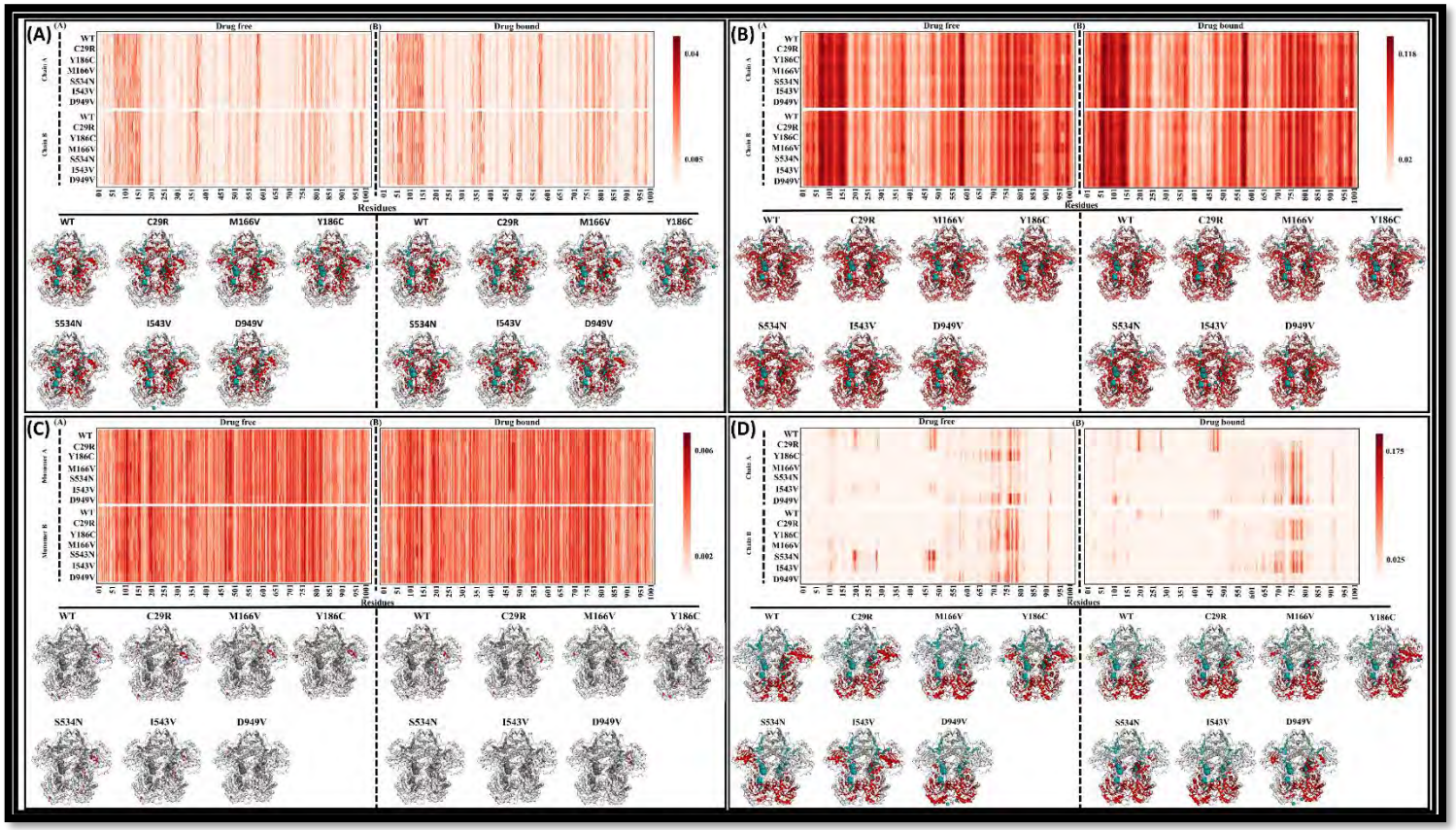


Figure S3.3.1. **Heatmap and 3D general mapping of high centrality residues A) *BC* representation B) *CC* representation C) *EC* representation and D) *KC* representation.** The blue components in the 3D figures represents the oxidoreduction catalytic channel (NADPH, FAD, $Fe^{2+}{}_4S^{-2}{}_4$ clusters FMN and 5-FU)

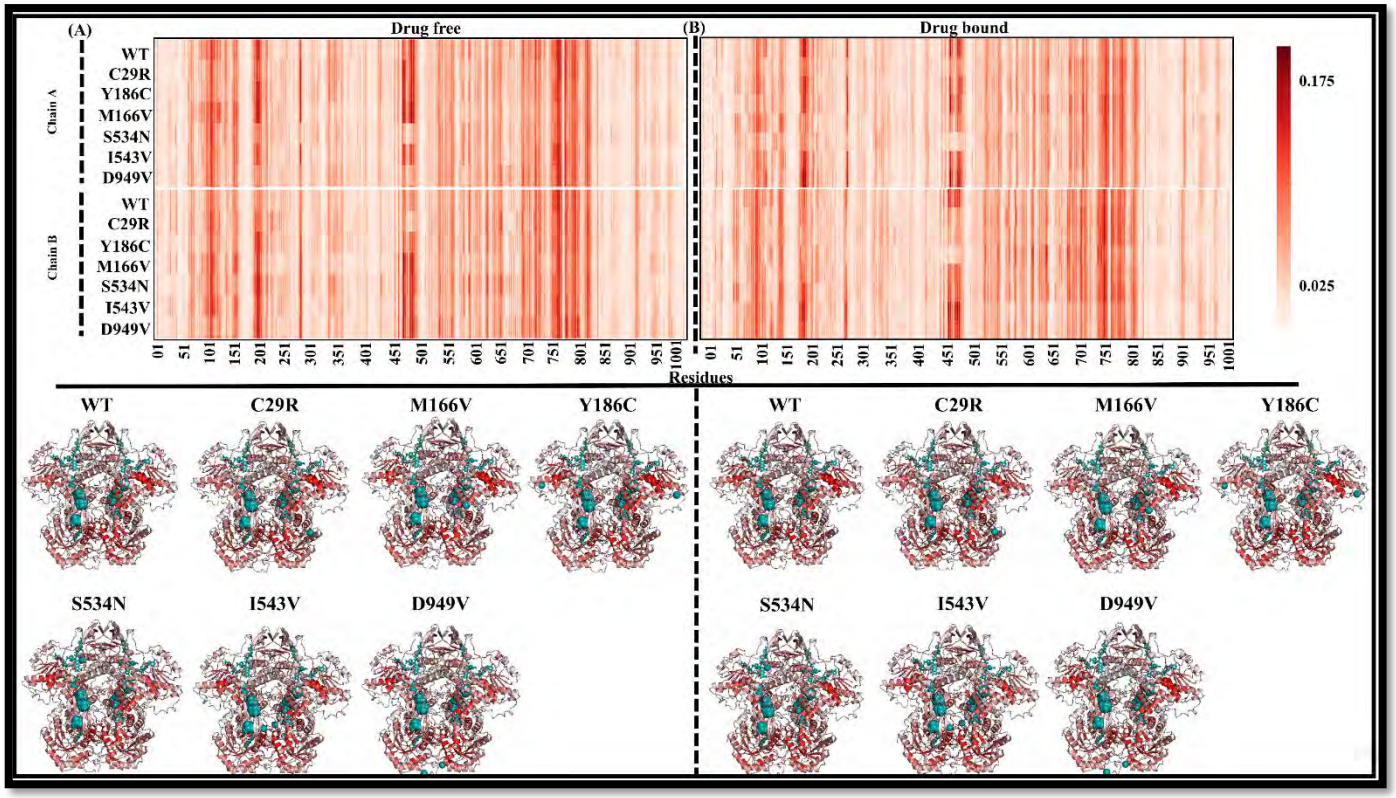


Figure S3.3.2. **Heatmap and 3D general mapping of high degree *KC* residues. In the heatmaps, x-axis represents residue numbers and y-axis represents WT and mutations systems.** The blue components in the 3D figures represents the oxidoreduction catalytic channel (NADPH, FAD, $Fe^{2+}{}_4S^{-2}{}_4$ clusters FMN and 5-FU).

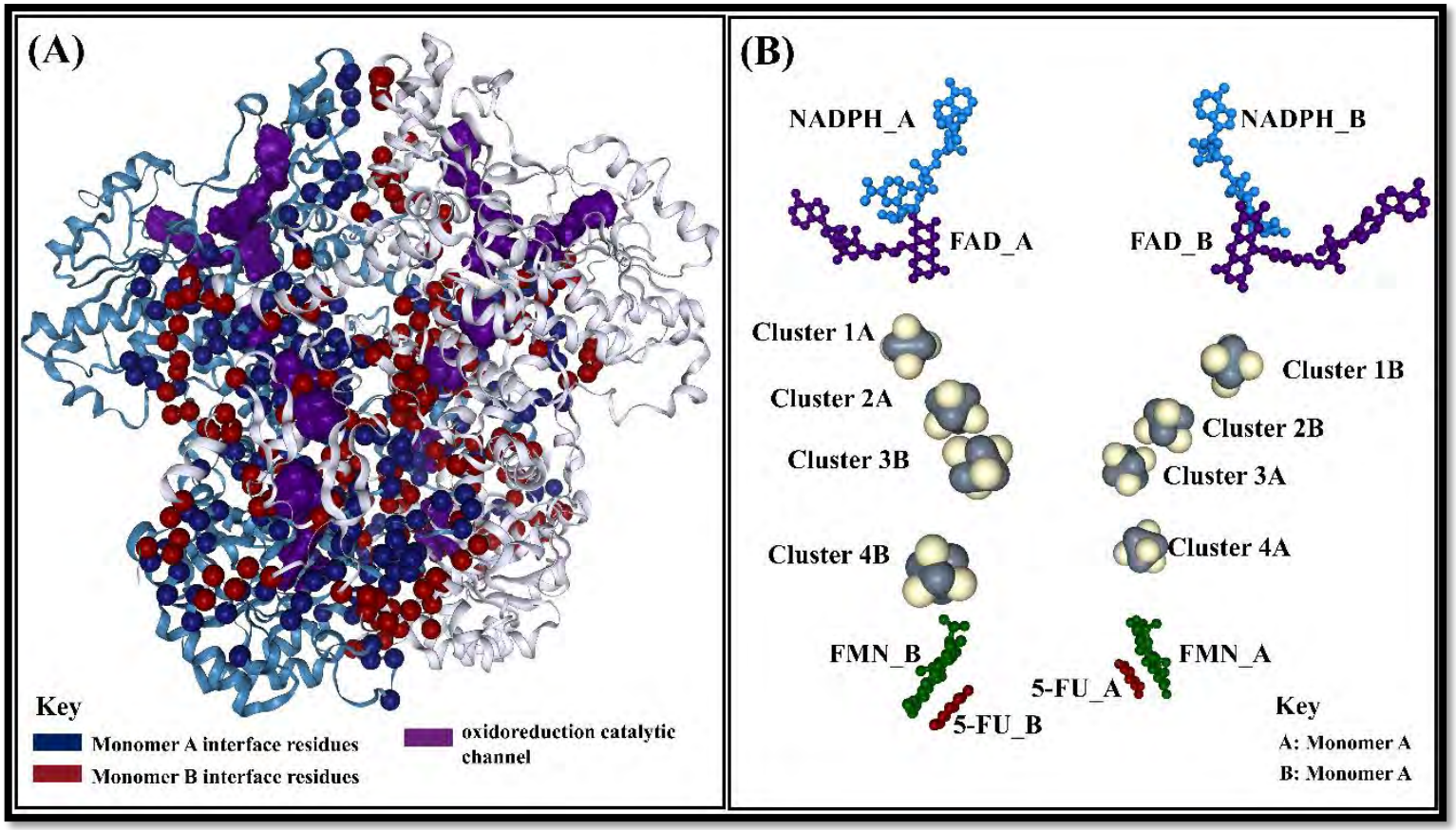


Figure S3.4. 3D Illustration of A) Interface residues for Monomer A in blue, Monomer B in red and oxidoreduction catalytic channel represented in indigo. **B)** Oxidoreduction channel whereby NADPH is blue, FAD is in indigo, $Fe^{2+}{}_4S^{-2}{}_4$ clusters in grey and yellow, FMN in green and 5-FU in maroon

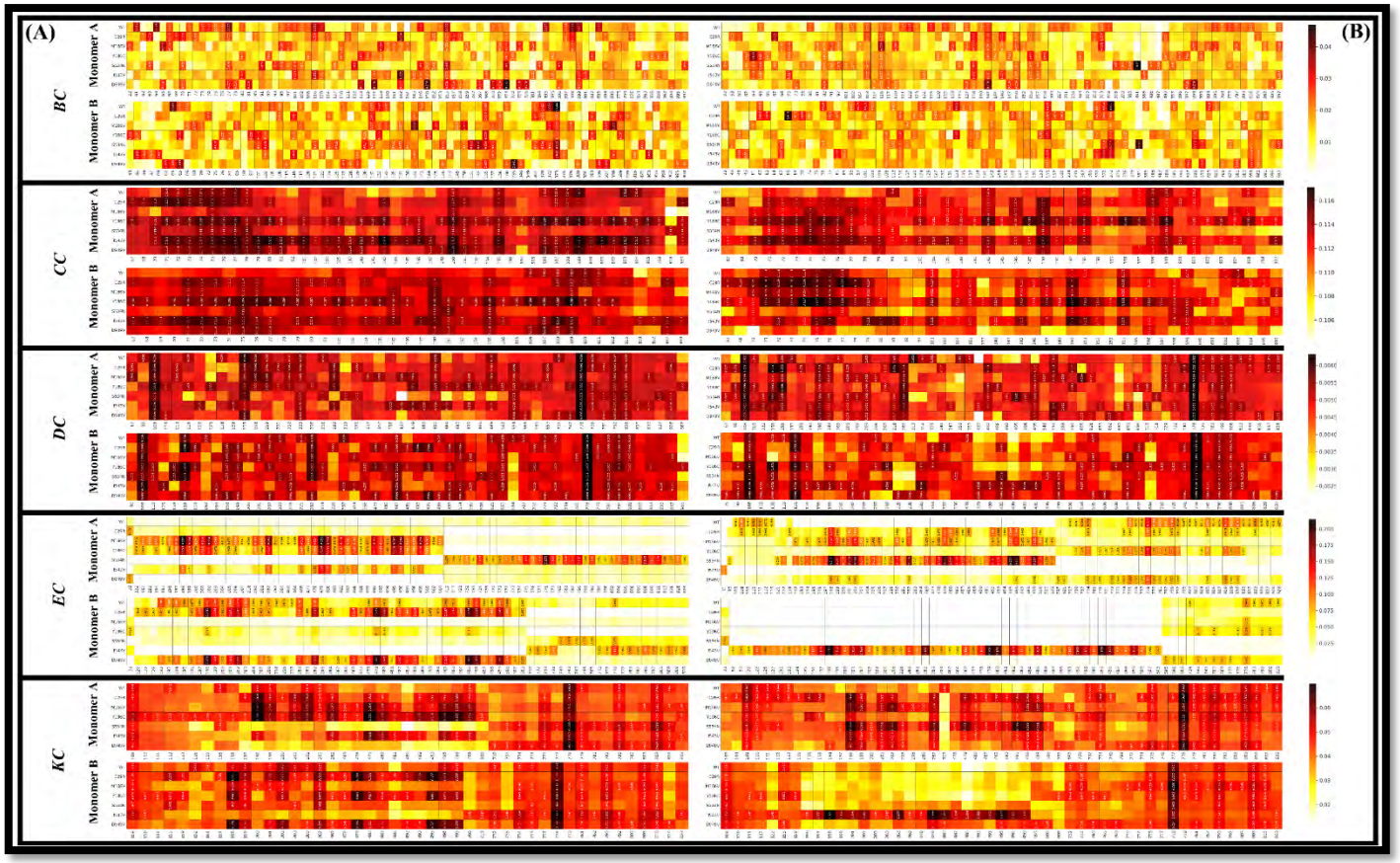


Figure S3.5.1. Centrality metrics heatmaps representation of the A) drug free and B) drug bound ensemble showing the top 2.0 % global significant hubs (x-axis) of the WT and six mutations (y-axis). The hubs are color-coded from low to high value (light-yellow to dark-red) and annotated accordingly.

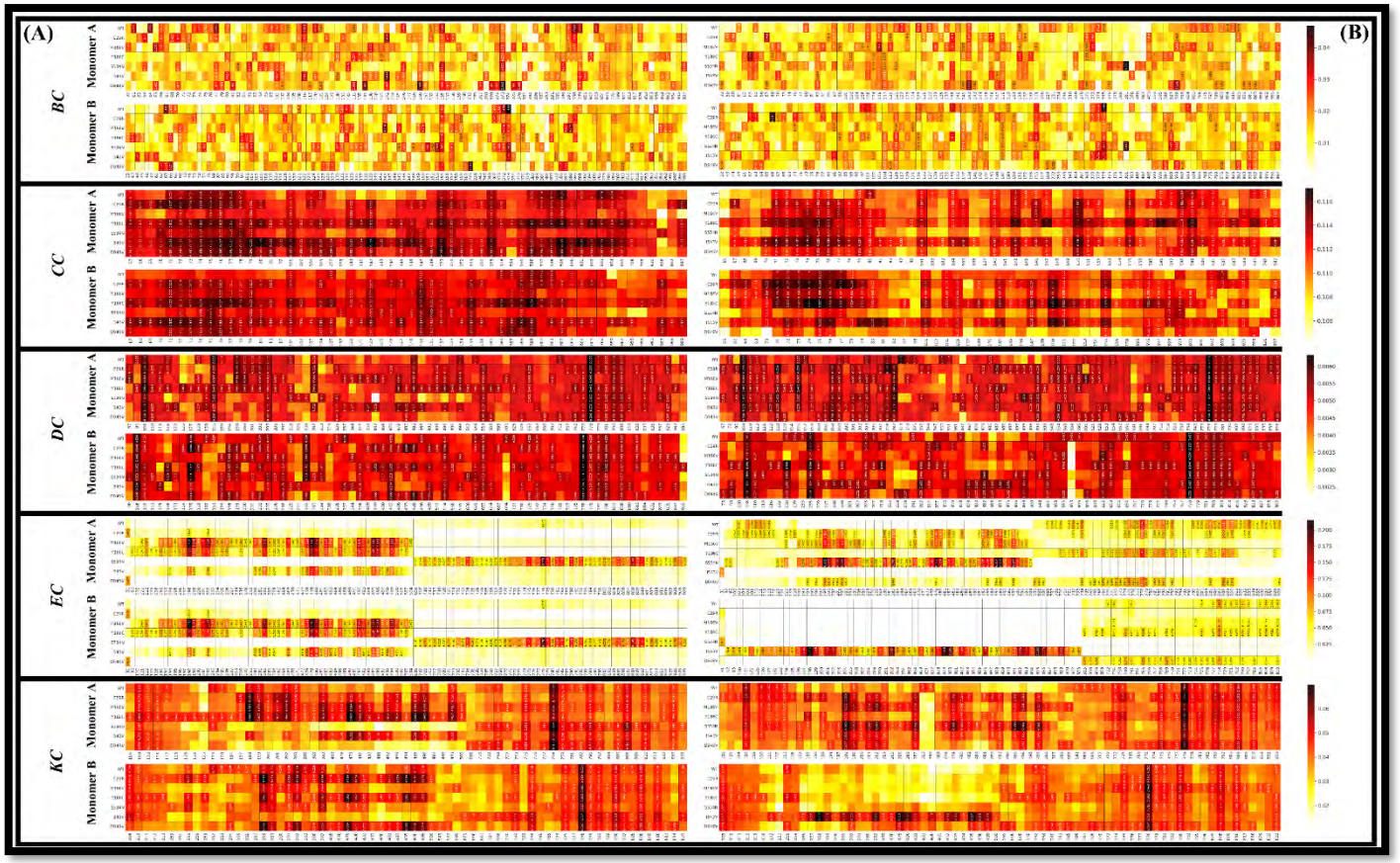


Figure 3.5.2. **Centrality metrics heatmaps representation of the A) drug free and B) drug bound ensemble showing the top 3.0 % global significant hubs (x-axis) of the WT and six mutations (y-axis).** The hubs are color-coded from low to high value (light-yellow to dark-red) and annotated accordingly.

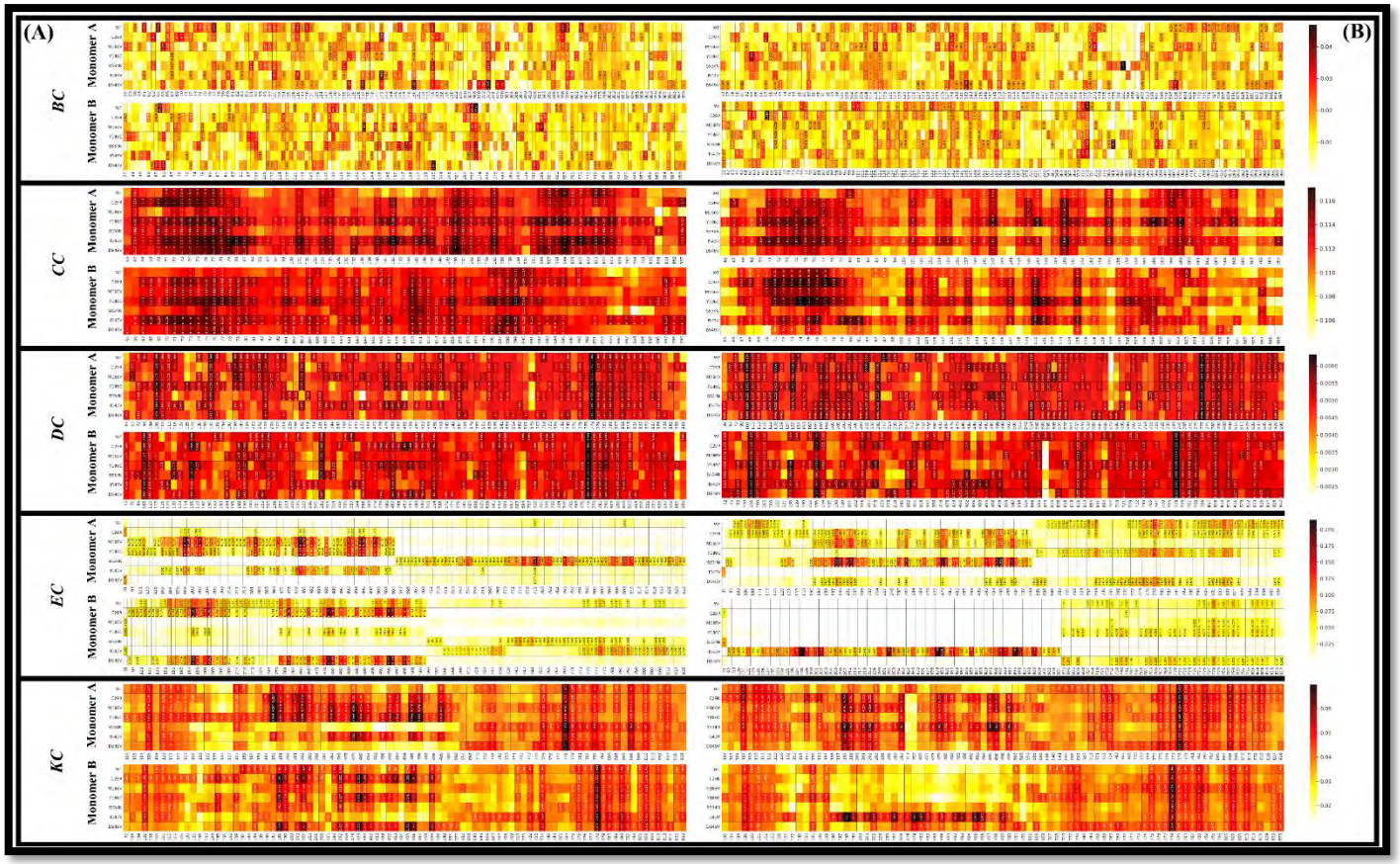


Figure S3.5.3. Centrality metrics heatmaps representation of the A) drug free and B) drug bound ensemble showing the top 4.0 % global significant hubs (x-axis) of the WT and six mutations (y-axis). The hubs are color-coded from low to high value (light-yellow to dark-red) and annotated accordingly.

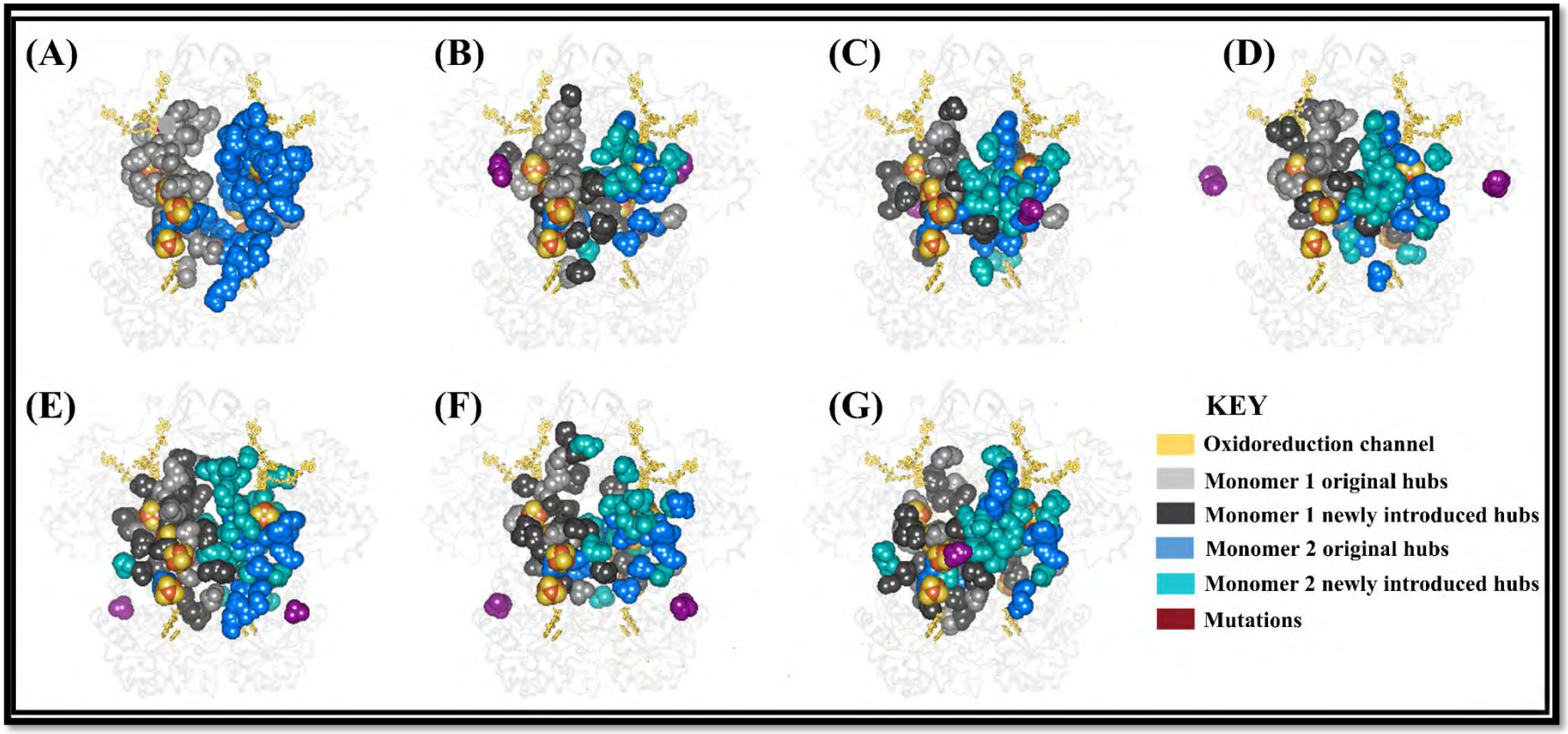


Figure S3.6.1. 3D representation of averaged *betweenness centrality* hubs of the activated (drug free) ensemble showing a pair of communication pathway proximal to oxidoreduction channel (Monomer A NADPH (yellow) to FAD (yellow) to 2 x $Fe^{2+}{}_4S^{2-}{}_4$ (yellow) clusters, proceeding to Monomer B 2 x $Fe^{2+}{}_4S^{2-}{}_4$ (yellow) to FMN (yellow) for the reduction of 5-FU (yellow)) and vice-versa. The mutations are colored in deep-purple while A = WT, B = C29R, C = M166V, D = Y186C, E = S543N, F = I543V and G = D949V

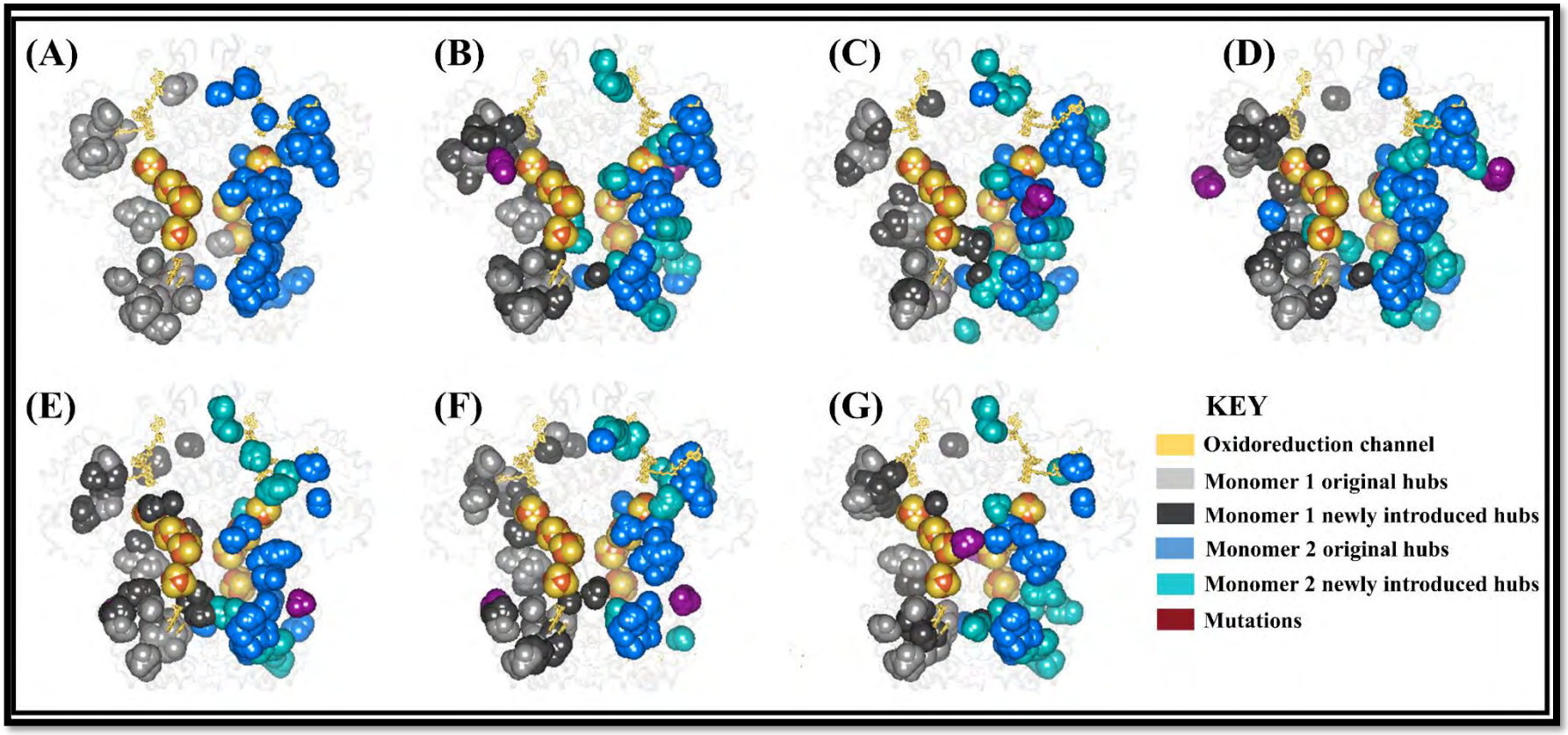


Figure S3.6.2. 3D representation of averaged *degree centrality* hubs of the activated (drug free) ensemble showing a pair of communication pathway proximal to oxidoreduction channel (Monomer A NADPH (yellow) to FAD (yellow) to 2 x $Fe^{2+}{}_4S^{2-}{}_4$ (yellow) clusters, proceeding to Monomer B 2 x $Fe^{2+}{}_4S^{2-}{}_4$ (yellow) to FMN (yellow) for the reduction of 5-FU (yellow)) and vice-versa. The mutations are colored in deep-purple while A = WT, B = C29R, C = M166V, D = Y186C, E = S543N, F = I543V and G = D949V

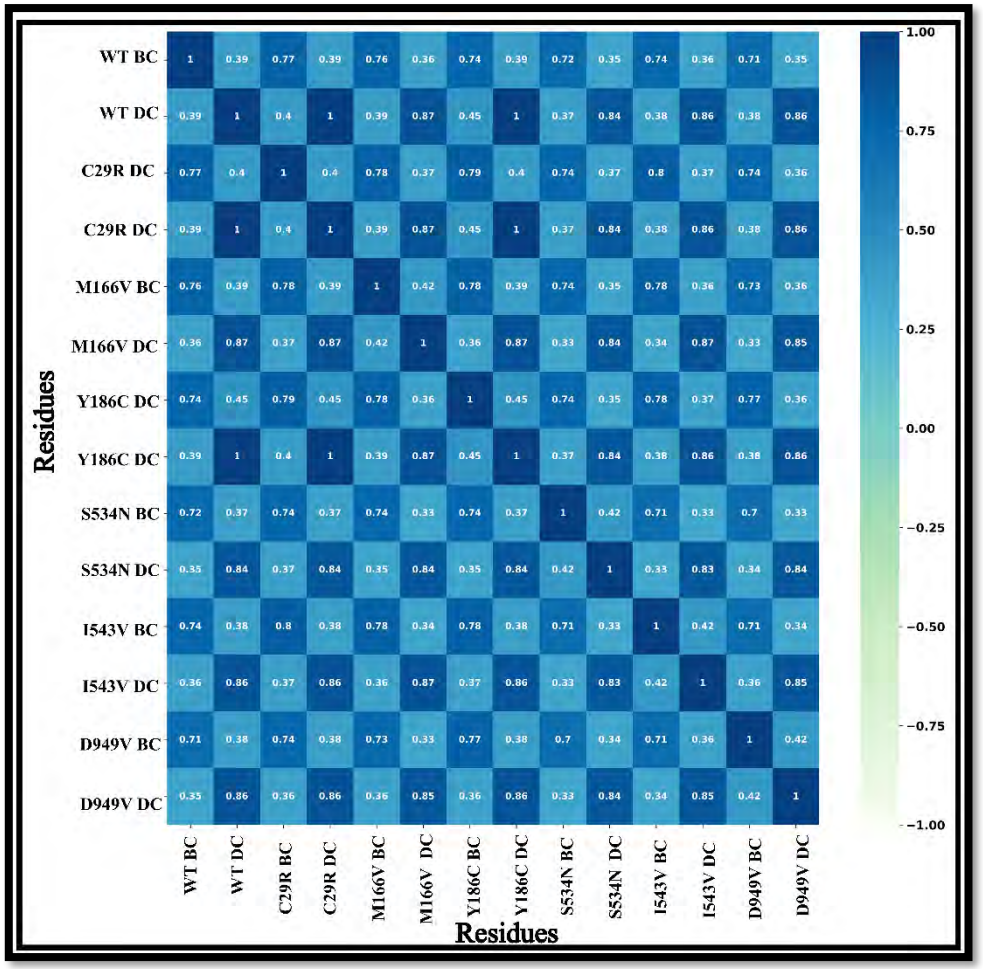


Figure S3.6.3. A direct cross-correlation (DCC) heatmap showing the relationship between WT and mutations in DPD drug free ensembles.

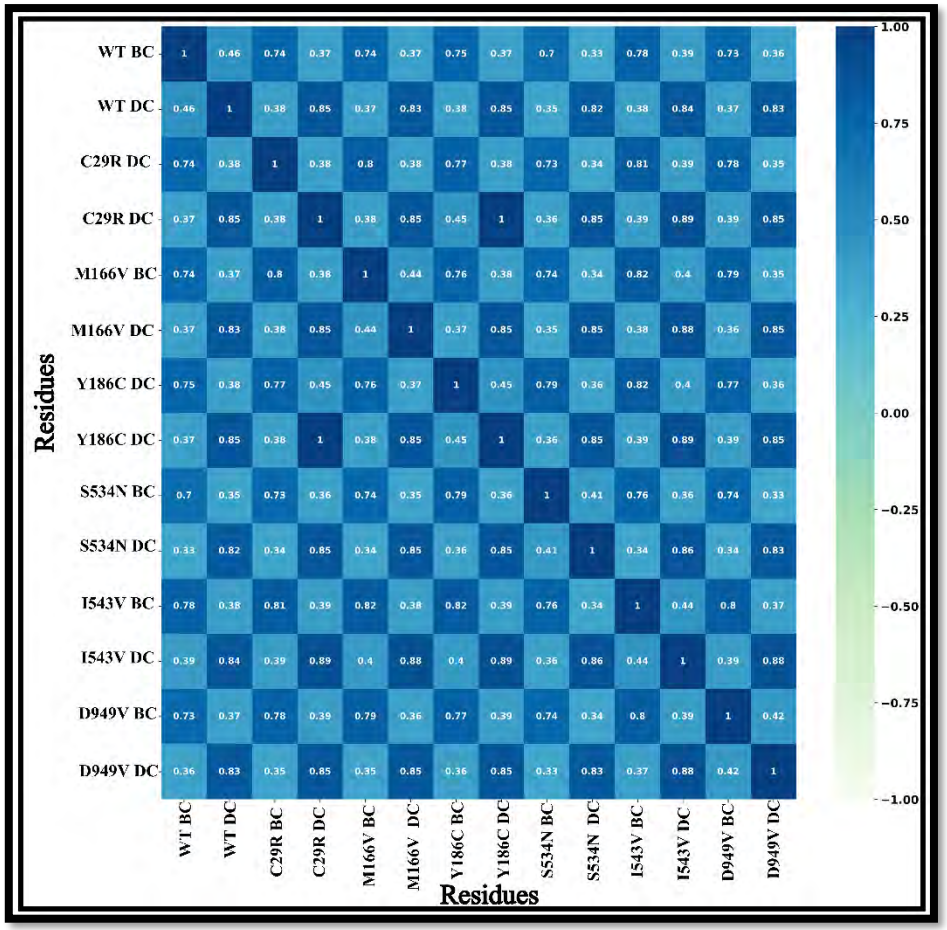


Figure S3.6.4. **A direct cross-correlation (DCC) heatmap showing the relationship between WT and mutations in DPD drug bound ensembles.**

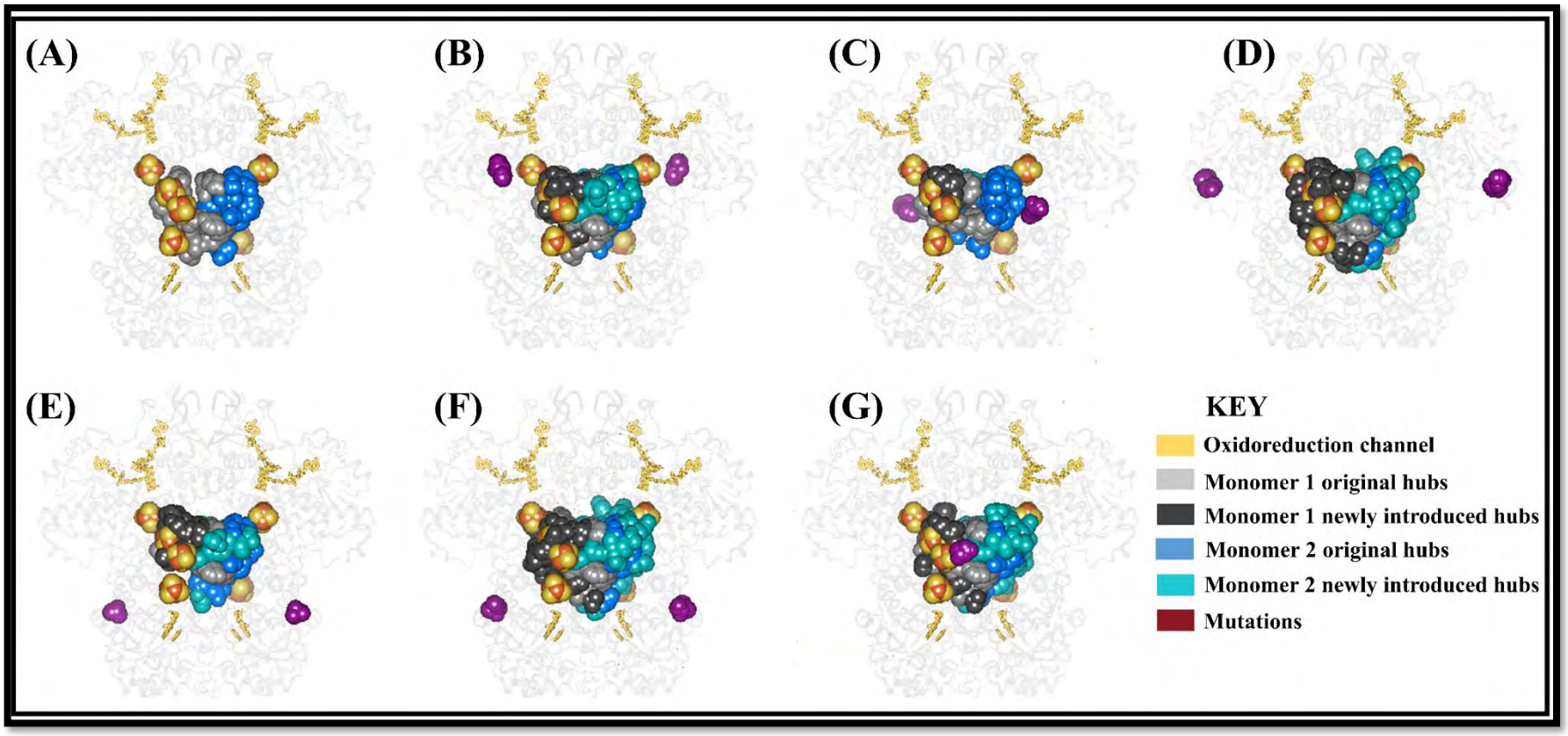


Figure S3.6.5. 3D representation of averaged *closeness centrality* hubs of the activated (drug free) ensemble showing a pair of communication pathway proximal to oxidoreduction channel (Monomer A NADPH (yellow) to FAD (yellow) to 2 x $Fe^{2+}{}_4S^{2-}{}_4$ (yellow) clusters, proceeding to Monomer B 2 x $Fe^{2+}{}_4S^{2-}{}_4$ (yellow) to FMN (yellow) for the reduction of 5-FU (yellow)) and vice-versa. The mutations are colored in deep-purple while A = WT, B = C29R, C = M166V, D = Y186C, E = S543N, F = I543V and G = D949V

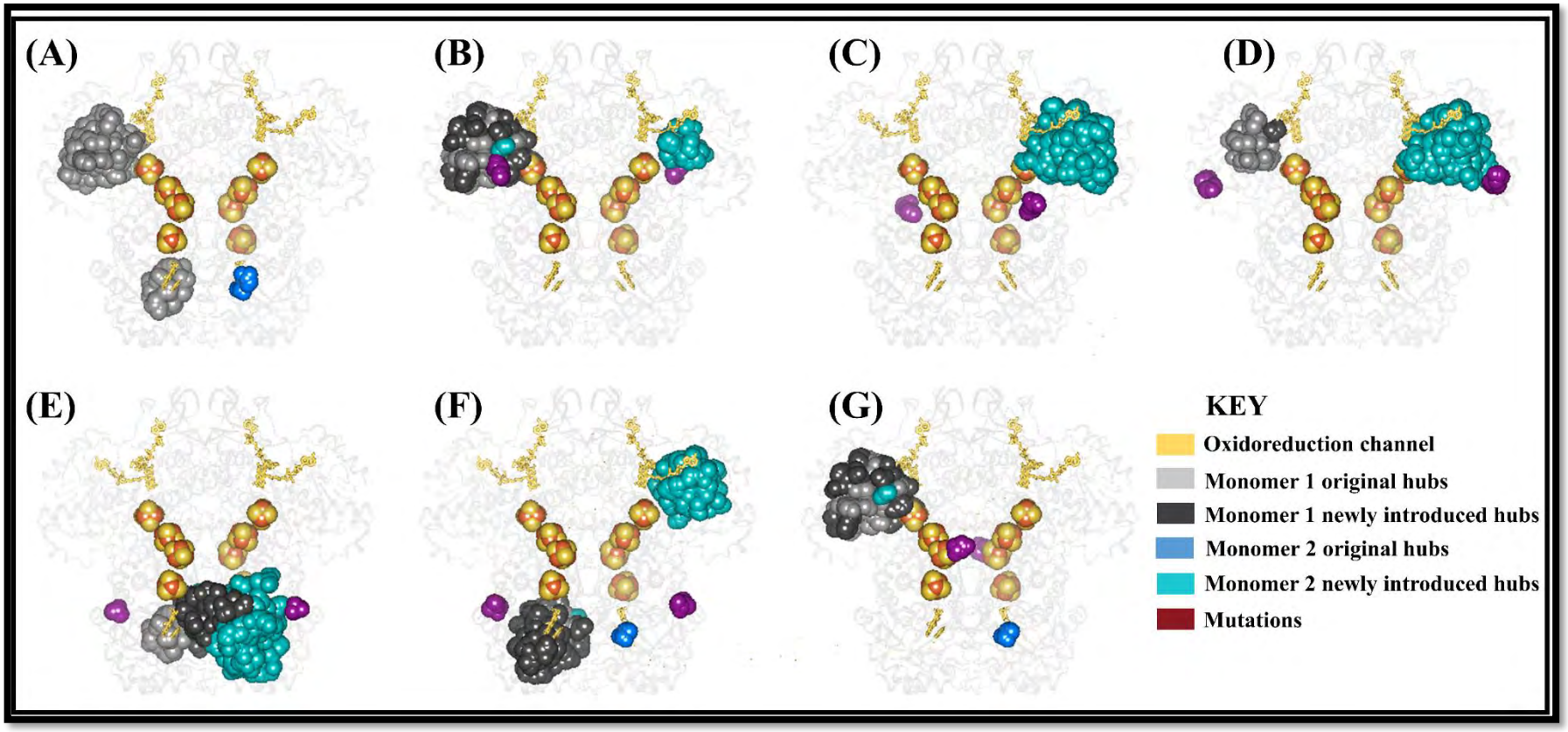


Figure S3.6.6. 3D representation of averaged *eigen centrality* hubs of the activated (drug free) ensemble showing a pair of communication pathway proximal to oxidoreduction channel (Monomer A NADPH (yellow) to FAD (yellow) to 2 x $Fe^{2+}{}_4S^{2-}{}_4$ (yellow) clusters, proceeding to Monomer B 2 x $Fe^{2+}{}_4S^{2-}{}_4$ (yellow) to FMN (yellow) for the reduction of 5-FU (yellow)) and vice-versa. The mutations are colored in deep-purple while A = WT, B = C29R, C = M166V, D = Y186C, E = S543N, F = I543V and G = D949V

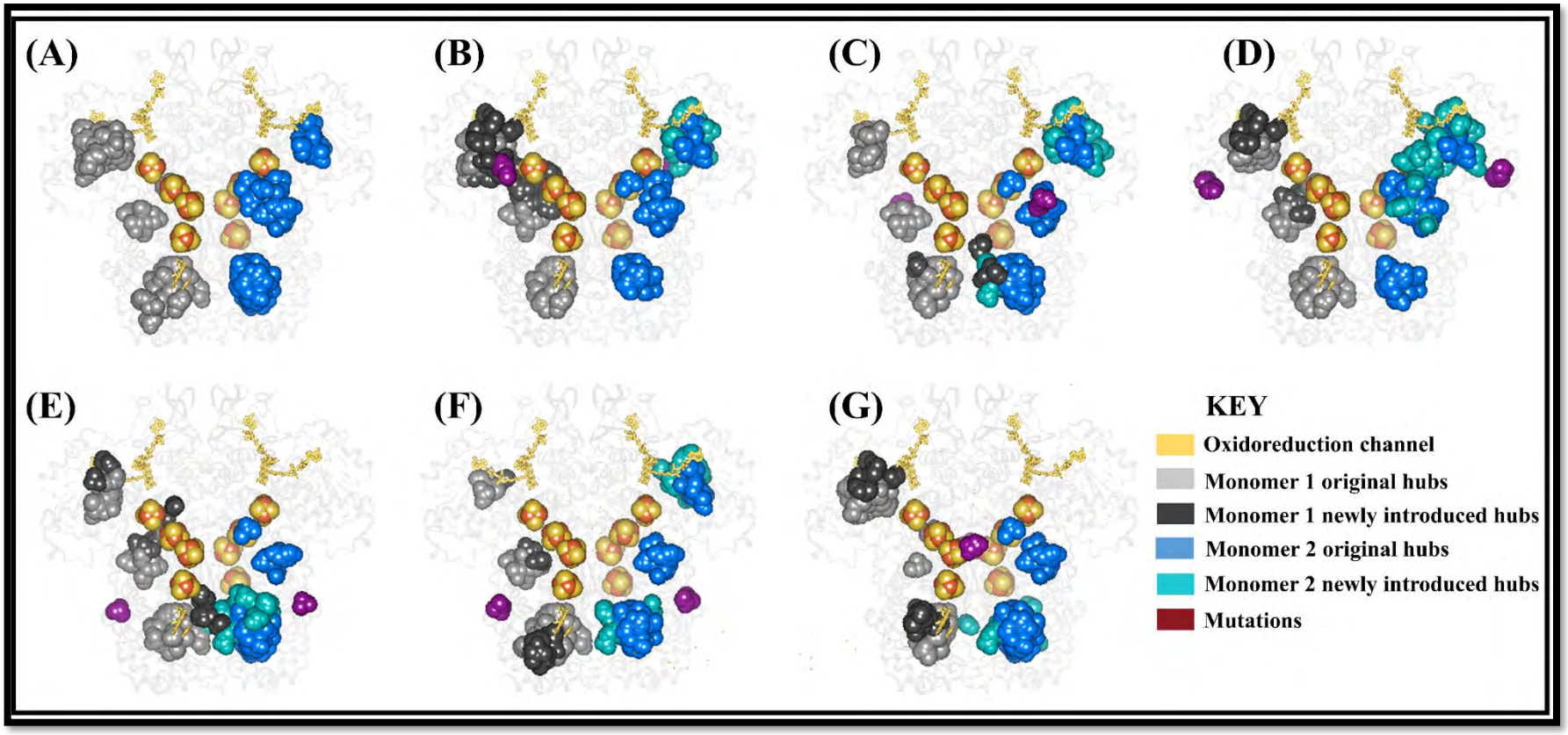


Figure S3.6.7. 3D representation of averaged *eigen centrality* hubs of the activated (drug free) ensemble showing a pair of communication pathway proximal to oxidoreduction channel (Monomer A NADPH (yellow) to FAD (yellow) to 2 x $Fe^{2+}{}_4S^{2-}{}_4$ (yellow) clusters, proceeding to Monomer B 2 x $Fe^{2+}{}_4S^{2-}{}_4$ (yellow) to FMN (yellow) for the reduction of 5-FU (yellow)) and vice-versa. The mutations are colored in deep-purple while A = WT, B = C29R, C = M166V, D = Y186C, E = S543N, F = I543V and G = D949V

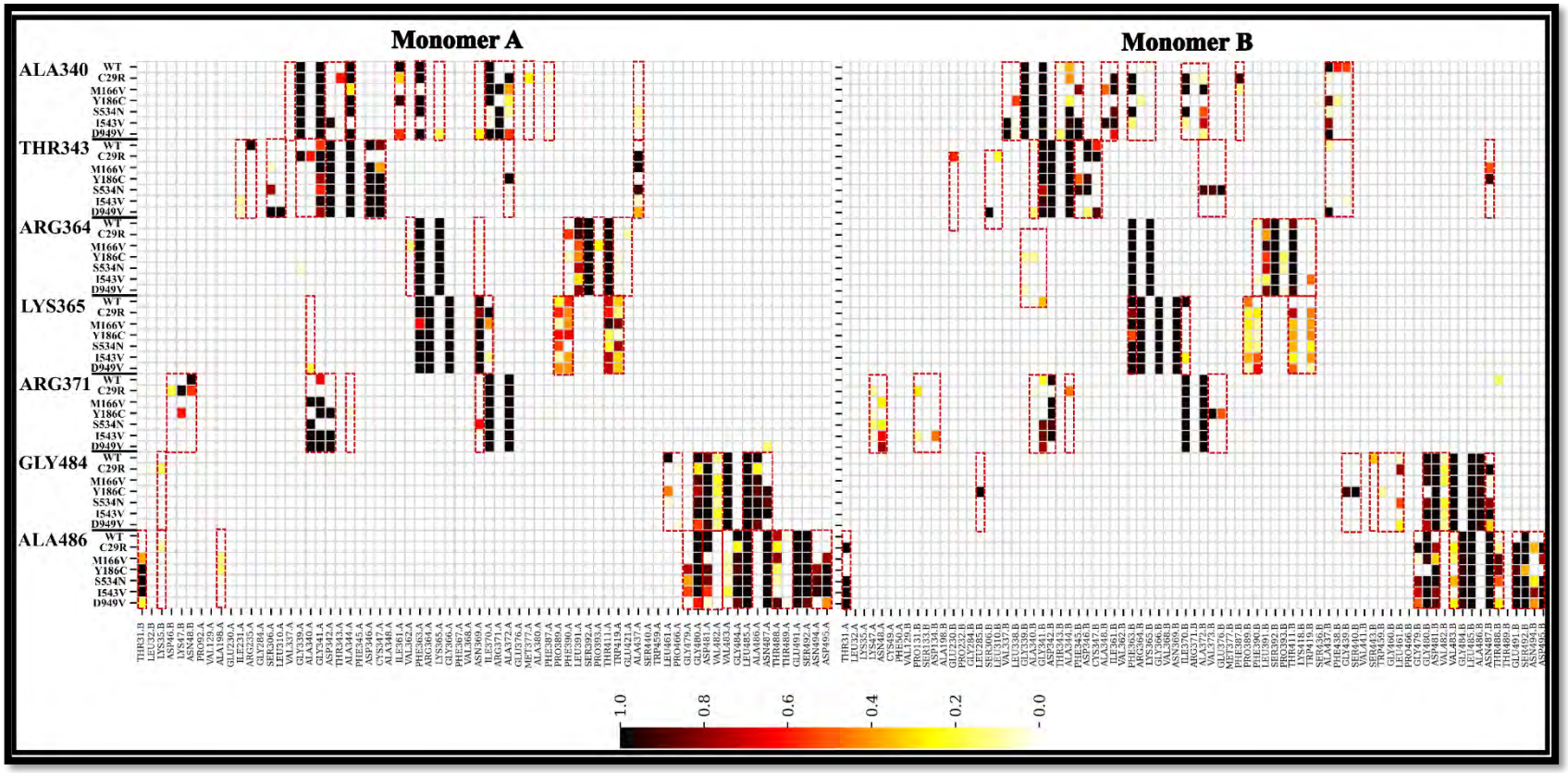


Figure S3.7.1: Activated DPD (drug bound) protein residue contact heatmap showing the interaction between Monomer A and Monomer B of NADPH. The shaded boxes in each frame represent significant differences in the frequency of interaction between the cofactor residues and the surrounding contacting residues in the WT versus the mutation. White (0) to dark-red (1.0) represent no-contact (0) to strong-contact (1.0).

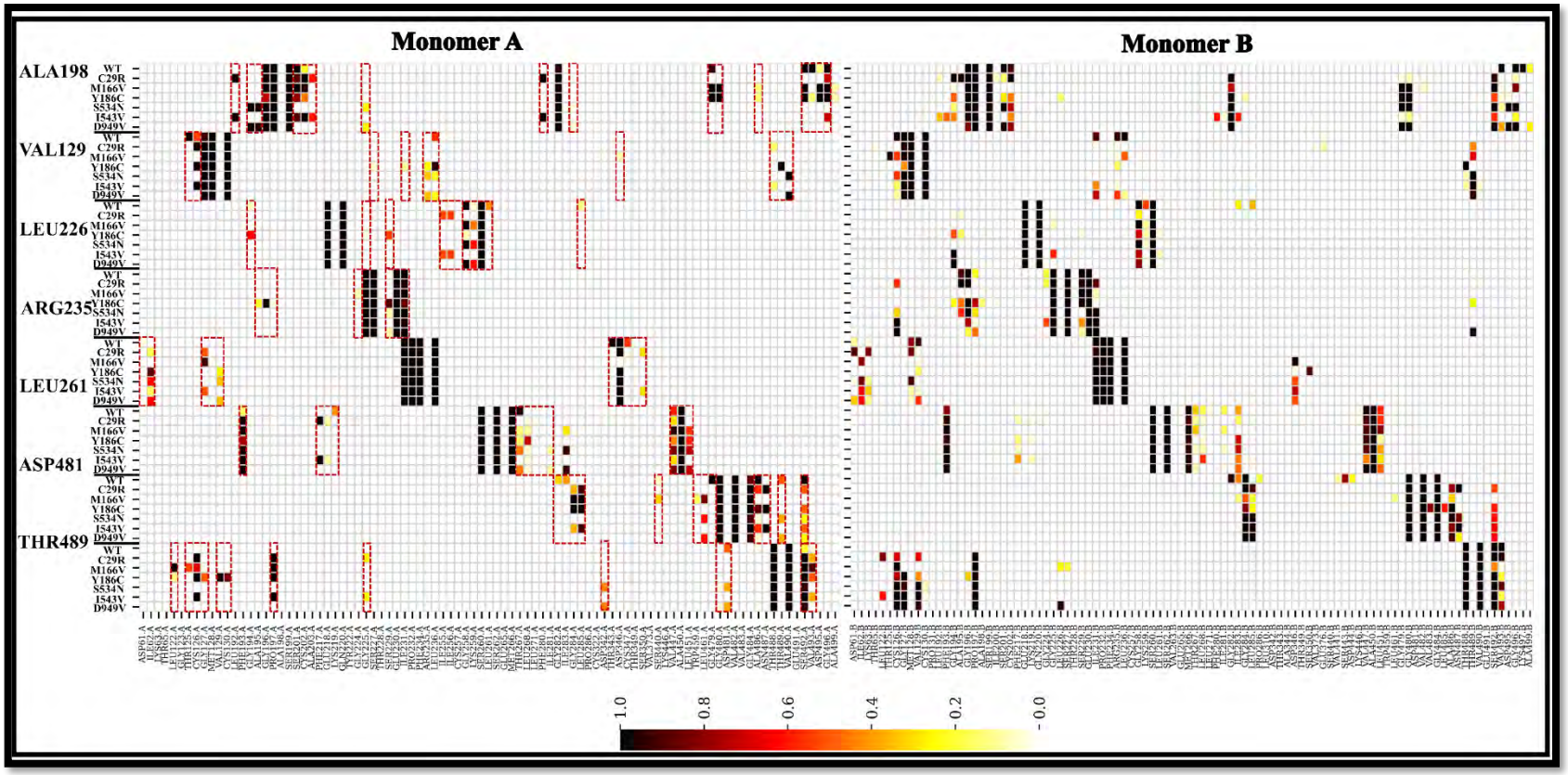


Figure S3.7.2: Activated DPD (drug bound) protein residue contact heatmap showing the interaction between Monomer A and Monomer B of FAD. The shaded boxes in each frame represent significant differences in the frequency of interaction between the cofactor residues and the surrounding contacting residues in the WT versus the mutation. White (0) to dark-red (1.0) represent no-contact (0) to strong-contact (1.0).

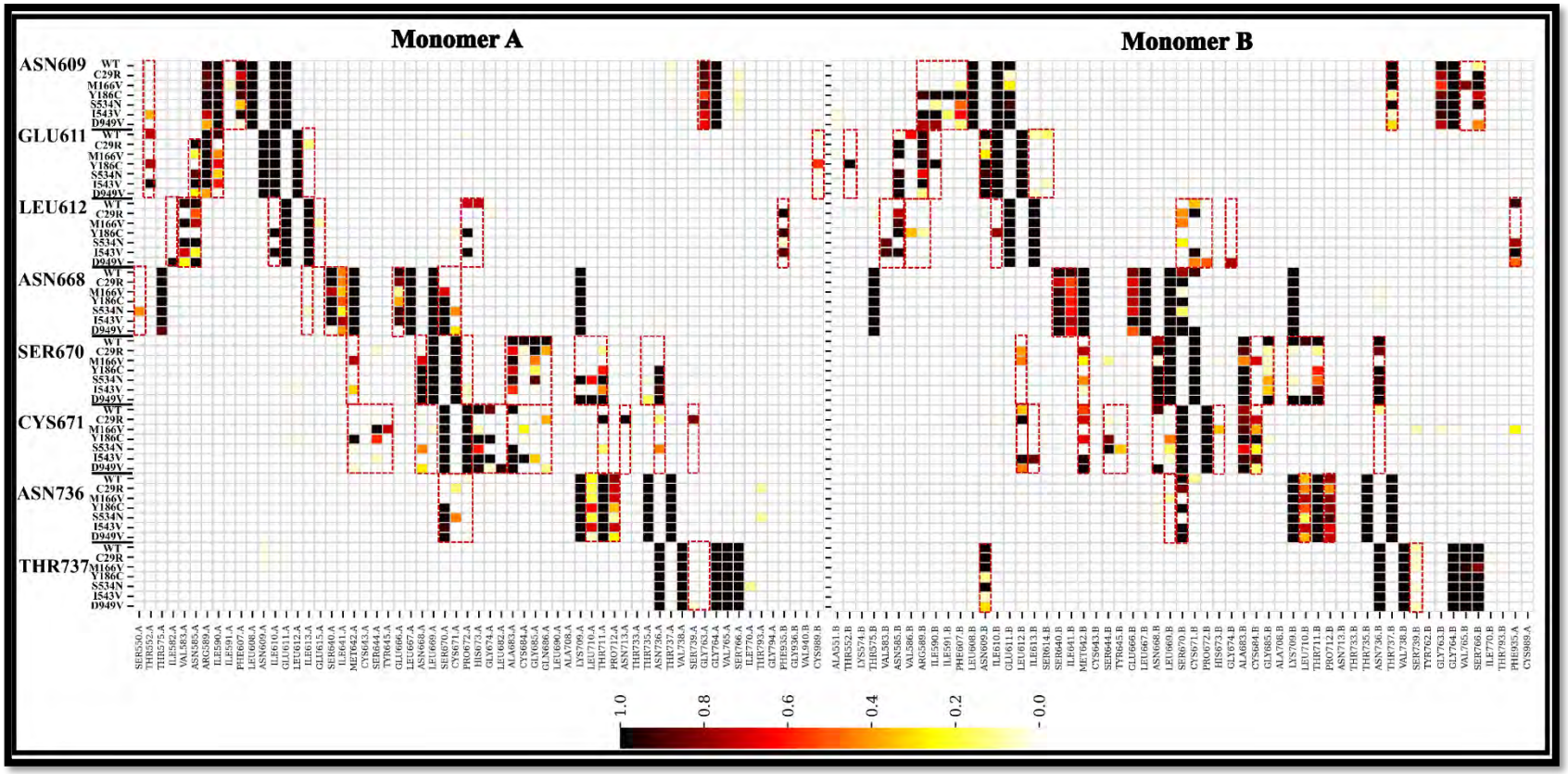


Figure S3.7.3: Activated DPD (drug bound) protein residue contact heatmap showing the interaction between Monomer A and Monomer B of FMN. The shaded boxes in each frame represent significant differences in the frequency of interaction between the cofactor residues and the surrounding contacting residues in the WT versus the mutation. White (0) to dark-red (1.0) represent no-contact (0) to strong-contact (1.0).

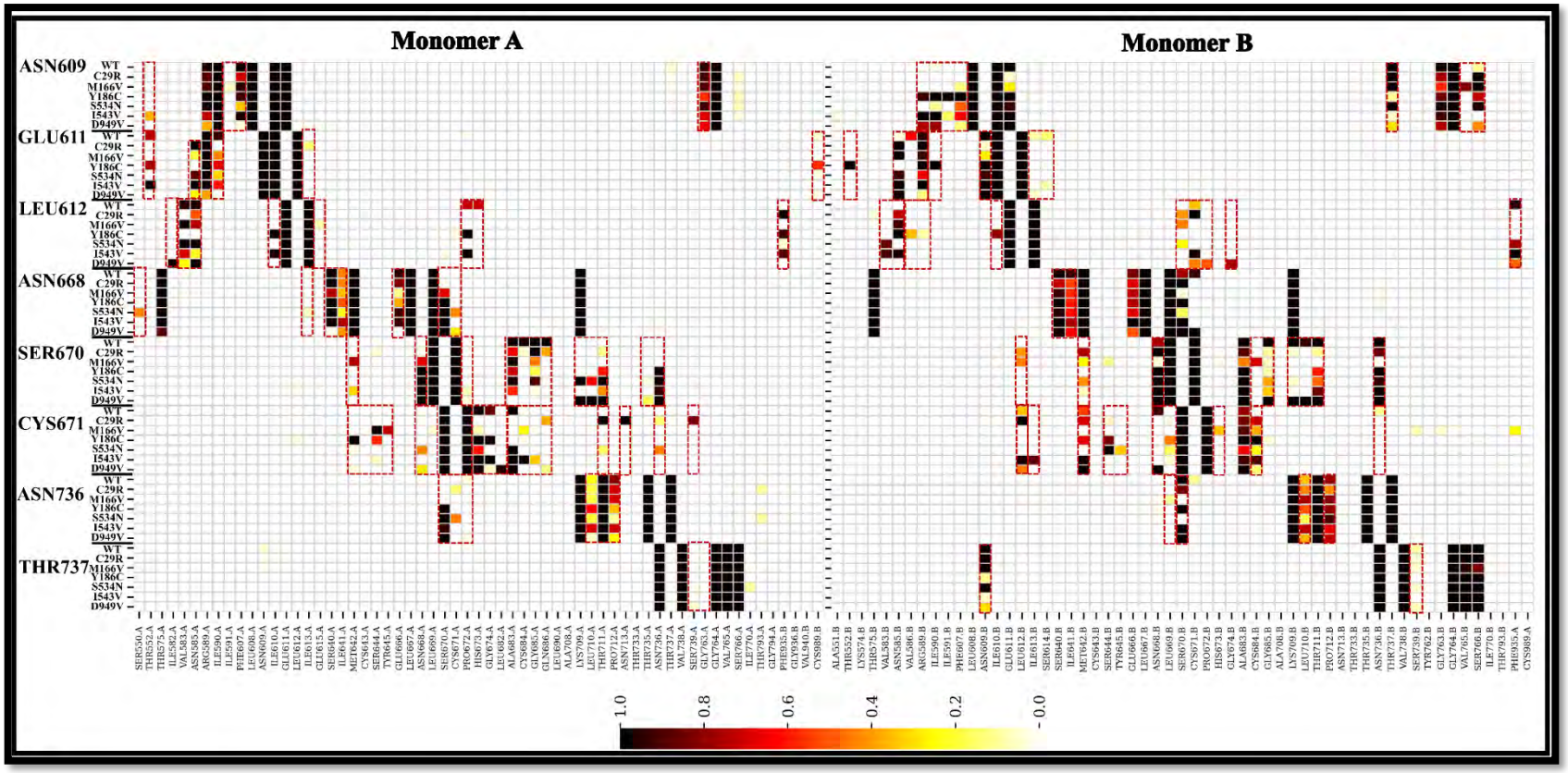


Figure S3.7.4: Activated DPD (drug bound) protein residue contact heatmap showing the interaction between Monomer A and Monomer B of 5-FU. The shaded boxes in each frame represent significant differences in the frequency of interaction between the cofactor residues and the surrounding contacting residues in the WT versus the mutation. White (0) to dark-red (1.0) represent no-contact (0) to strong-contact (1.0).

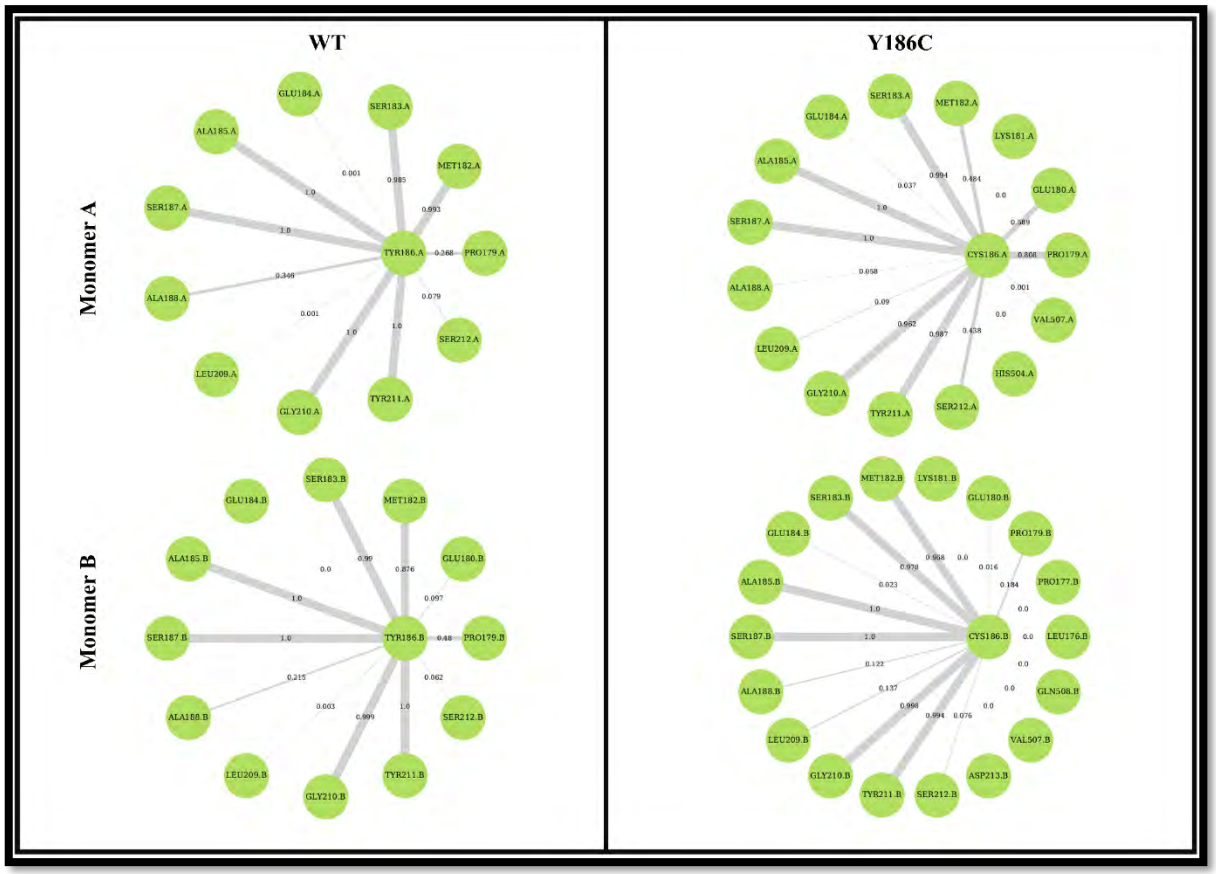


Figure S3.8.1. Contact-map representation of the frequency (%) of residue-residue interaction around the WT in both Monomer A and B versus the mutation Y186C.

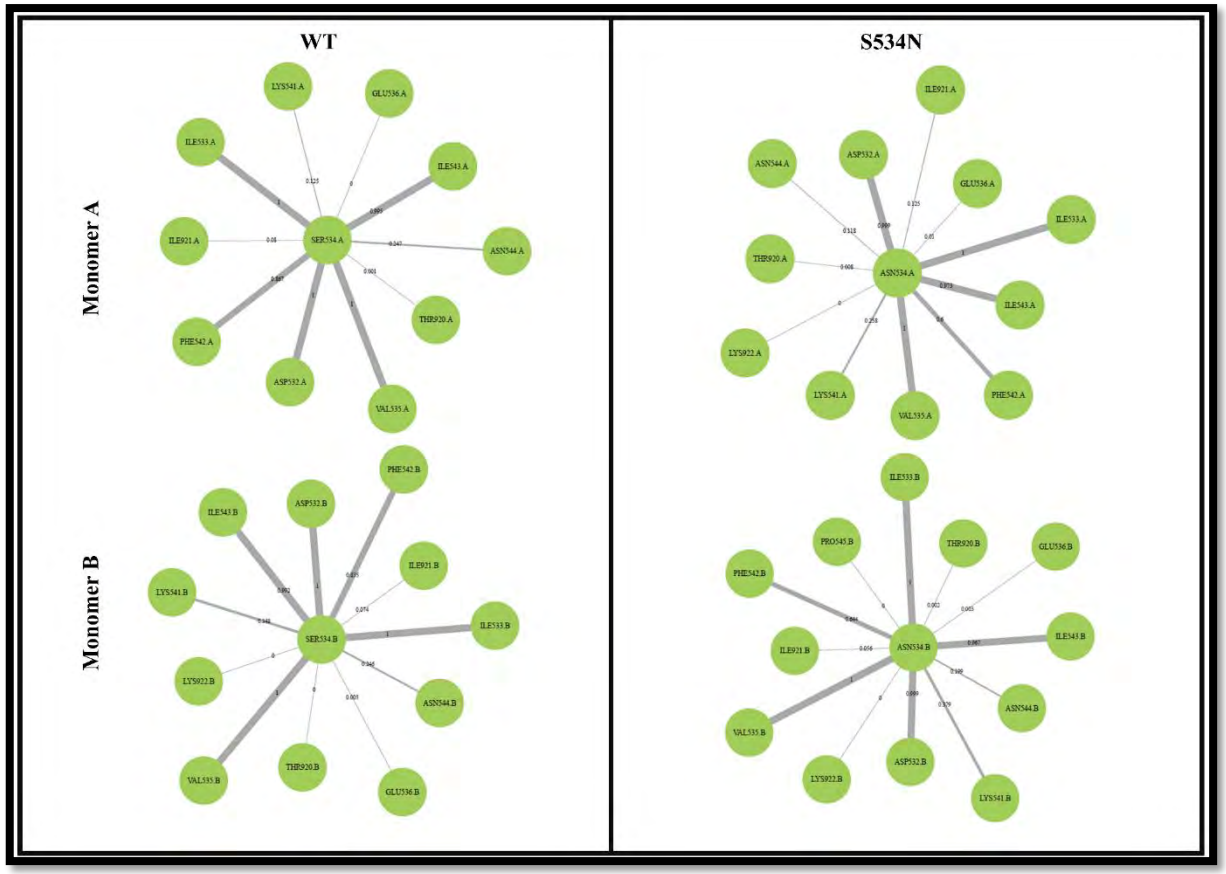


Figure S3.8.2. Contact-map representation of the frequency (%) of residue-residue interaction around the WT in both Monomer A and B versus the mutation S534N.

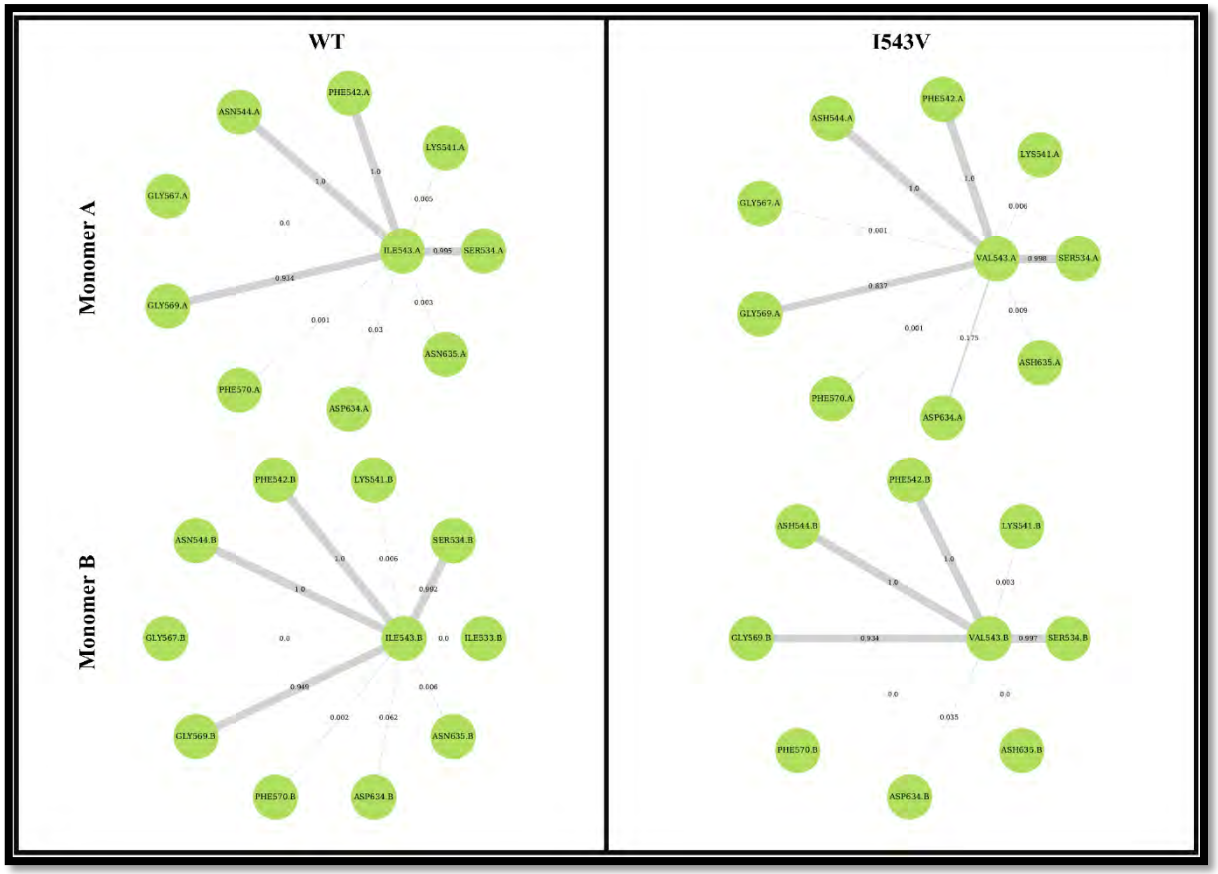


Figure S3.8.3. Contact-map representation of the frequency (%) of residue-residue interaction around the WT in both Monomer A and B versus the mutation I543V.

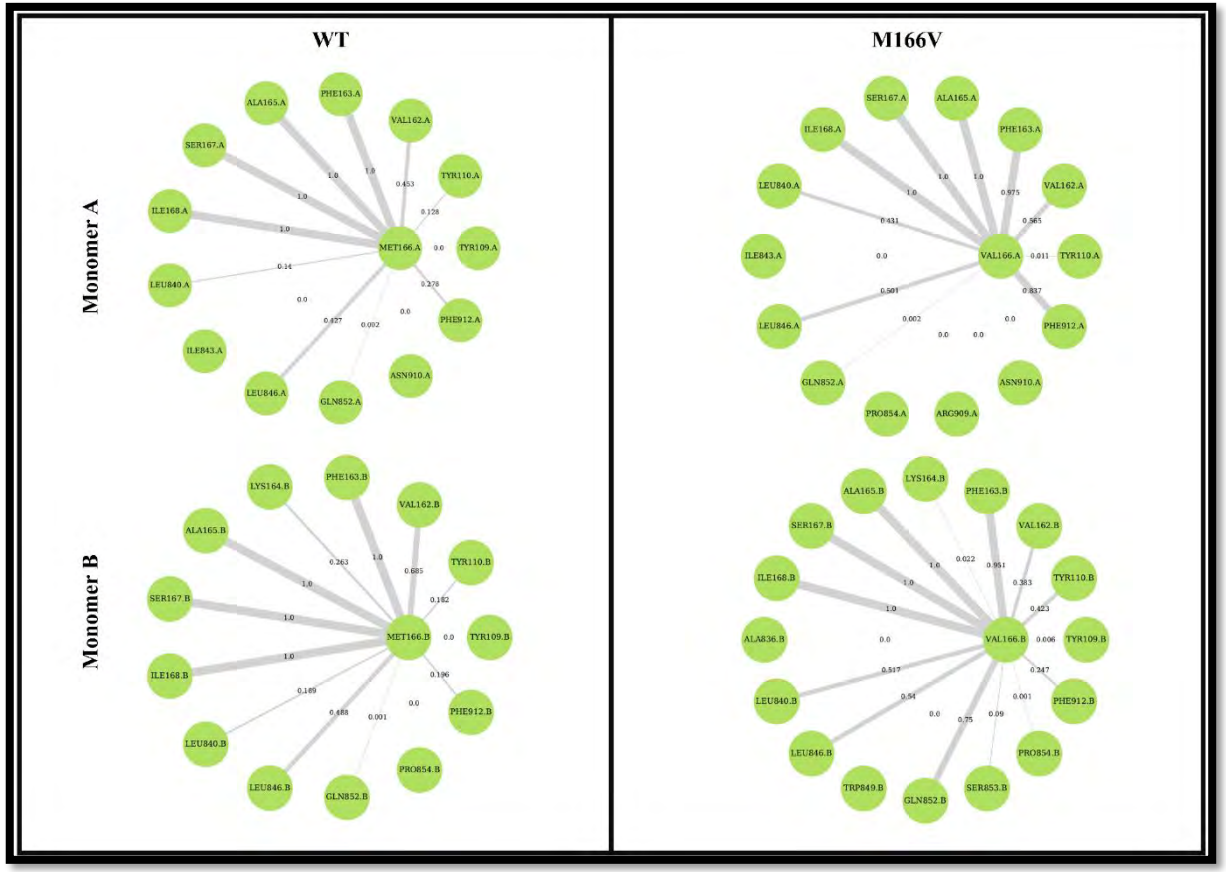


Figure S3.8.4. Contact-map representation of the frequency (%) of residue-residue interaction around the WT in both Monomer A and B versus the mutation M166V.

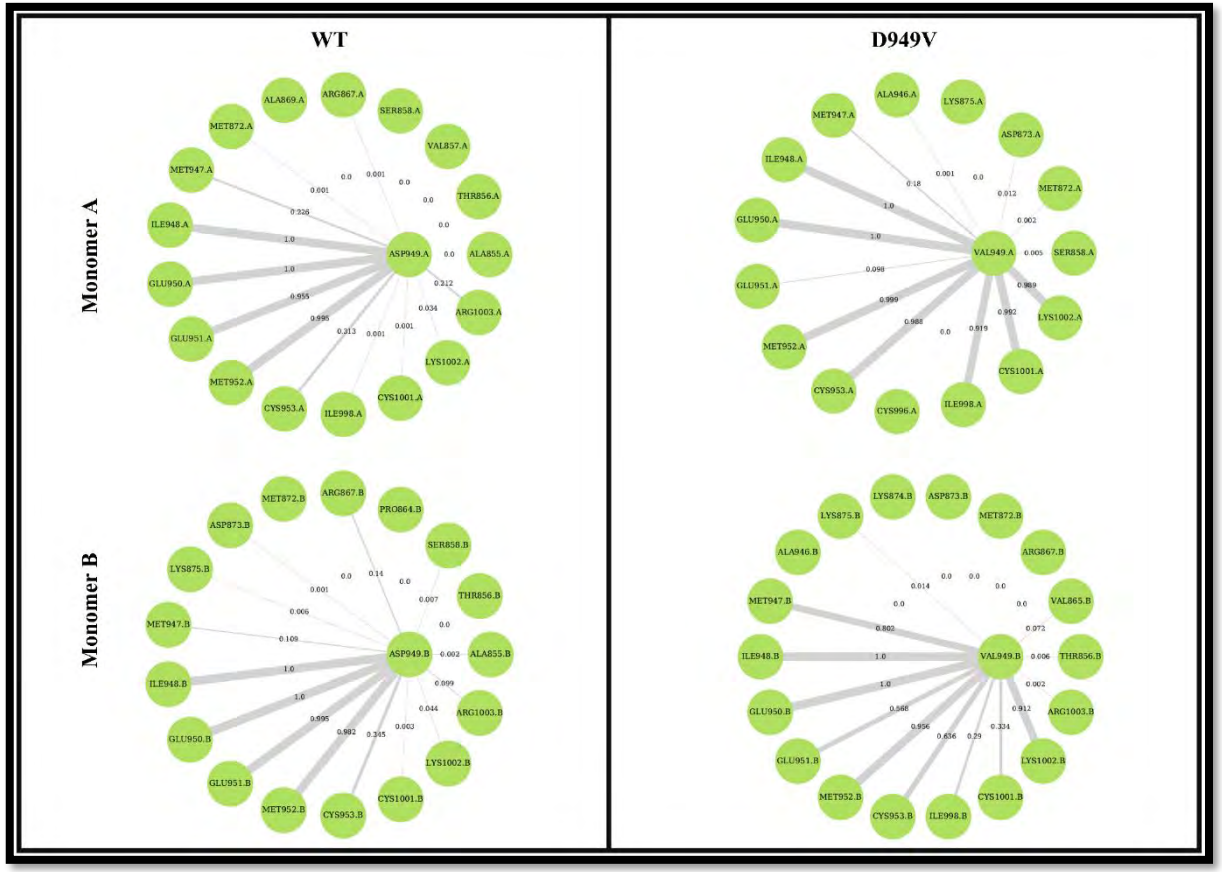


Figure S3.8.5. Contact-map representation of the frequency (%) of residue-residue interaction around the WT in both Monomer A and B versus the mutation D949V.

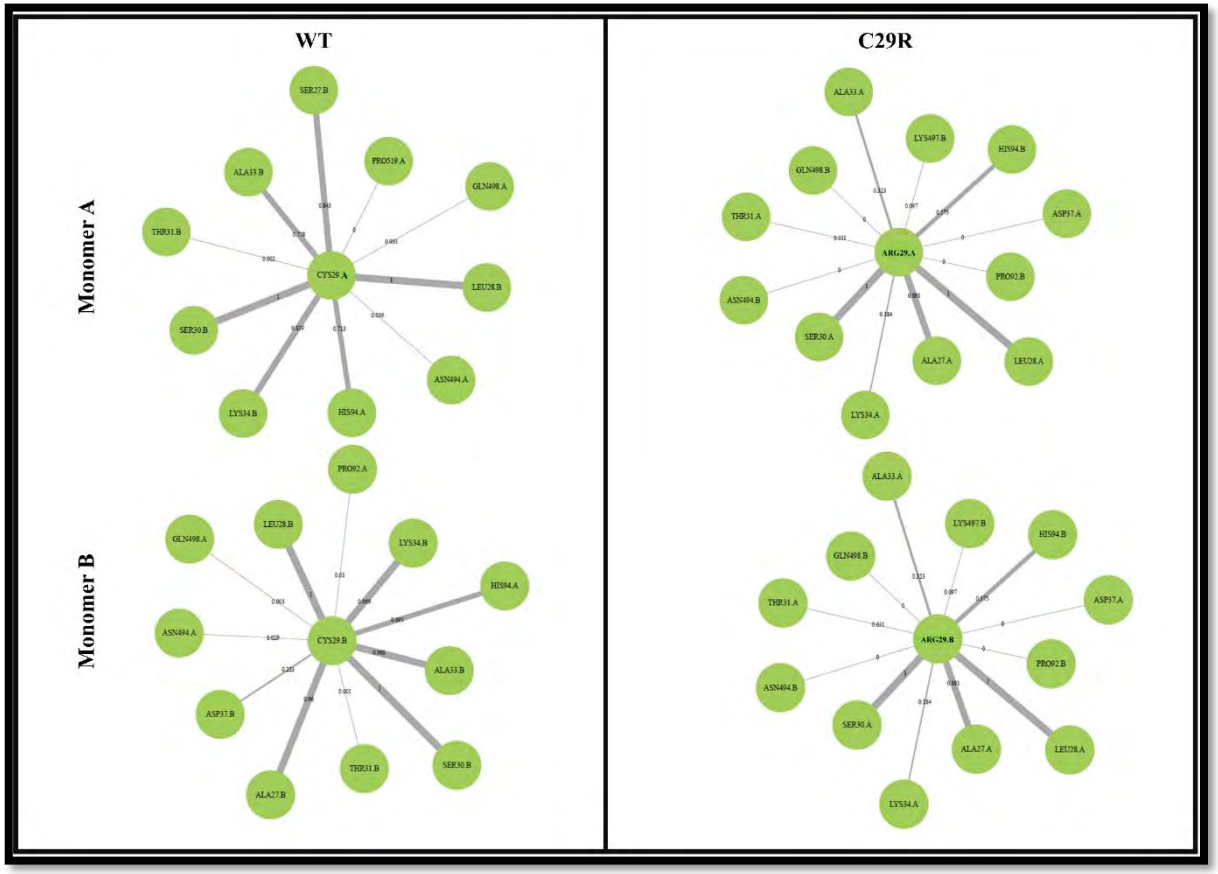


Figure S3.8.6. Contact-map representation of the frequency (%) of residue-residue interaction around the WT in both Monomer A and B versus the mutation C29R.

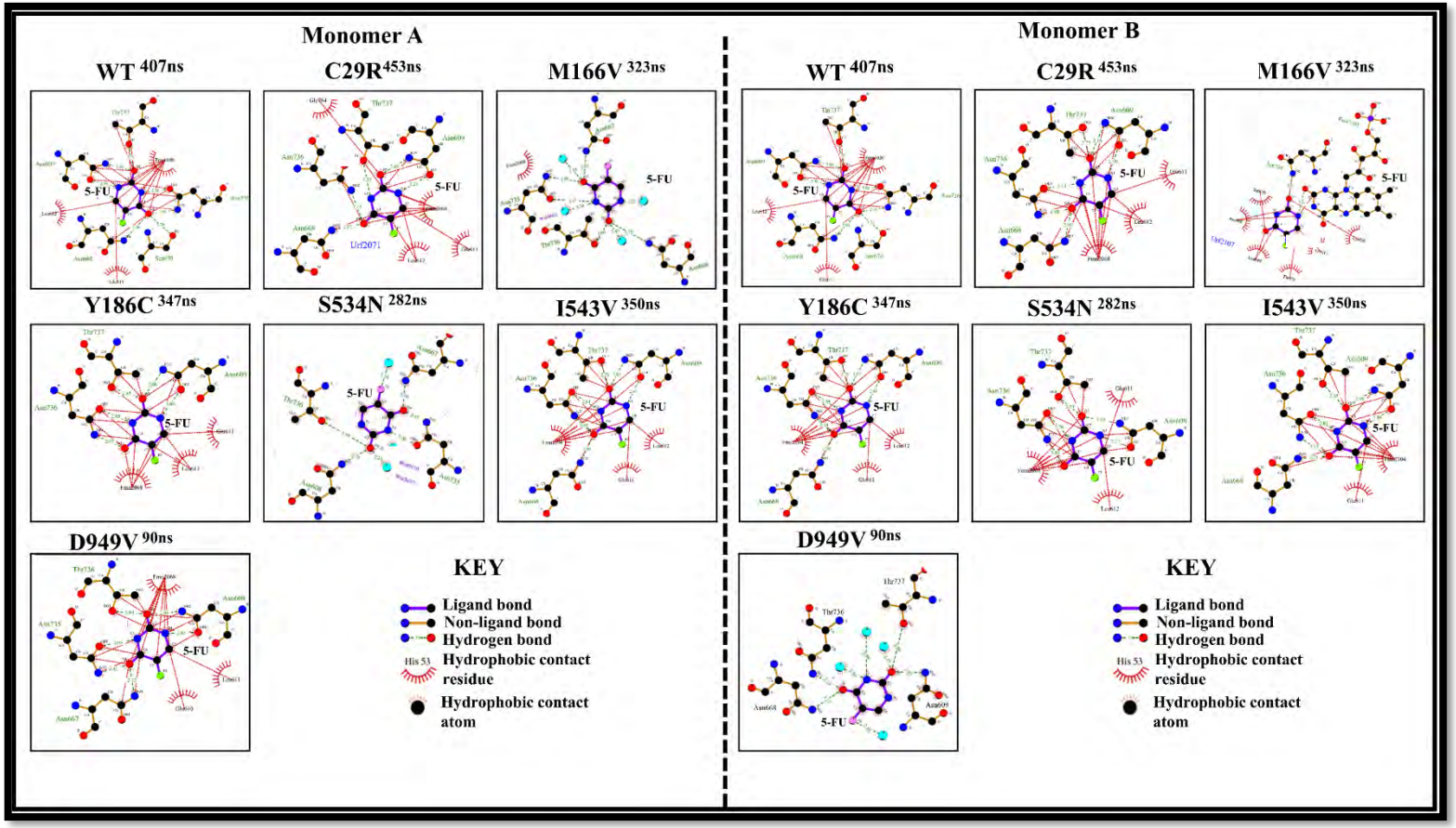


Figure S3.9. A 2D representation of Monomer A and Monomer B DPD-5FU binding residues. The structures were obtained from the trajectories' lowest energy minima, as specified on top of each picture frame.

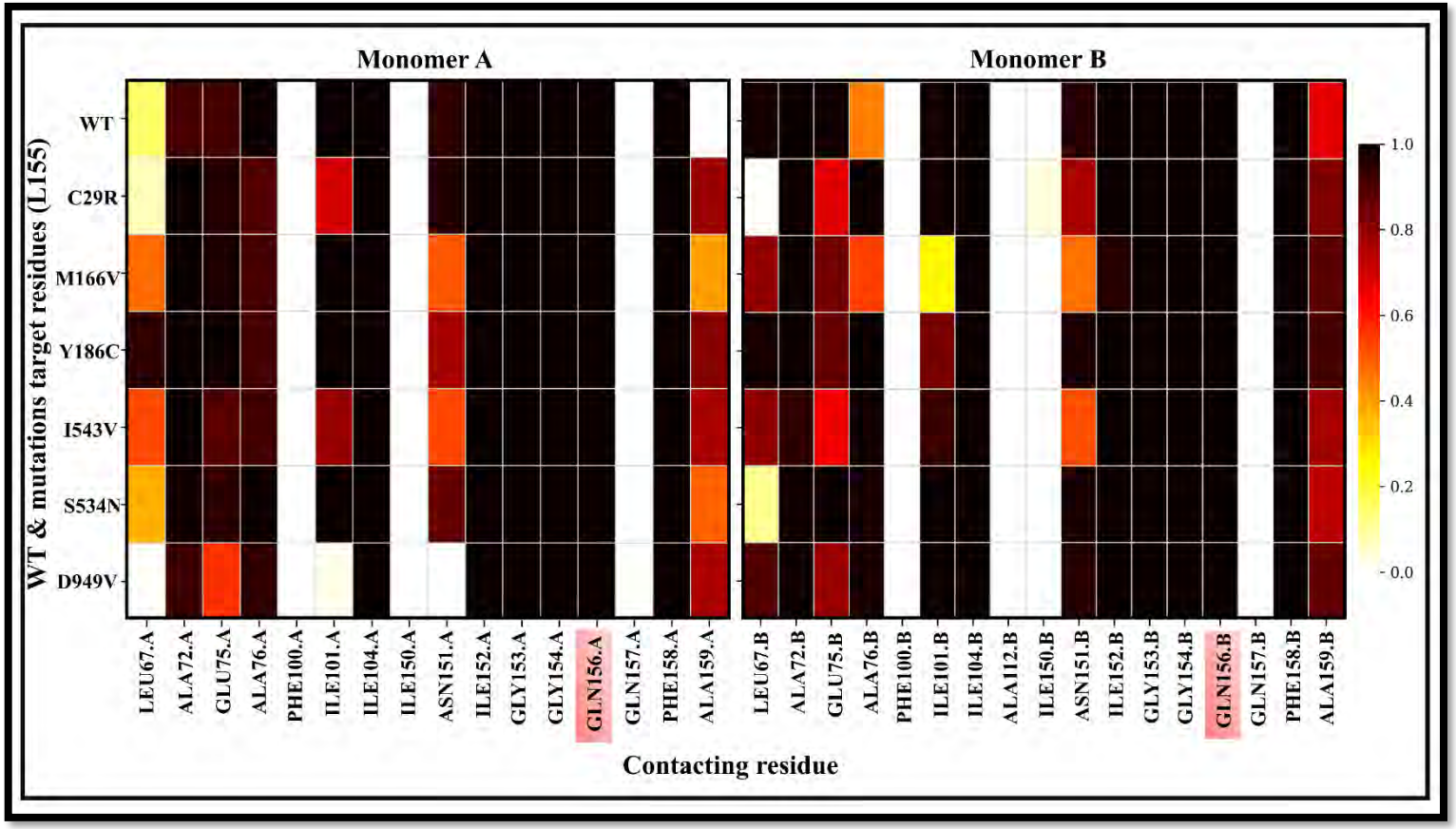


Figure S3.10.1: Contact heatmap representing contacting residue to high significant super-persistent hub L155 that connects to unique residue Q156, forming a bond with Fe^{2+}.

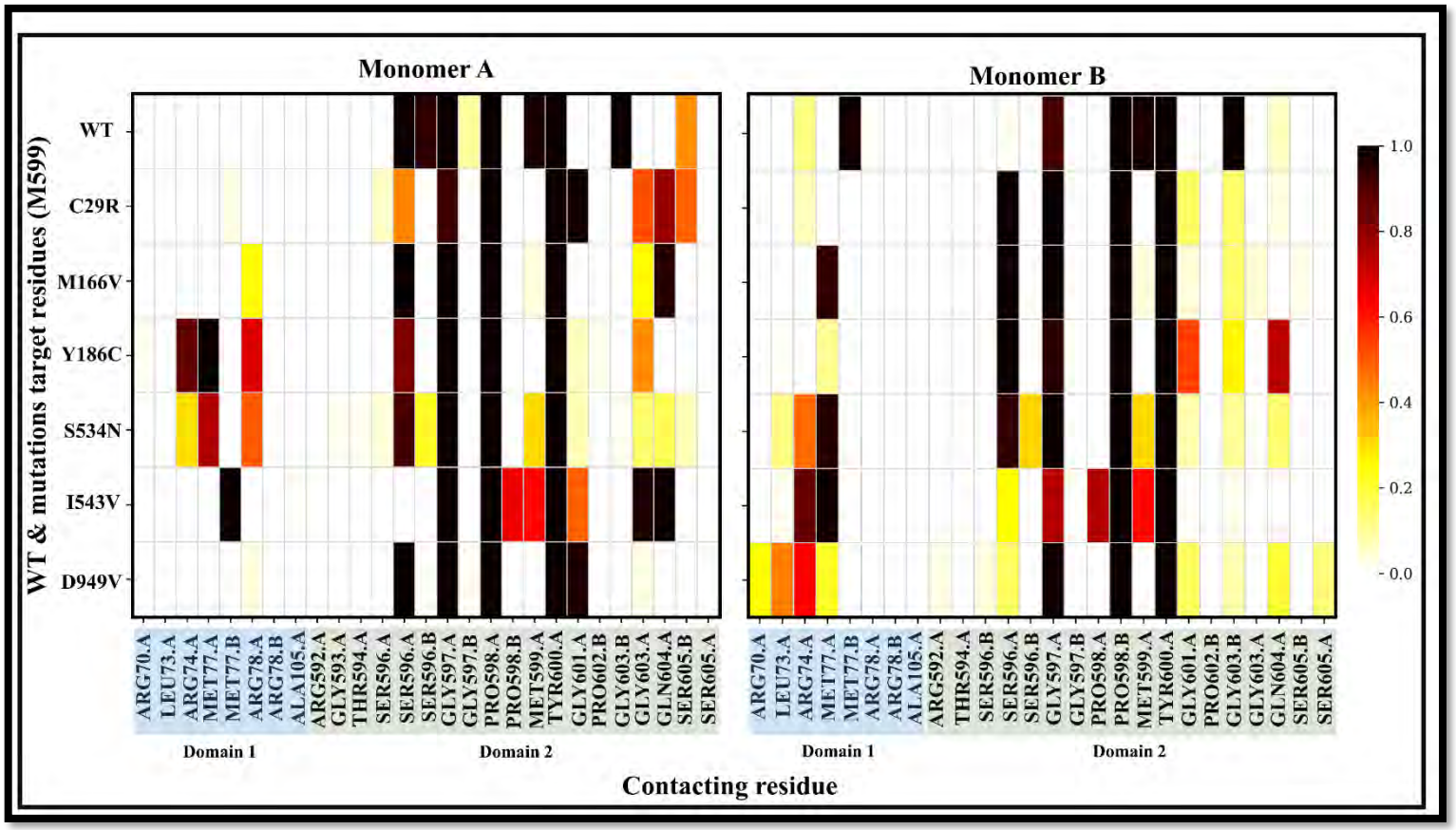


Figure S3.10.2: Contact heatmap representing contacting residue to high significant hub M599 that inter-connects to domain 1 and domain 4 (active site).

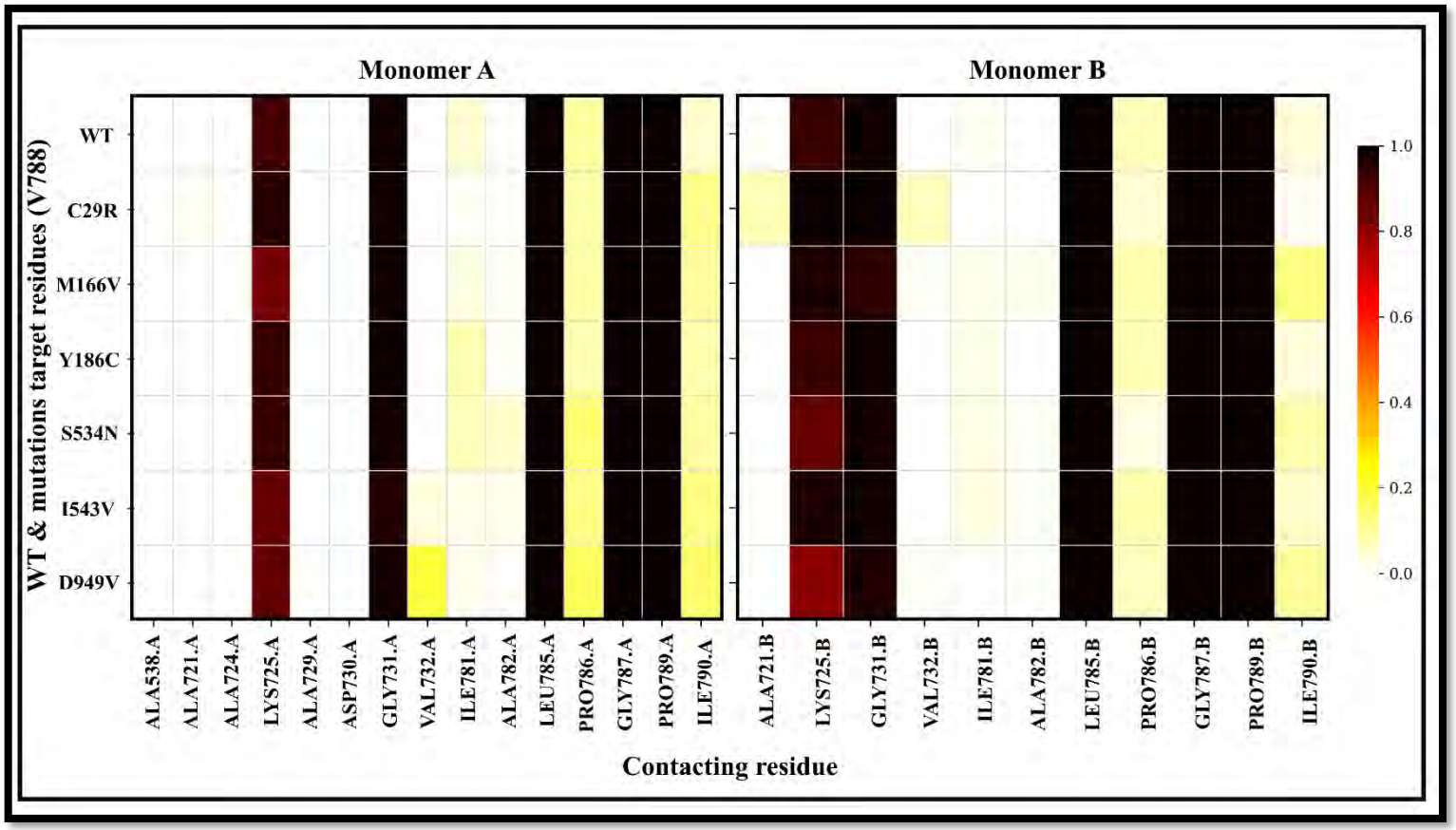


Figure S3.10.3: Contact heatmap representing contacting residue to high significant hub V778 that mediates interactions within the active site area.

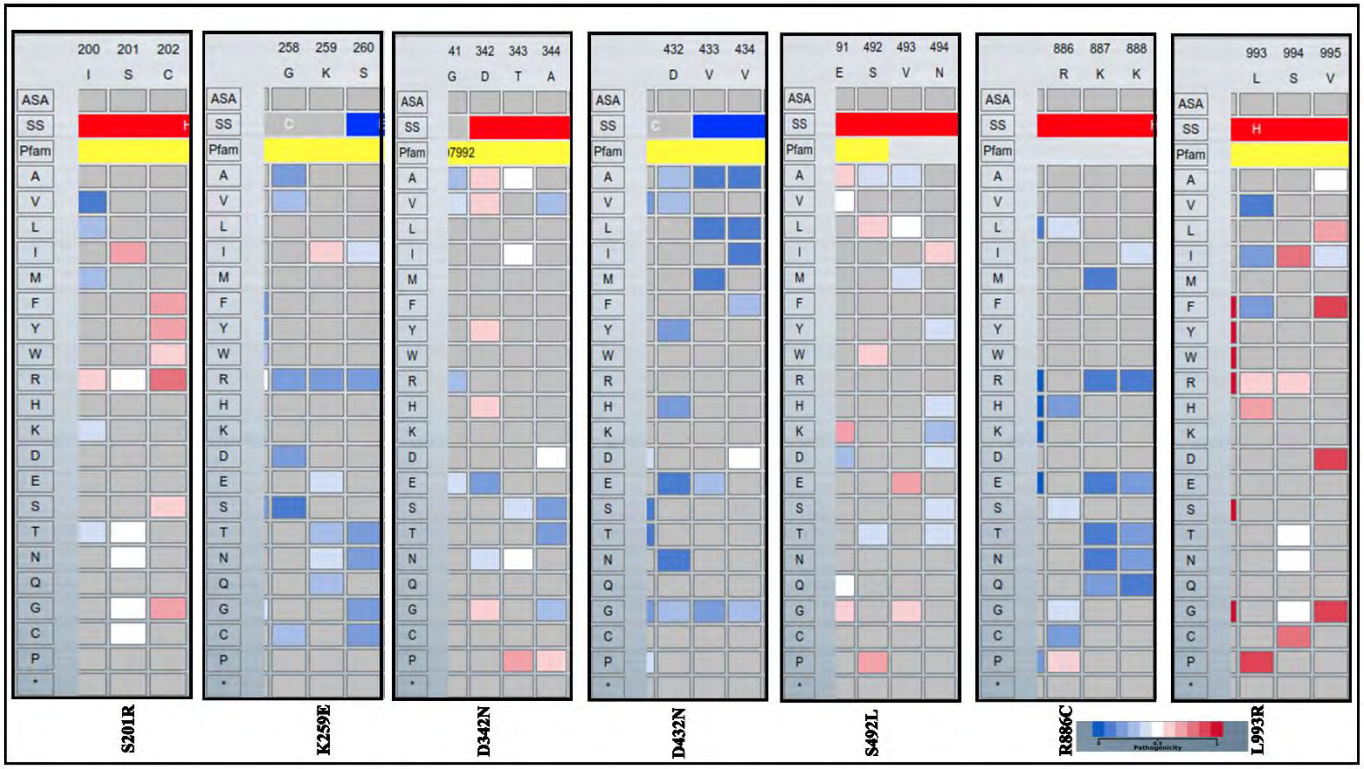


Figure S4.1.1: A heatmap displaying the sequence-based pathogenicity effects probability for remote mutants of uncertain significance (MUS) (S201R, K259E, D342N, D432N, S492L, R886C and L993R) generated by the Ensembl Variant Effect Predictor (EVE). The heatmap is color-coded from blue to white, representing a range from 0 to 1, indicating an ascending likelihood of pathogenicity. Mutations assigned a value of 0.5 on the heatmap are considered moderately pathogenic, while those assigned a value of 1 are highly likely to be pathogenic. Conversely, mutations with a value of 0 are less likely to be pathogenic.

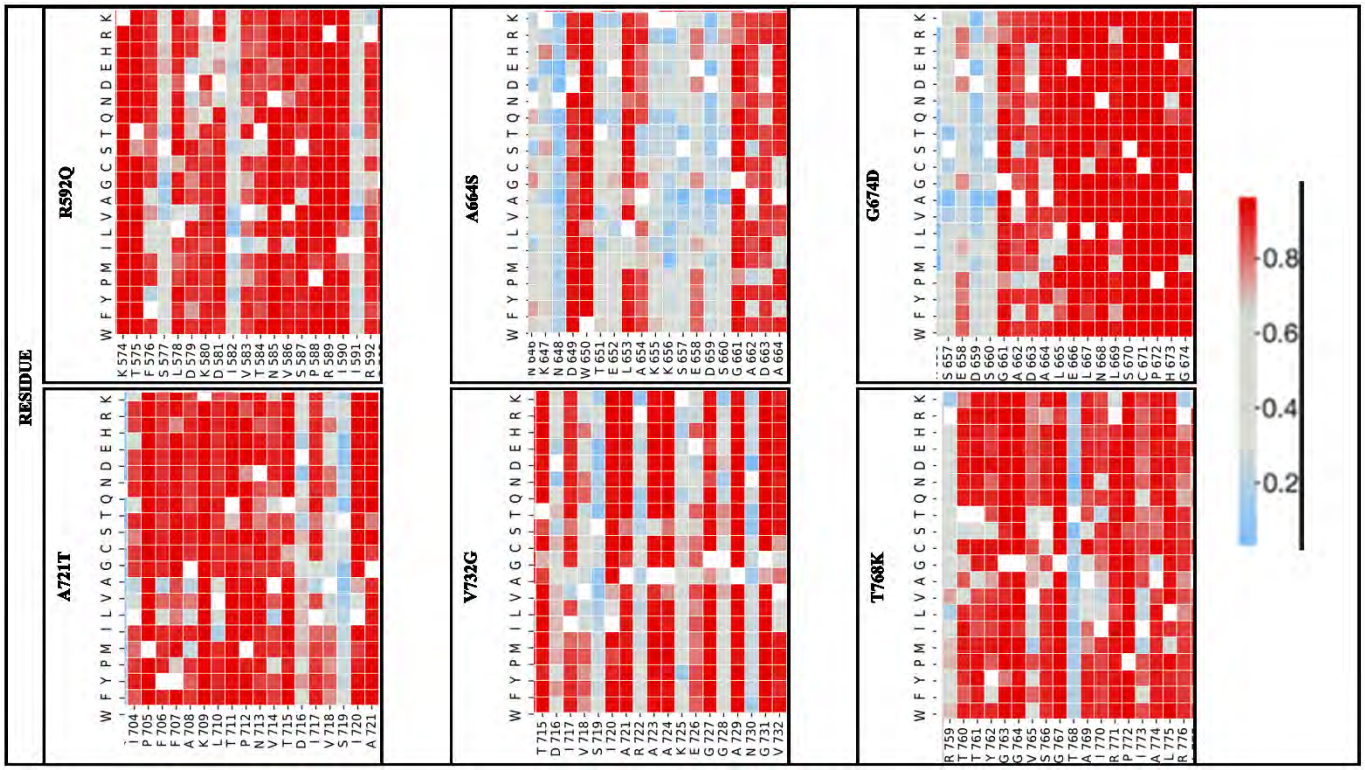


Figure S4.1.2: A heatmap displaying the sequence-based pathogenicity effects probability for remote mutants of uncertain significance (MUS) (S201R, K259E, D342N, D432N, S492L, R886C and L993R) generated by the Variant Annotation Integrator (VARITY). The heatmap is color-coded from blue to white, representing a range from 0 to 1, indicating an ascending likelihood of pathogenicity. Mutations assigned a value of 0.5 on the heatmap are considered moderately pathogenic, while those assigned a value of 1 are highly likely to be pathogenic. Conversely, mutations with a value of 0 are less likely to be pathogenic.

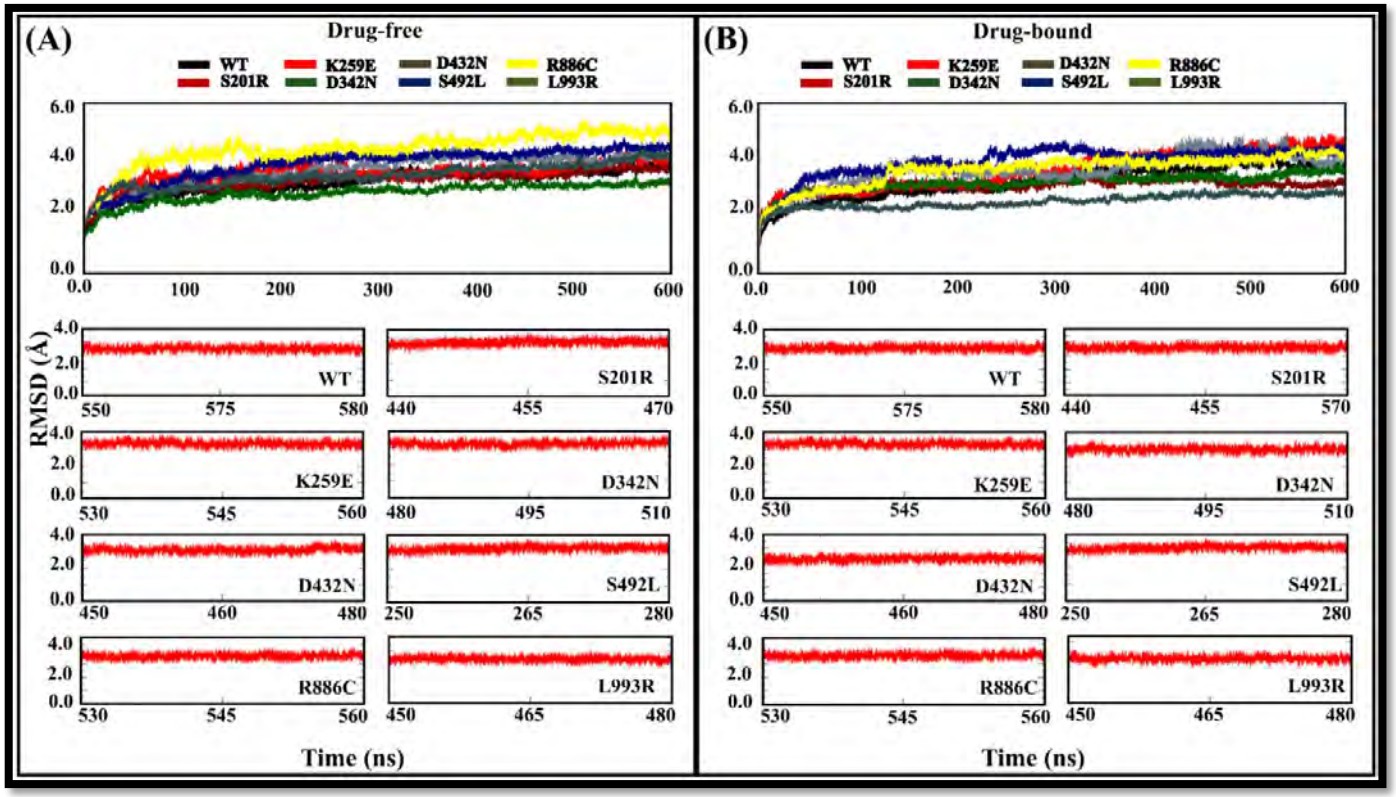


Figure S4.2.1. Line graph representing 600 ns RMSD simulation for the activated (drug free) and inactivated (drug bound) system. The red lines indicate the equilibrated regions of the protein for (A) the drug free and (B) drug bound ensembles, that were extracted for direct network analysis (DRN) calculation.

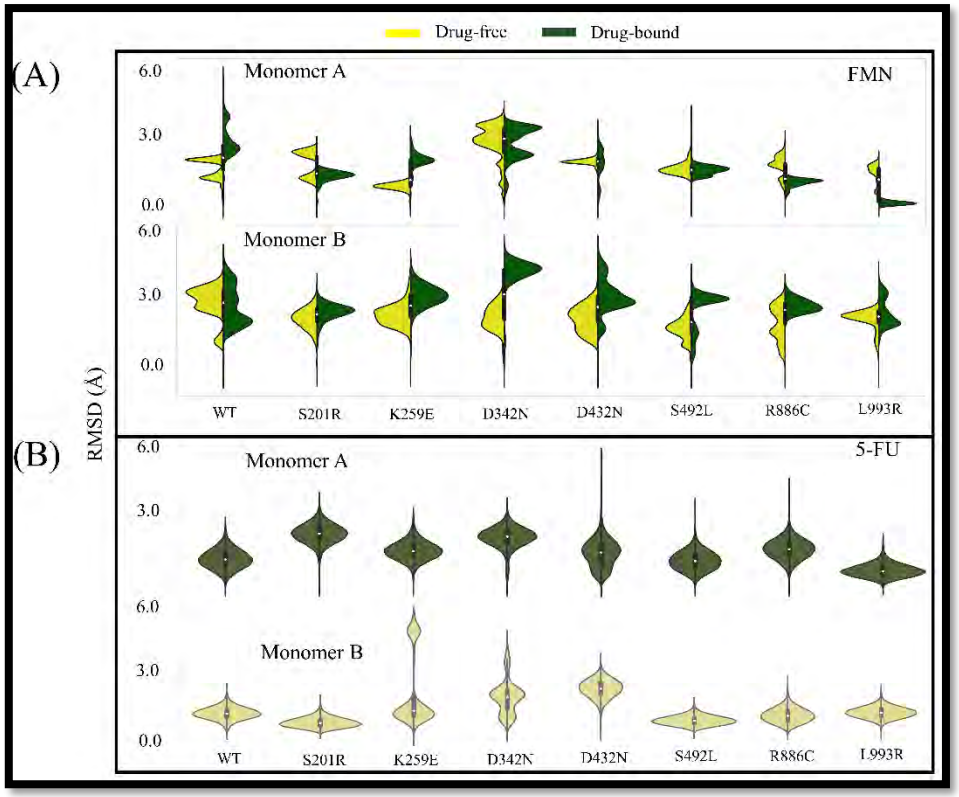


Figure S4.2.2. split Violin plots depicting 600 ns RMSD simulation for drug free and drug bound systems of cofactor (FMN) and ligand (5-FU)

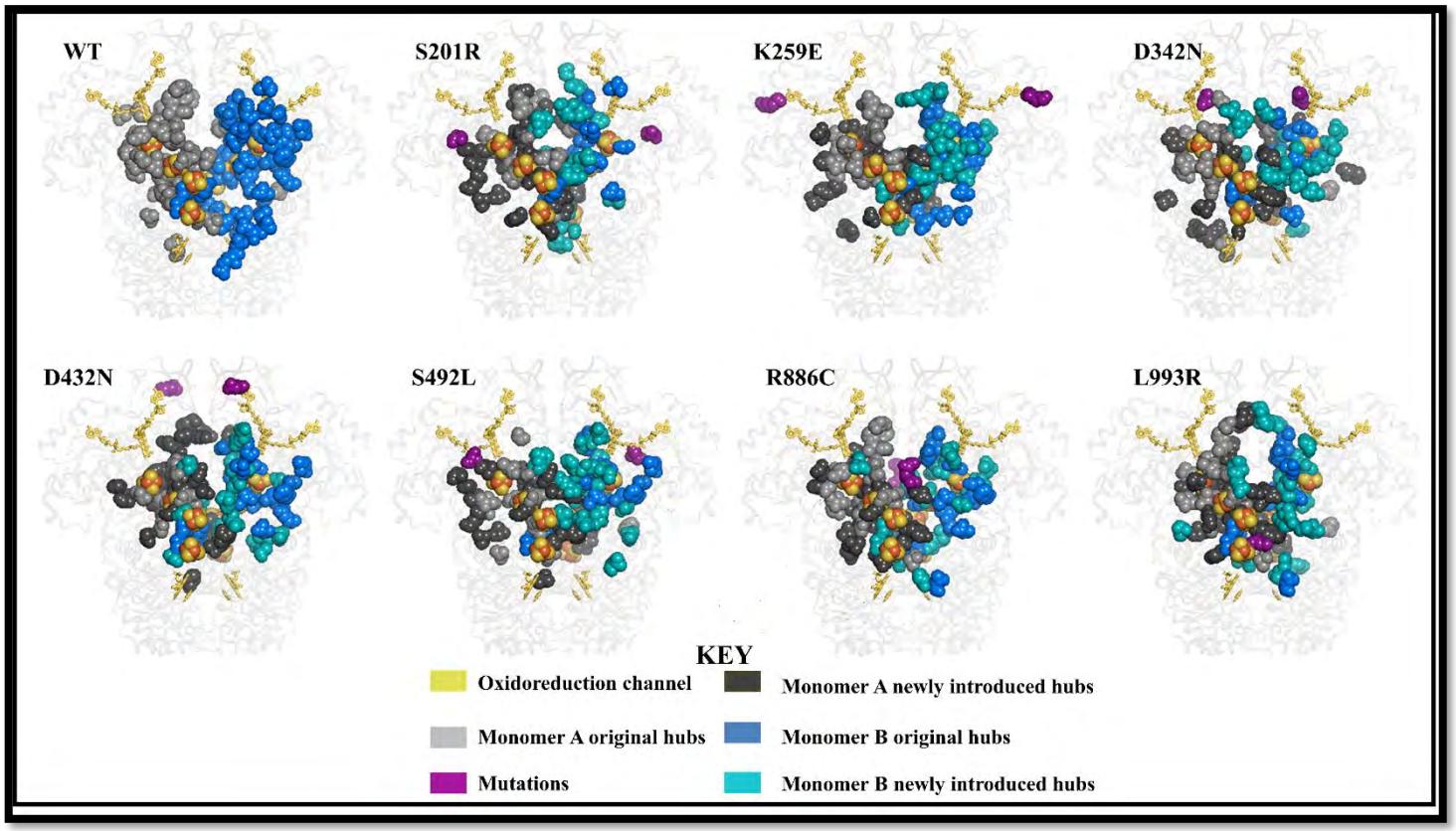


Figure S4.3.1: 3D cartoon representation of the distribution of *betweenness centrality* hubs of the drug free ensembles along the oxidoreduction catalytic channel. Monomer A WT *BC* hubs are denoted as light-grey, Monomer B WT *BC* hubs as blue. Mutation-induced newly introduced hubs shown in dark-grey in Monomer A and sky-blue Monomer B. Oxidoreduction catalytic channel (NADPH, FAD, $Fe^{2+}{}_4S^{2-}{}_4$ clusters, FMN and 5FU is shown in yellow. The mutations are represented in indigo.

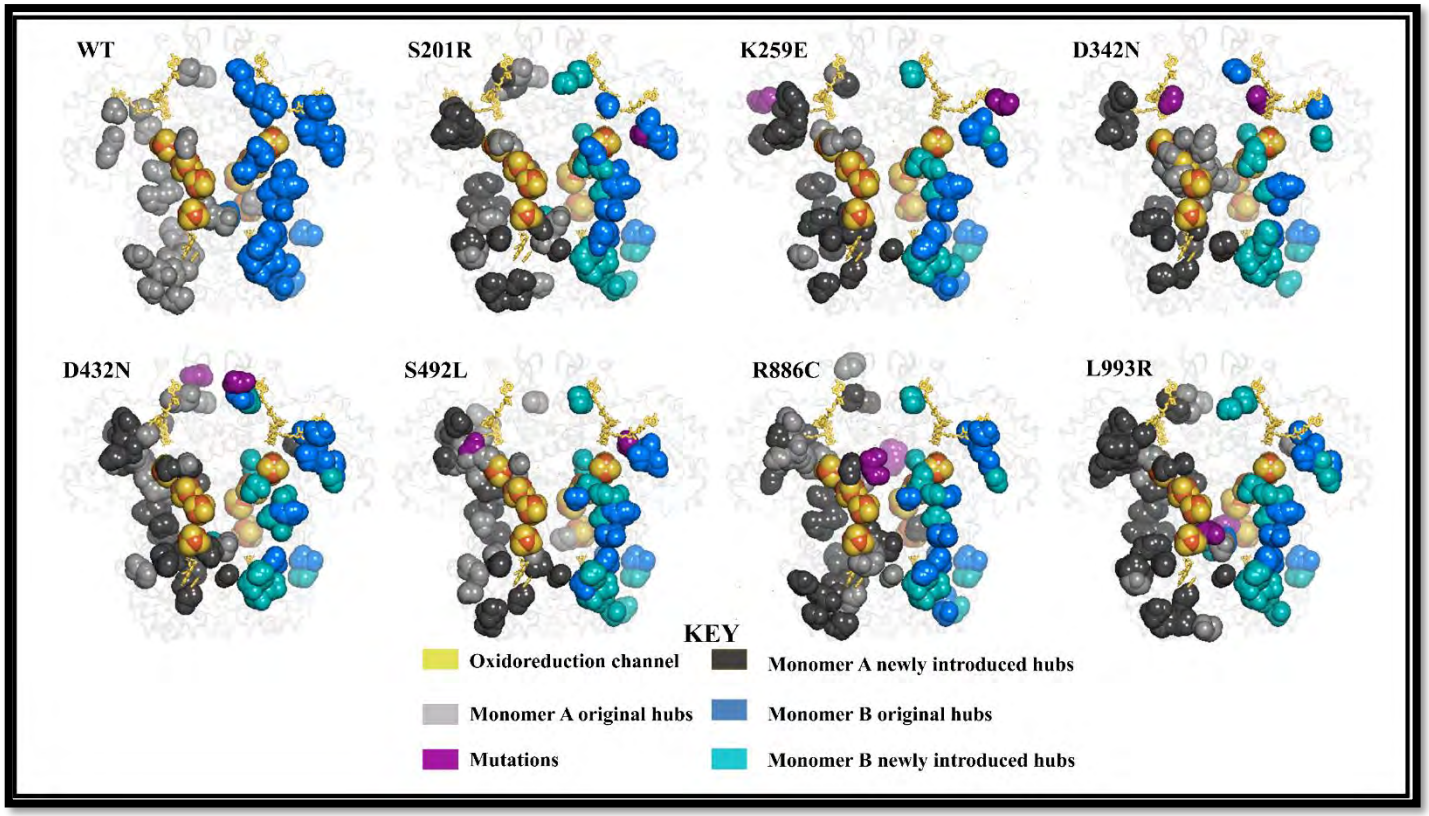


Figure S4.3.2: 3D cartoon representation of the distribution of *degree of centrality* hubs of the drug free ensembles along the oxidoreduction catalytic channel. Monomer A WT *DC* hubs are denoted as light-grey, Monomer B WT *DC* hubs as blue. Mutation-induced newly introduced hubs shown in dark-grey in Monomer A and sky-blue Monomer B. Oxidoreduction catalytic channel (NADPH, FAD, $Fe^{2+}{}_4S^{2-}{}_4$ clusters, FMN and 5FU is shown in yellow. The mutations are represented in indigo.

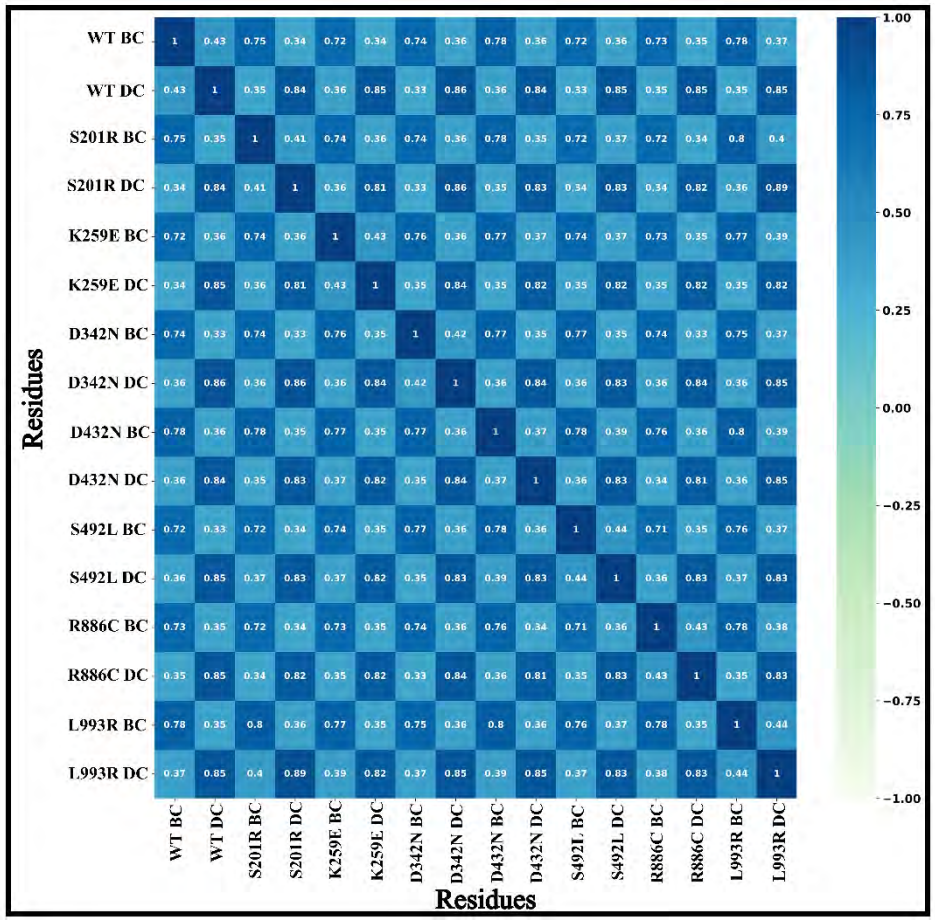


Figure S4.3.3. A direct cross-correlation (DCC) heatmap showing the relationship between WT and mutations in DPD drug free ensembles.

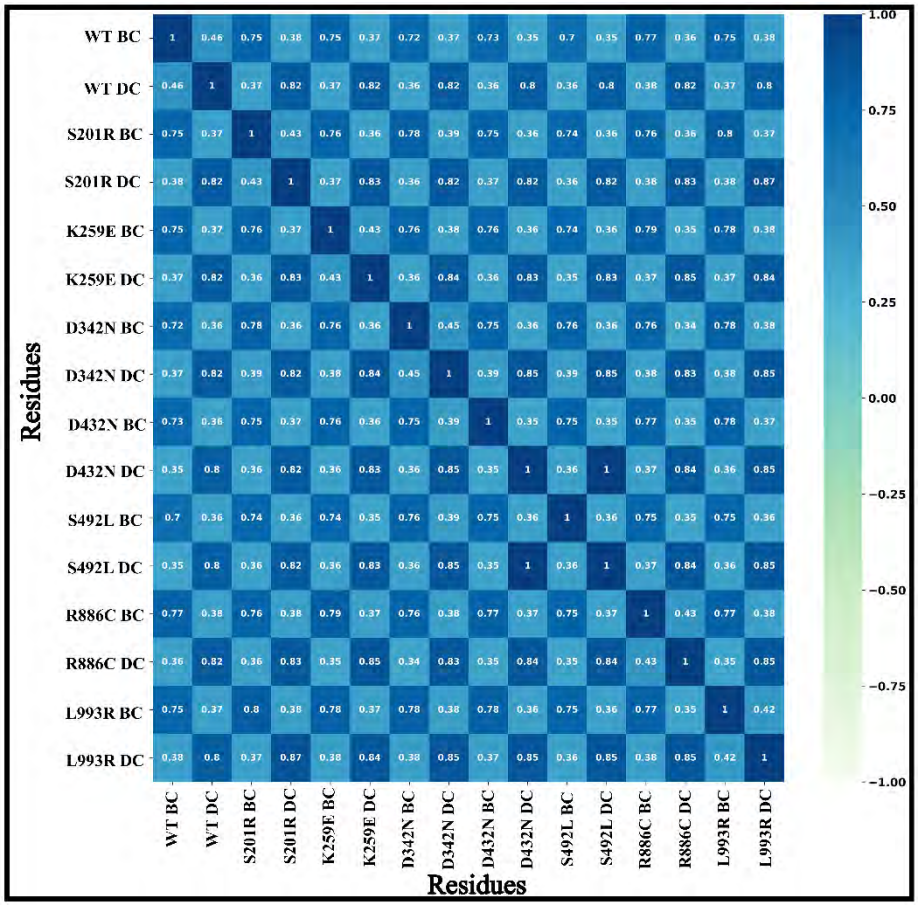


Figure S4.3.4. A direct cross-correlation (DCC) heatmap showing the relationship between WT and mutations in DPD drug bound ensembles.

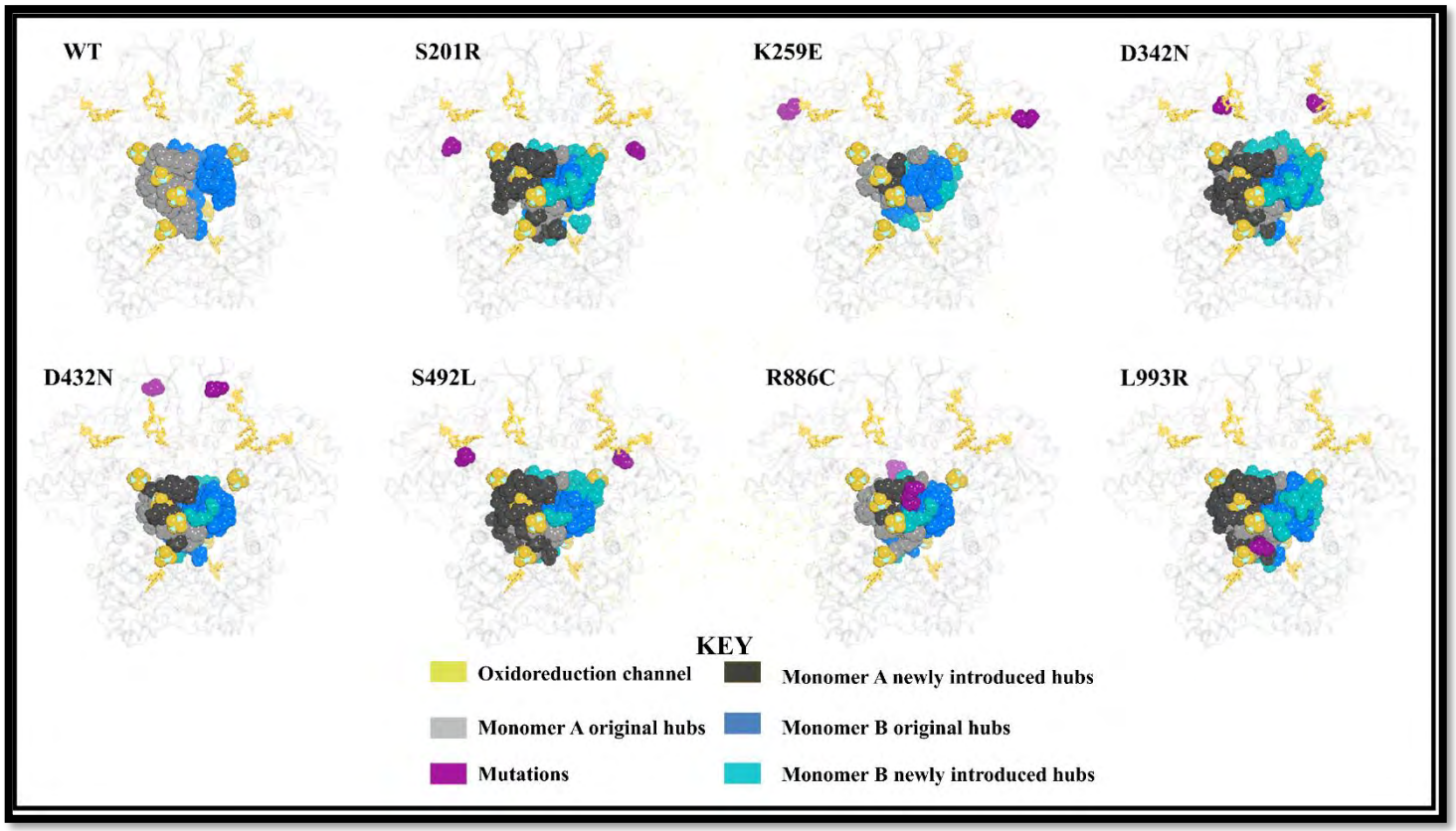


Figure S4.3.5: 3D cartoon representation of the distribution of *closeness centrality* hubs of the drug free ensembles along the oxidoreduction catalytic channel. Monomer A WT *CC* hubs are denoted as light-grey, Monomer B WT *CC* hubs as blue. Mutation-induced newly introduced hubs shown in dark-grey in Monomer A and sky-blue Monomer B. Oxidoreduction catalytic channel (NADPH, FAD, $Fe^{2+}{}_4S^{2-}{}_4$ clusters, FMN and 5FU is shown in yellow. The mutations are represented in indigo.

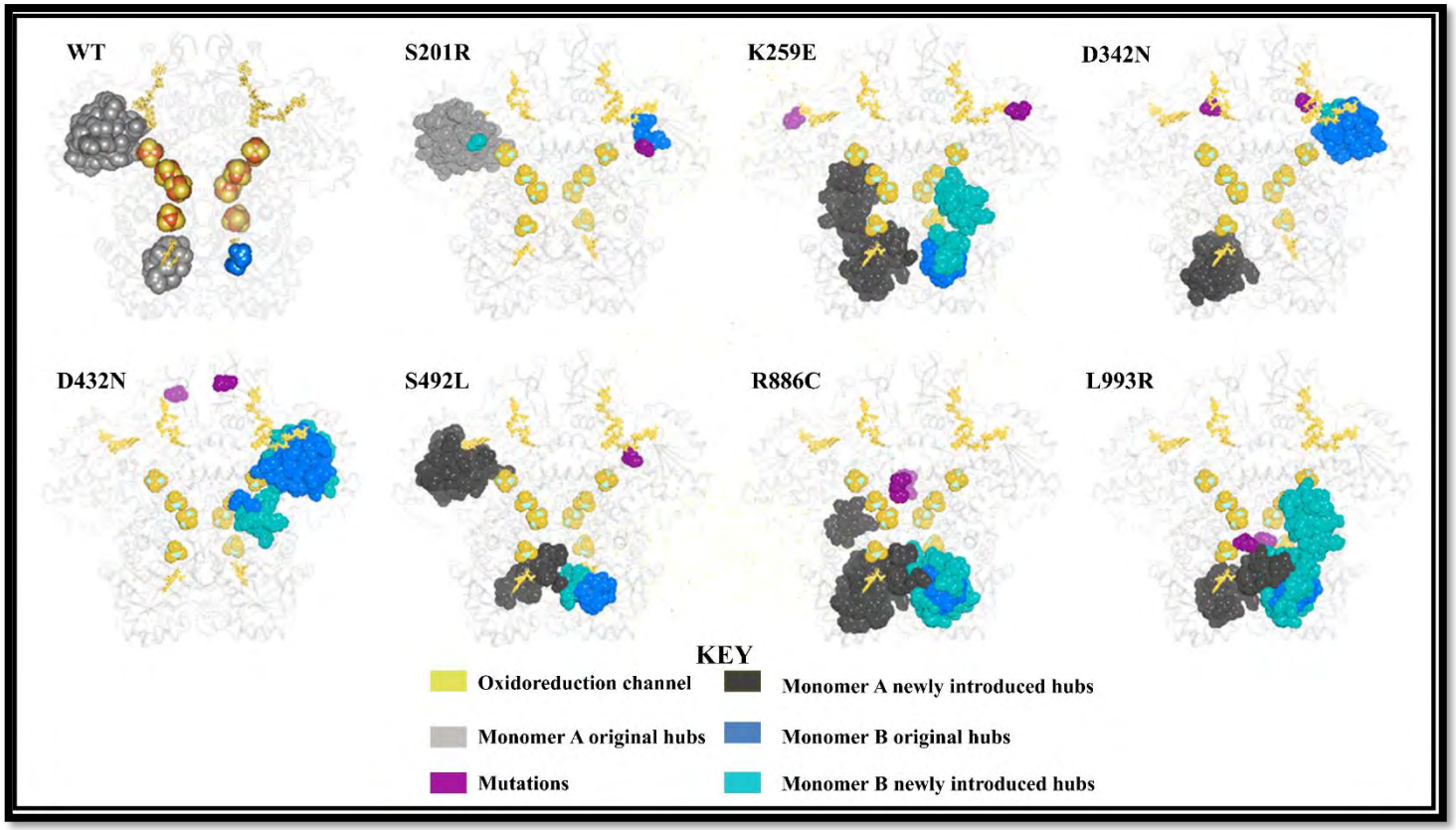


Figure S4.3.6. 3D cartoon representation of the distribution of *eigen centrality* hubs of the drug free ensembles along the oxidoreduction catalytic channel. Monomer A WT *EC* hubs are denoted as light-grey, Monomer B WT *EC* hubs as blue. Mutation-induced newly introduced hubs shown in dark-grey in Monomer A and sky-blue Monomer B. Oxidoreduction catalytic channel (NADPH, FAD, $Fe^{2+}{}_4S^{2-}{}_4$ clusters, FMN and 5FU is shown in yellow. The mutations are represented in indigo.

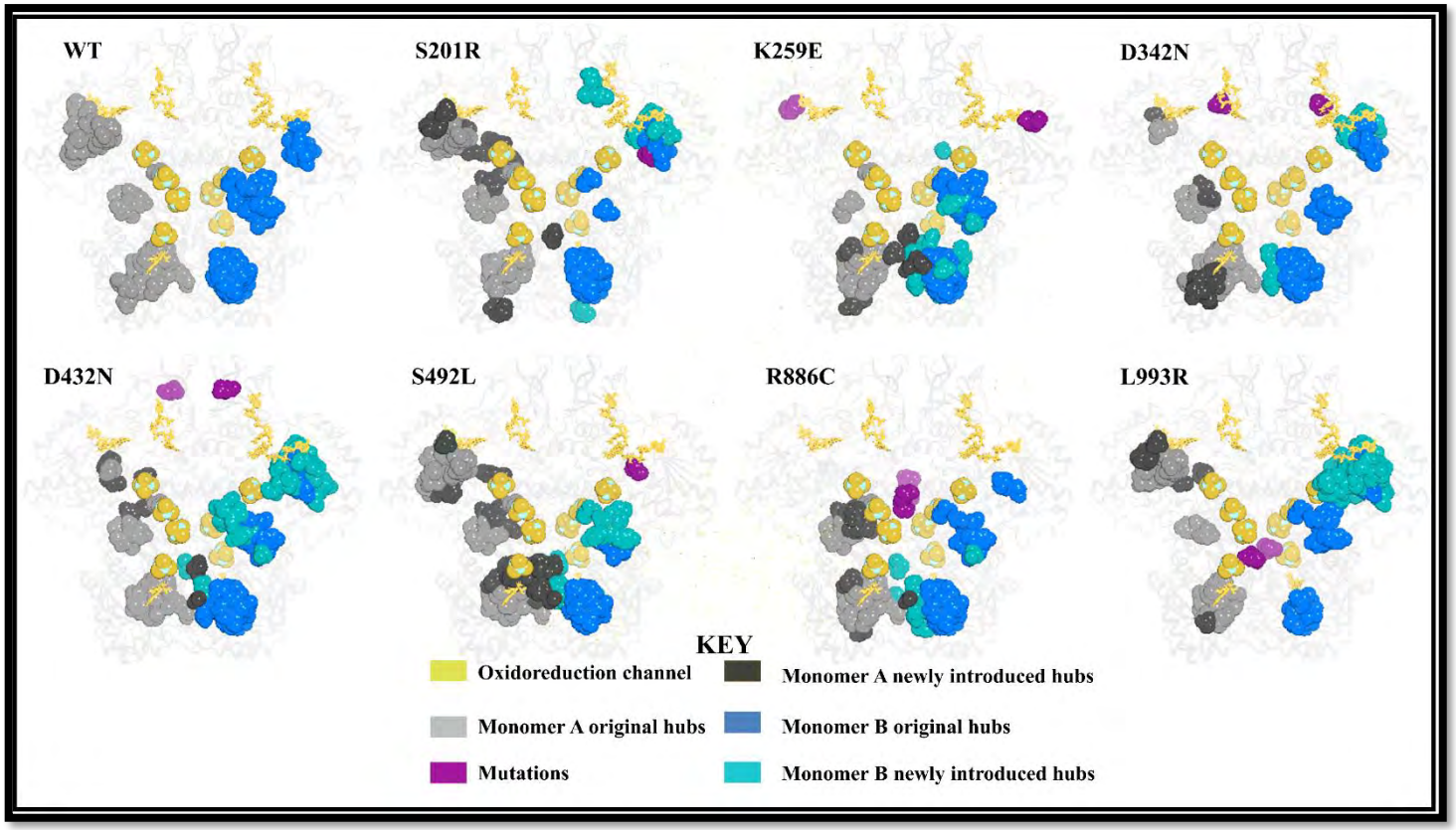


Figure S4.3.7: 3D cartoon representation of the distribution of *katz centrality* hubs of the drug free ensembles along the oxidoreduction catalytic channel. Monomer A WT *KC* hubs are denoted as light-grey, Monomer B WT *KC* hubs as blue. Mutation-induced newly introduced hubs shown in dark-grey in Monomer A and sky-blue Monomer B. Oxidoreduction catalytic channel (NADPH, FAD, $Fe^{2+}{}_4S^{2-}{}_4$ clusters, FMN and 5FU is shown in yellow. The mutations are represented in indigo.

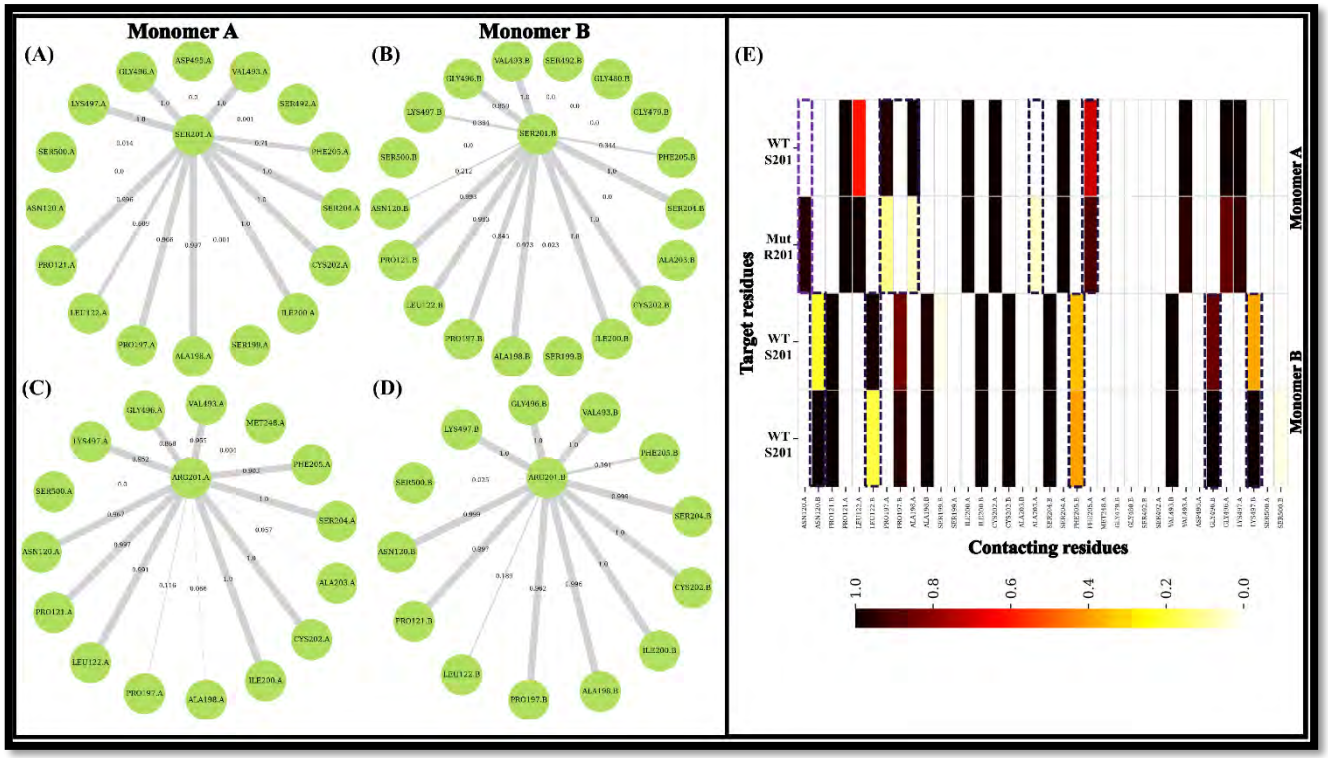


Figure S4.4.1. Contact-map representing frequency (%) of residue-residue interaction around Monomer A and B S201of the (WT) in comparison to the mutation (S201R). (A) Weighted residues contact map around S201 (WT) of Monomer A, (B) Weighted residue contact map around S201 (WT) of Monomer B, (C) Weighted residue contact map around R201 (mutation) of Monomer A, (D) Weighted residue contact map around R201 (mutation) of Monomer B and (E) Weighted residues contact heatmap showing the inter-residue interaction around S201 (WT) and R201 (mutation) of Monomer A and Monomer B of DPD. The shaded boxes in each frame represent significant differences in the frequency of interaction between the S201 residues and the surrounding contacting residues in the WT versus the mutations. White (0) to dark-red (1.0) represent no-contact (0) to strong-contact (1.0)

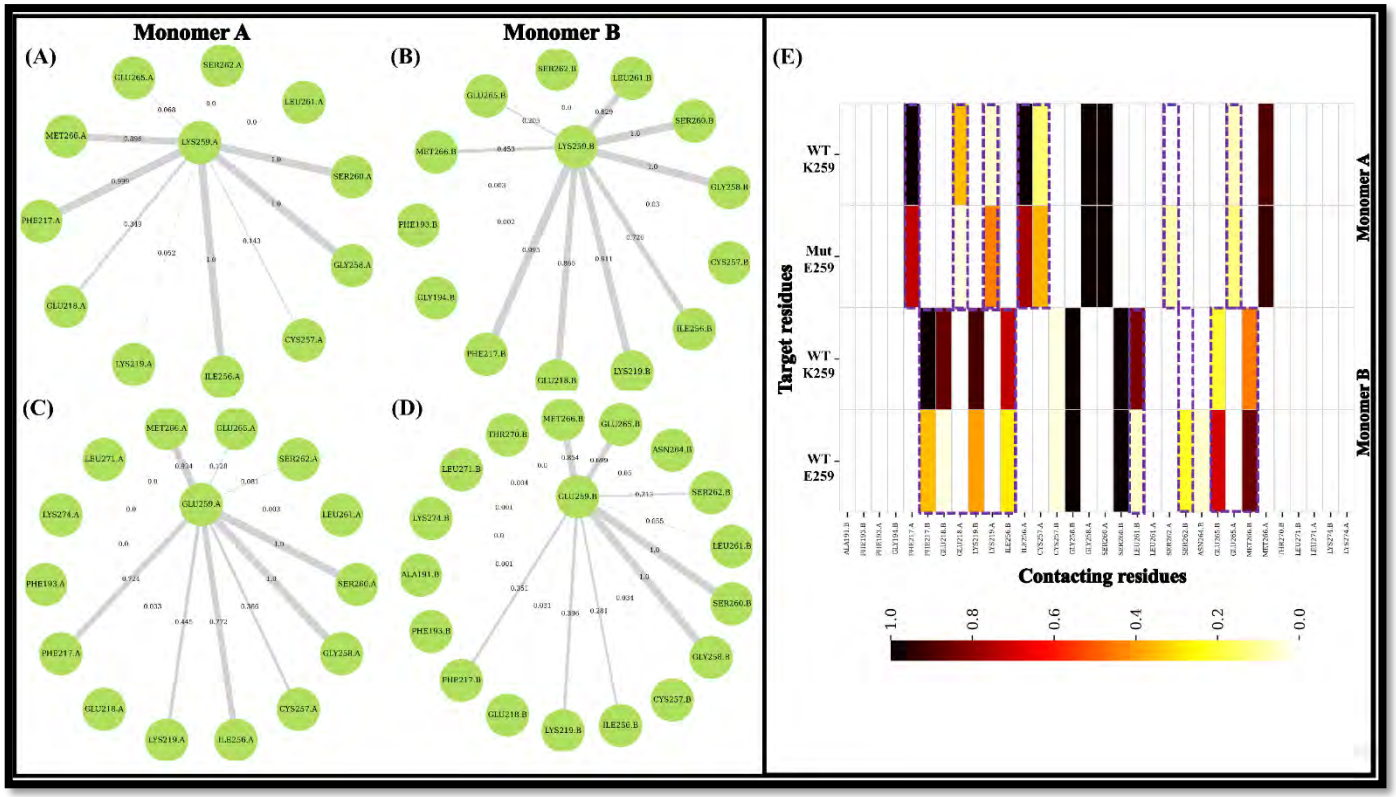


Figure S4.4.2. Contact-map representing frequency (%) of residue-residue interaction around Monomer A and B K259 of the (WT) in comparison to the mutation (K259E). (A) Weighted residues contact map around K259 (WT) of Monomer A, (B) Weighted residue contact map around K259 (WT) of Monomer B, (C) Weighted residue contact map around E259 (mutation) of Monomer A, (D) Weighted residue contact map around E259 (mutation) of Monomer B and (E) Weighted residues contact heatmap showing the inter-residue interaction around K259 (WT) and E259 (mutation) of Monomer A and Monomer B of DPD. The shaded boxes in each frame represent significant differences in the frequency of interaction between the K259 residues and the surrounding contacting residues in the WT versus the mutations. White (0) to dark-red (1.0) represent no-contact (0) to strong-contact (1.0)

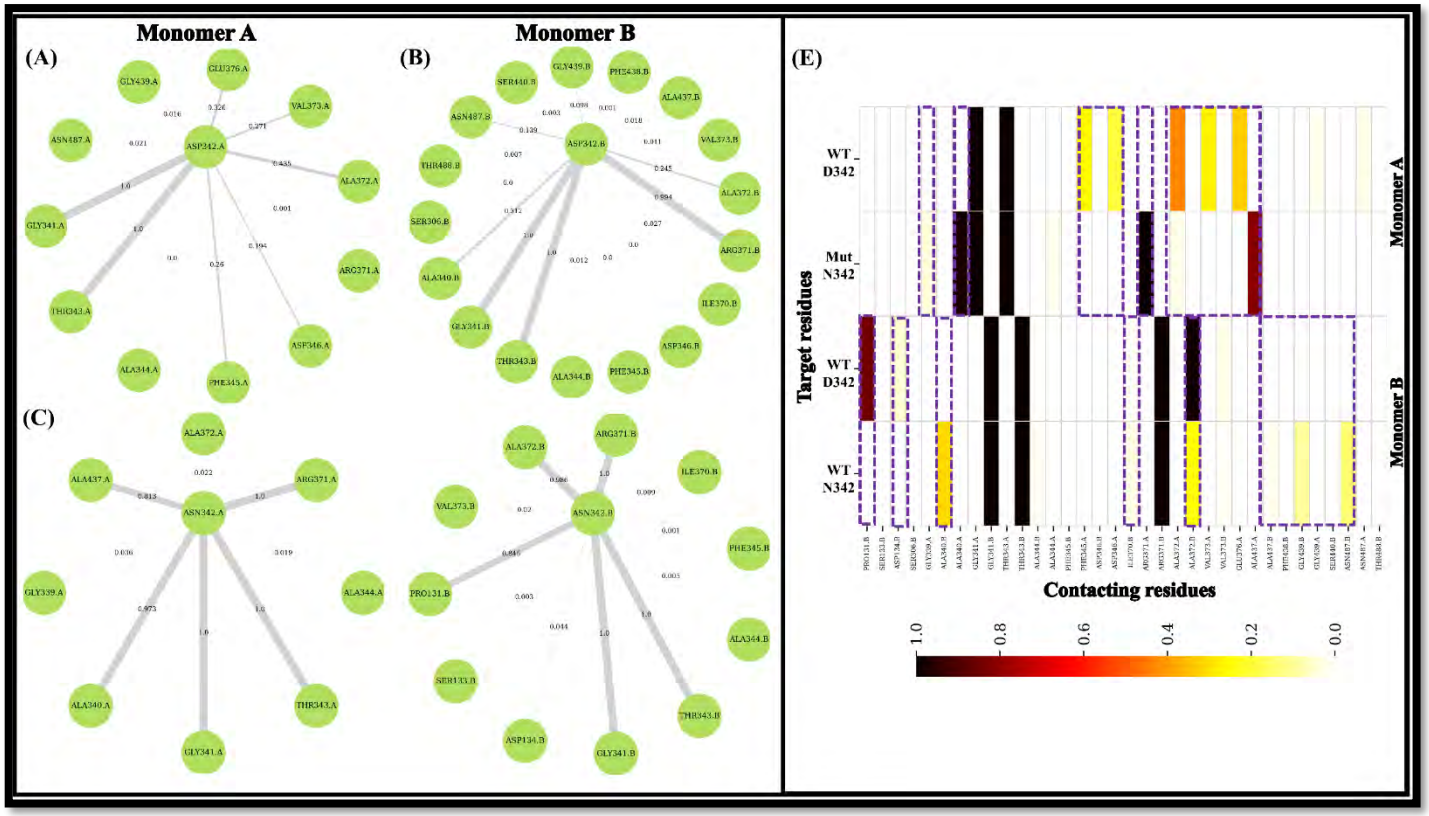


Figure S4.4.3. Contact-map representing frequency (%) of residue-residue interaction around Monomer A and B D342 of the (WT) in comparison to the mutation (D342N). (A) Weighted residues contact map around D342 (WT) of Monomer A, (B) Weighted residue contact map around D342 (WT) of Monomer B, (C) Weighted residue contact map around D342 (mutation) of Monomer A, (D) Weighted residue contact map around D342 (mutation) of Monomer B and (E) Weighted residues contact heatmap showing the inter-residue interaction around D342 (WT) and D342 (mutation) of Monomer A and Monomer B of DPD. The shaded boxes in each frame represent significant differences in the frequency of interaction between the D342 residues and the surrounding contacting residues in the WT versus the mutations. White (0) to dark-red (1.0) represent no-contact (0) to strong-contact (1.0)

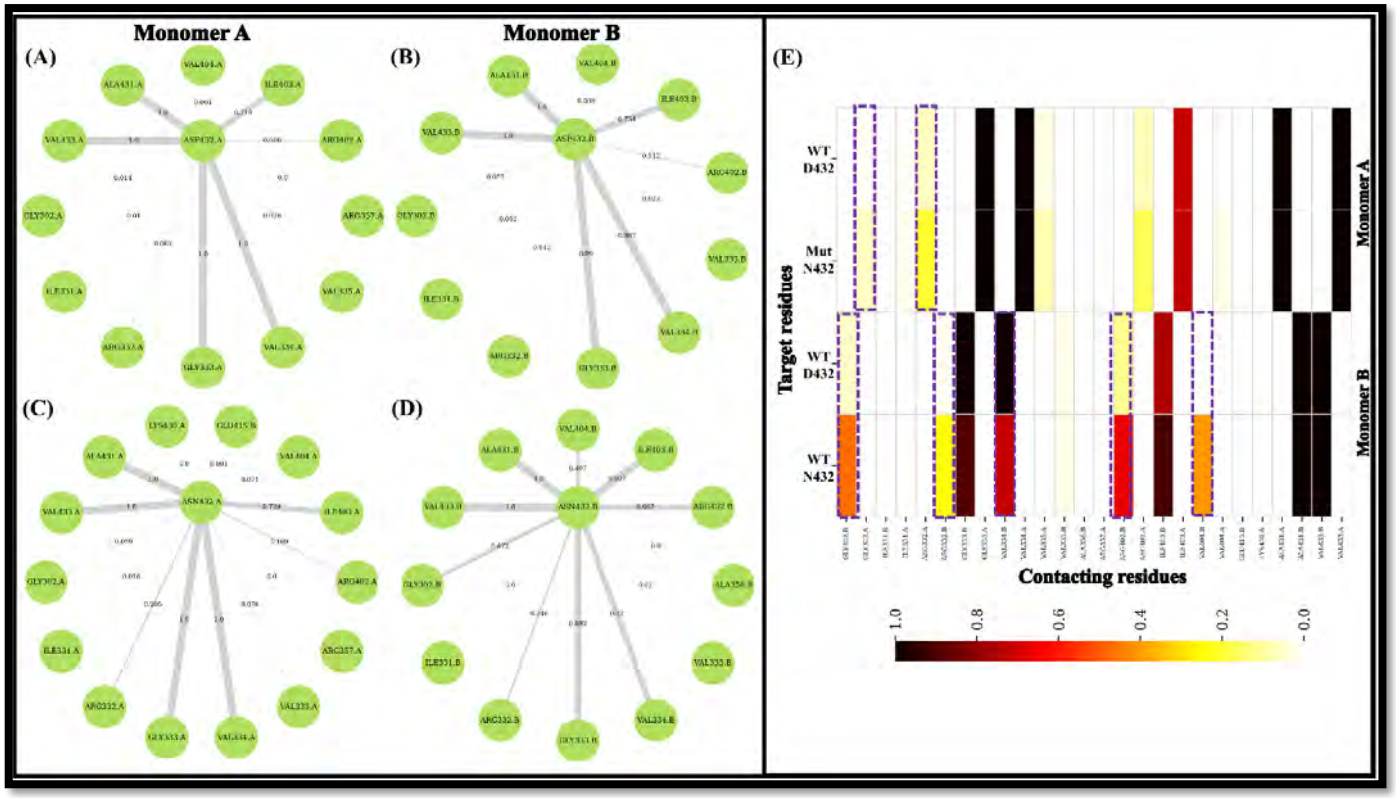


Figure S4.4.4. Contact-map representing frequency (%) of residue-residue interaction around Monomer A and B D432 of the (WT) in comparison to the mutation (D432N). (A) Weighted residues contact map around D432 (WT) of Monomer A, (B) Weighted residue contact map around D432 (WT) of Monomer B, (C) Weighted residue contact map around D432 (mutation) of Monomer A, (D) Weighted residue contact map around D432 (mutation) of Monomer B and (E) Weighted residues contact heatmap showing the inter-residue interaction around D432 (WT) and D432 (mutation) of Monomer A and Monomer B of DPD. The shaded boxes in each frame represent significant differences in the frequency of interaction between the D432 residues and the surrounding contacting residues in the WT versus the mutations. White (0) to dark-red (1.0) represent no-contact (0) to strong-contact (1.0)

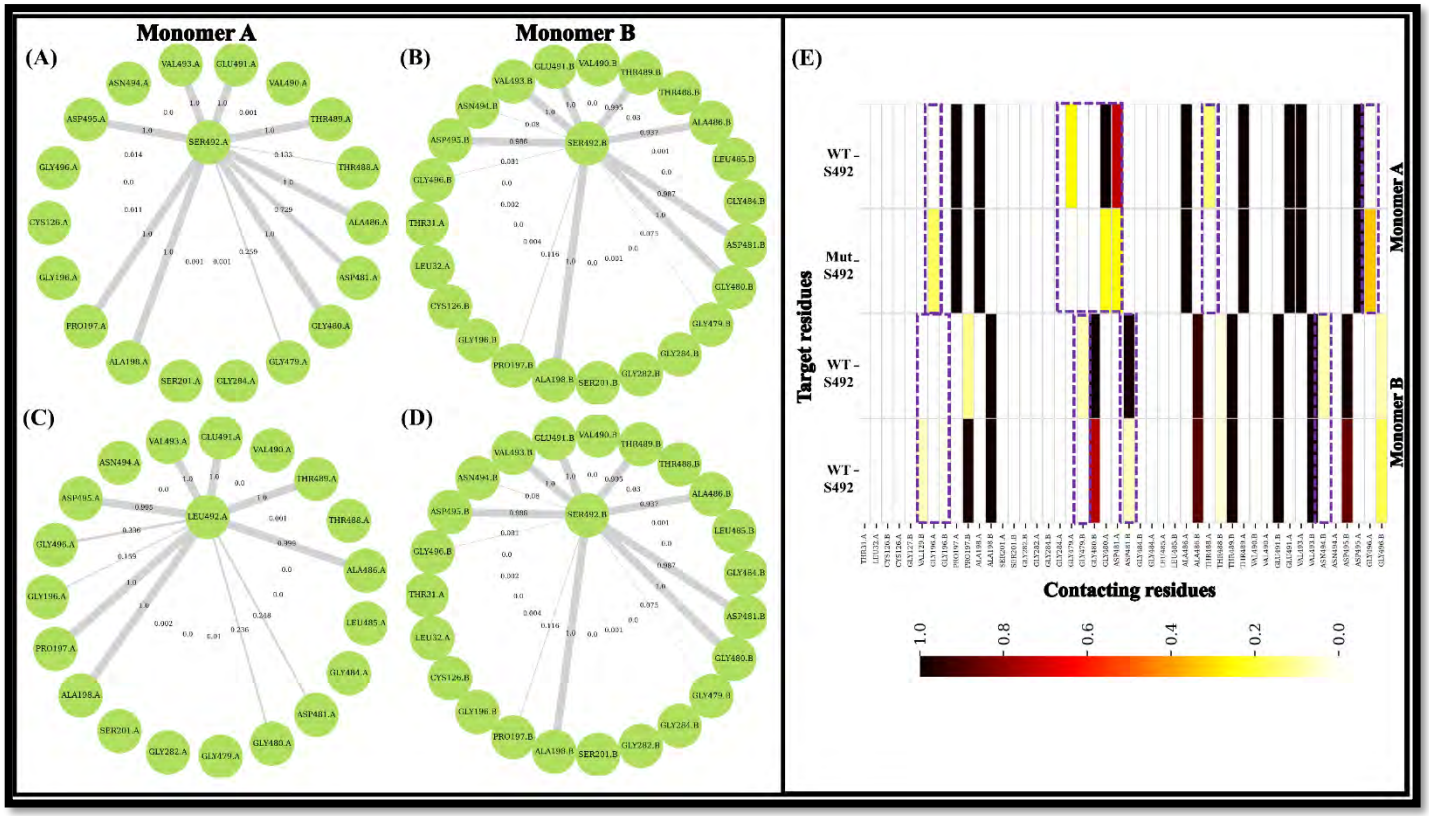


Figure S4.4.5. Contact-map representing frequency (%) of residue-residue interaction around Monomer A and B S492 of the (WT) in comparison to the mutation (S492L). (A) Weighted residues contact map around S492 (WT) of Monomer A, (B) Weighted residue contact map around S492 (WT) of Monomer B, (C) Weighted residue contact map around L492 (mutation) of Monomer A, (D) Weighted residue contact map around L492 (mutation) of Monomer B and (E) Weighted residues contact heatmap showing the inter-residue interaction around S492 (WT) and L492 (mutation) of Monomer A and Monomer B of DPD. The shaded boxes in each frame represent significant differences in the frequency of interaction between the S492 residues and the surrounding contacting residues in the WT versus the mutations. White (0) to dark-red (1.0) represent no-contact (0) to strong-contact (1.0)

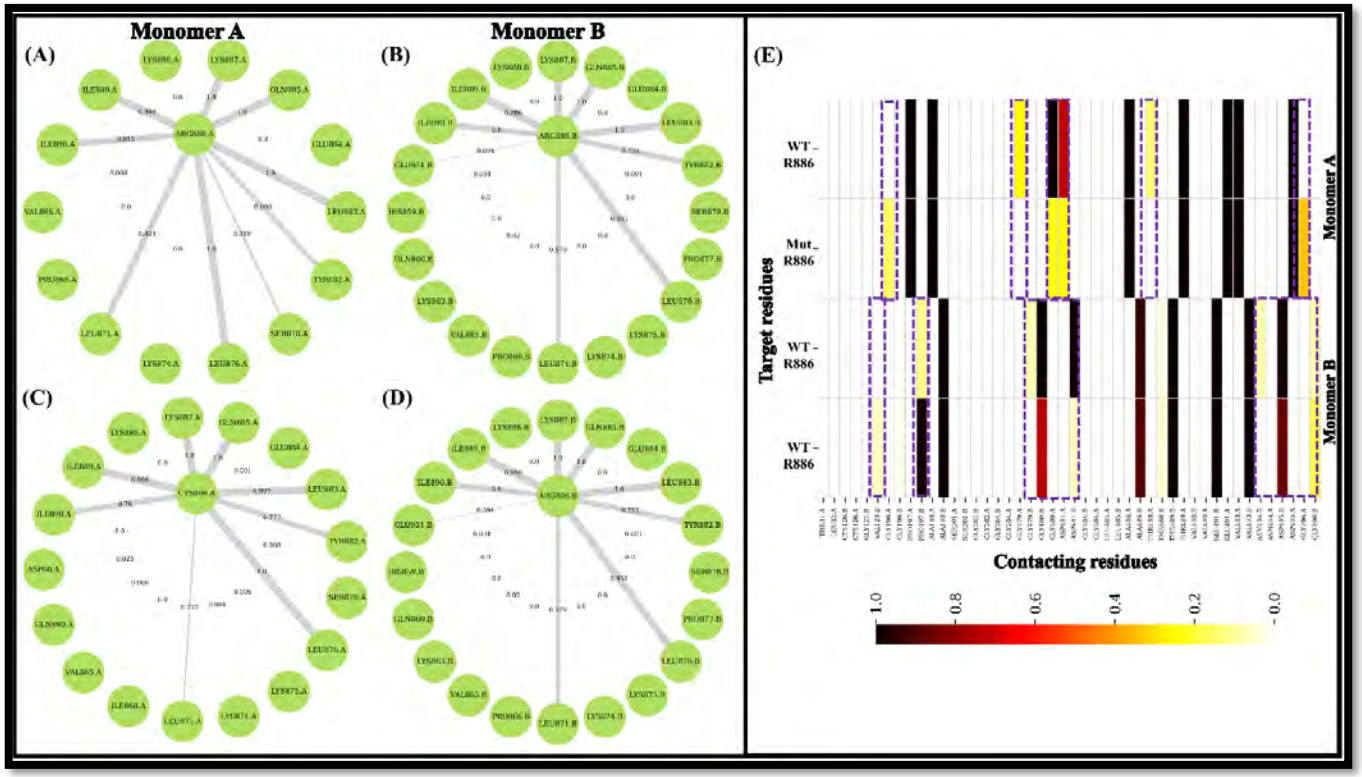


Figure S4.4.6. Contact-map representing frequency (%) of residue-residue interaction around Monomer A and B R886 of the (WT) in comparison to the mutation (R886C). (A) Weighted residues contact map around R886 (WT) of Monomer A, (B) Weighted residue contact map around R886 (WT) of Monomer B, (C) Weighted residue contact map around C886 (mutation) of Monomer A, (D) Weighted residue contact map around C886 (mutation) of Monomer B and (E) Weighted residues contact heatmap showing the inter-residue interaction around R886 (WT) and C886 (mutation) of Monomer A and Monomer B of DPD. The shaded boxes in each frame represent significant differences in the frequency of interaction between the R886 residues and the surrounding contacting residues in the WT versus the mutations. White (0) to dark-red (1.0) represent no-contact (0) to strong-contact (1.0)

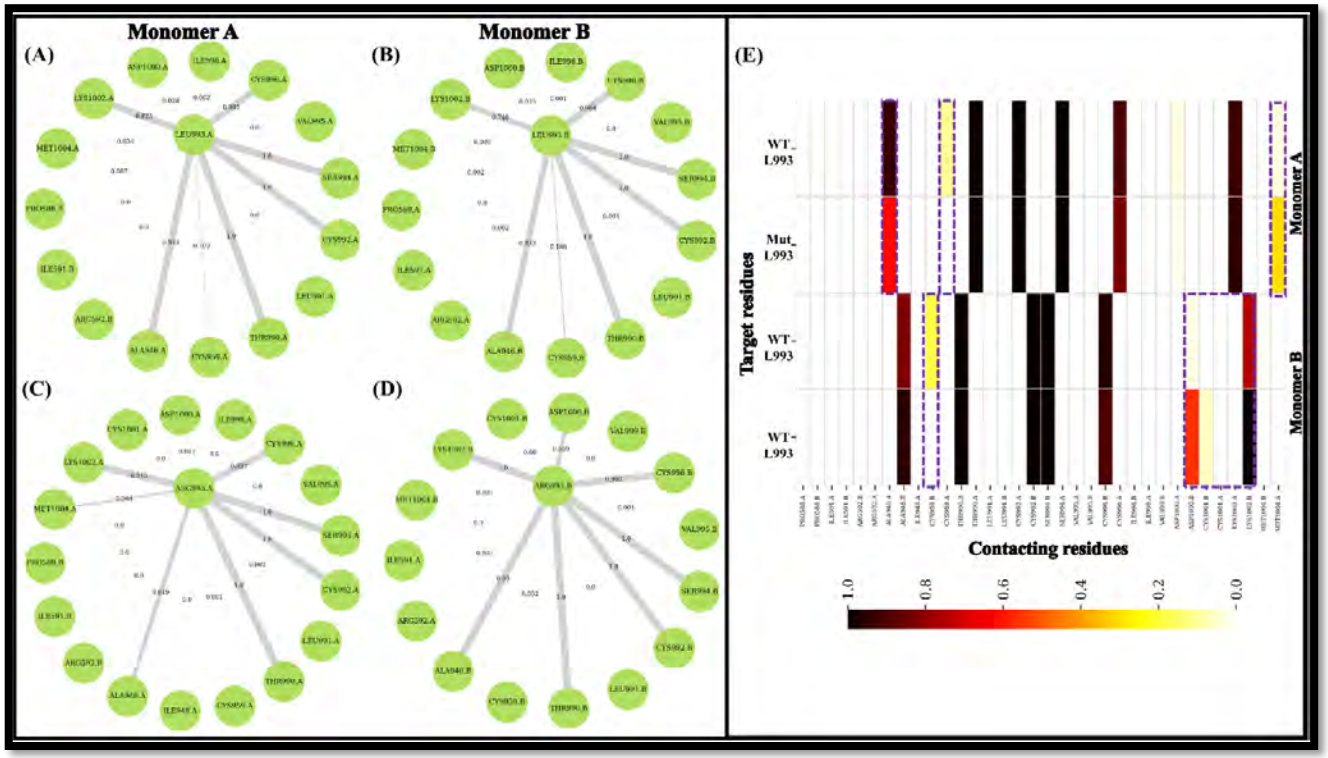


Figure S4.4.7. Contact-map representing frequency (%) of residue-residue interaction around Monomer A and B L993 of the (WT) in comparison to the mutation (L993R). (A) Weighted residues contact map around L993 (WT) of Monomer A, (B) Weighted residue contact map around L993 (WT) of Monomer B, (C) Weighted residue contact map around R993 (mutation) of Monomer A, (D) Weighted residue contact map around R993 (mutation) of Monomer B and (E) Weighted residues contact heatmap showing the inter-residue interaction around L993 (WT) and R993 (mutation) of Monomer A and Monomer B of DPD. The shaded boxes in each frame represent significant differences in the frequency of interaction between the L993 residues and the surrounding contacting residues in the WT versus the mutations. White (0) to dark-red (1.0) represent no-contact (0) to strong-contact (1.0)

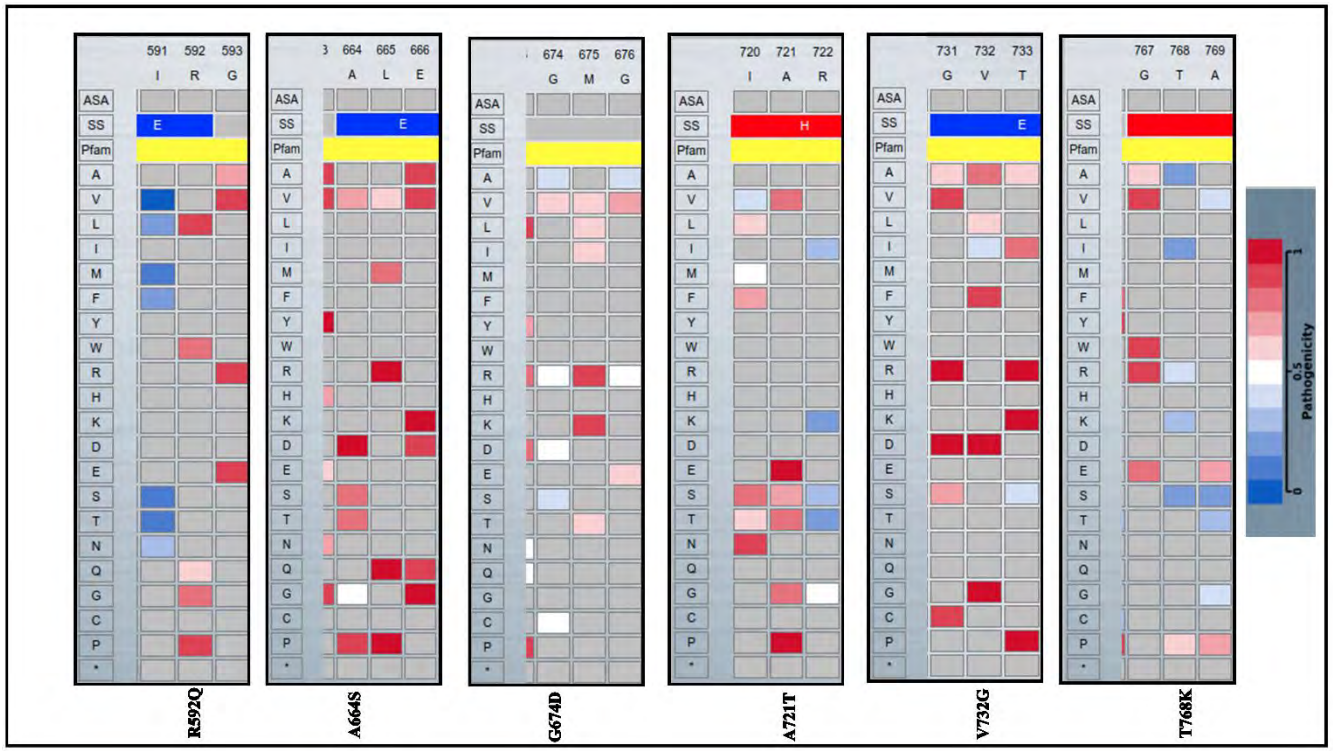


Figure S5.1.1: A heatmap displaying the sequence-based pathogenicity effects probability for active site mutants of uncertain significance (MUS) (R592Q, A664S, G674D, A721T, V732G, T768K) generated by the Ensembl Variant Effect Predictor (EVE). The heatmap is color-coded from blue to white, representing a range from 0 to 1, indicating an ascending likelihood of pathogenicity. Mutations assigned a value of 0.5 on the heatmap are considered moderately pathogenic, while those assigned a value of 1 are highly likely to be pathogenic. Conversely, mutations with a value of 0 are less likely to be pathogenic.

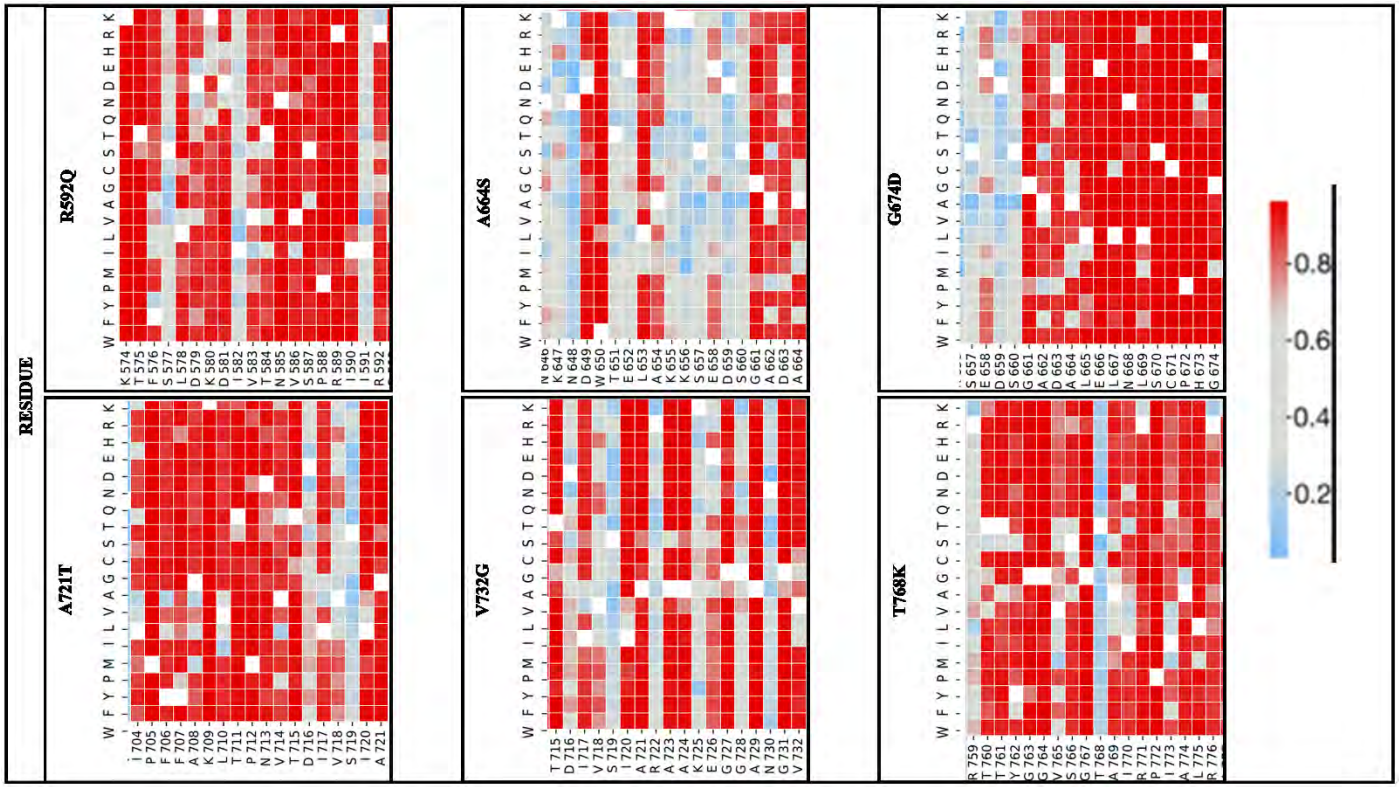


Figure S5.1.2: A heatmap displaying the sequence-based pathogenicity effects probability for active site mutants of uncertain significance (MUS) (R592Q, A664S, G674D, A721T, V732G, T768K) generated by the Variant Annotation Integrator (VARITY). The heatmap is color-coded from blue to white, representing a range from 0 to 1, indicating an ascending likelihood of pathogenicity. Mutations assigned a value of 0.5 on the heatmap are considered moderately pathogenic, while those assigned a value of 1 are highly likely to be pathogenic. Conversely, mutations with a value of 0 are less likely to be pathogenic.

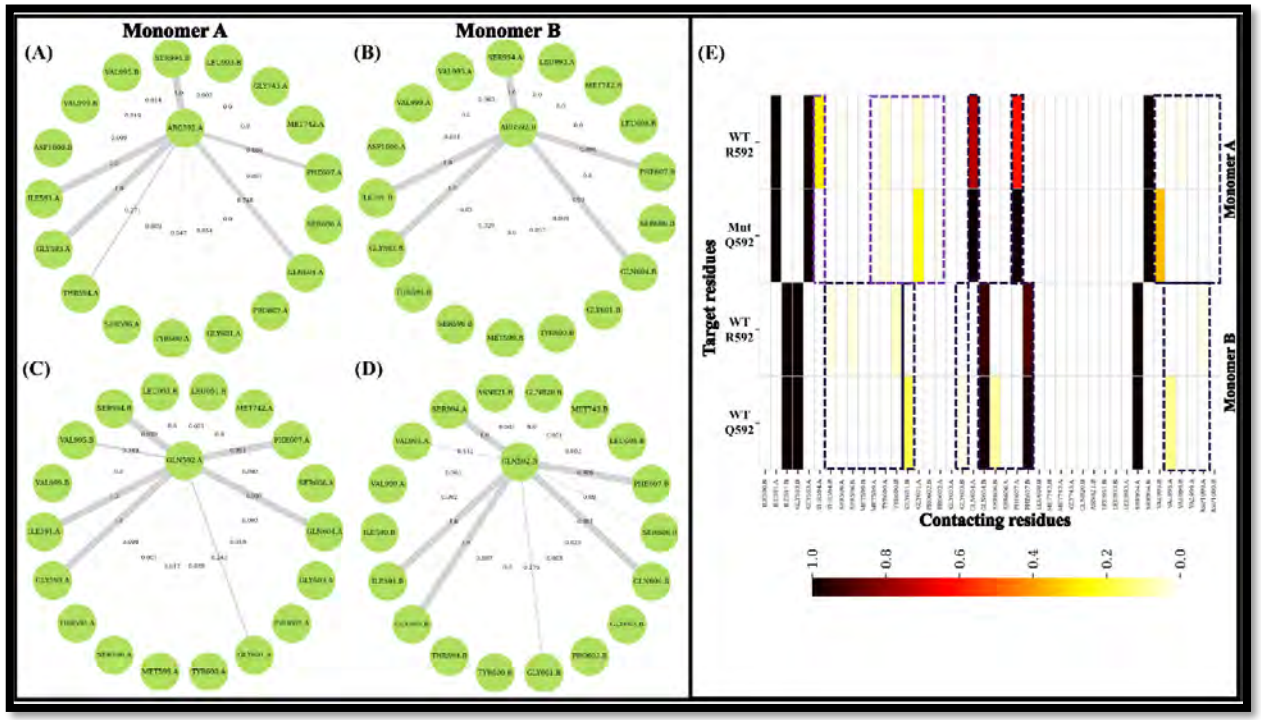


Figure S5.2.1. Contact-map representing frequency (%) of residue-residue interaction around Monomer A and B R592 of the (WT) in comparison to the mutation (R592Q). (A) Weighted residues contact map around R592 (WT) of Monomer A, (B) Weighted residue contact map around R592 (WT) of Monomer B, (C) Weighted residue contact map around R592 (mutation) of Monomer A, (D) Weighted residue contact map around R592 (mutation) of Monomer B and (E) Weighted residues contact heatmap showing the inter-residue interaction around R592 (WT) and R592 (mutation) of Monomer A and Monomer B of DPD. The shaded boxes in each frame represent significant differences in the frequency of interaction between the R592residues and the surrounding contacting residues in the WT versus the mutations. White (0) to dark-red (1.0) represent no-contact (0) to strong-contact (1.0)

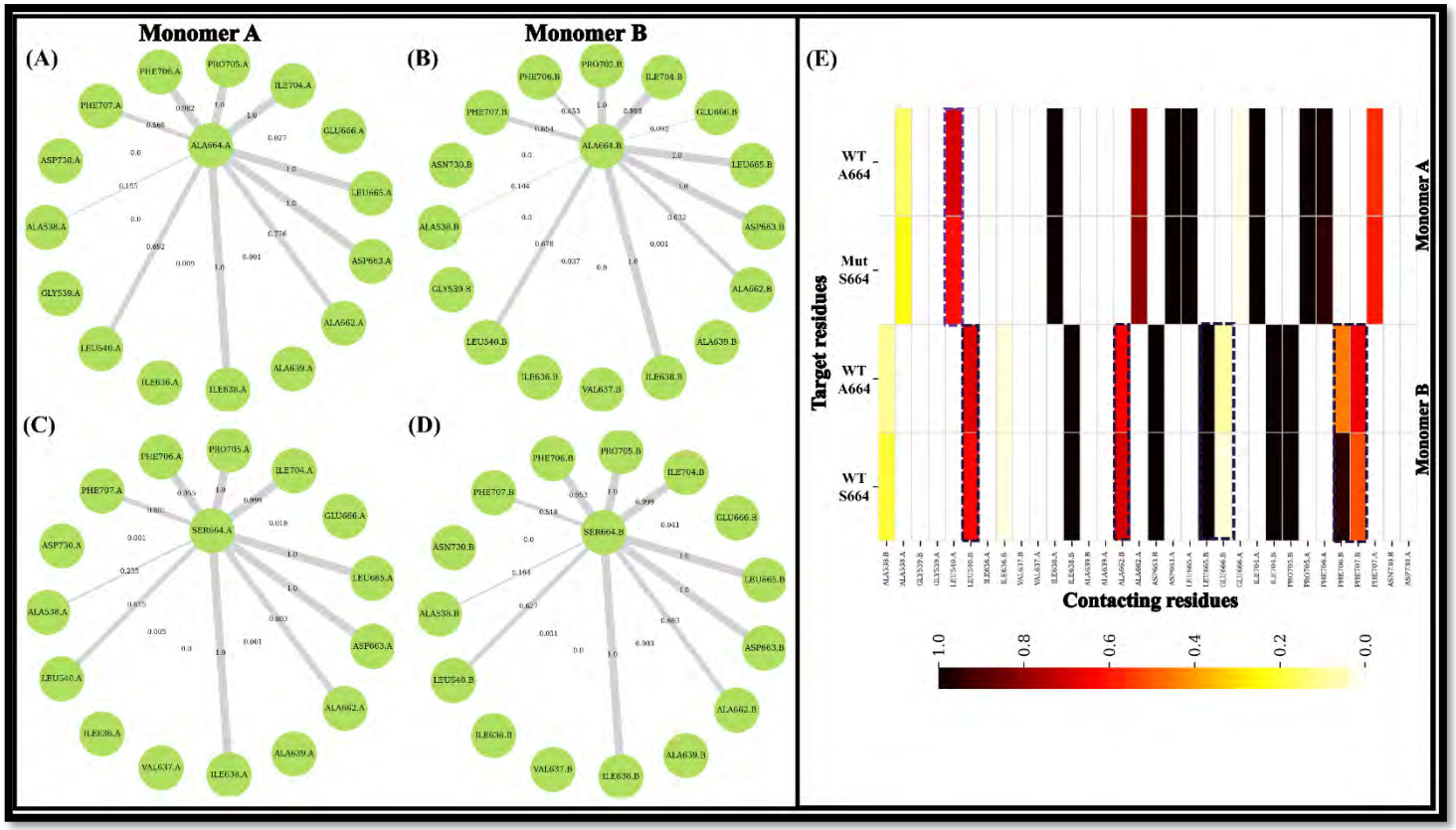


Figure S5.2.2. Contact-map representing frequency (%) of residue-residue interaction around Monomer A and B A664 of the (WT) in comparison to the mutation (A664S). (A) Weighted residues contact map around A664 (WT) of Monomer A, (B) Weighted residue contact map around A664 (WT) of Monomer B, (C) Weighted residue contact map around A664 (mutation) of Monomer A, (D) Weighted residue contact map around A664 (mutation) of Monomer B and (E) Weighted residues contact heatmap showing the inter-residue interaction around A664 (WT) and A664 (mutation) of Monomer A and Monomer B of DPD. The shaded boxes in each frame represent significant differences in the frequency of interaction between the A664 residues and the surrounding contacting residues in the WT versus the mutations. White (0) to dark-red (1.0) represent no-contact (0) to strong-contact (1.0).

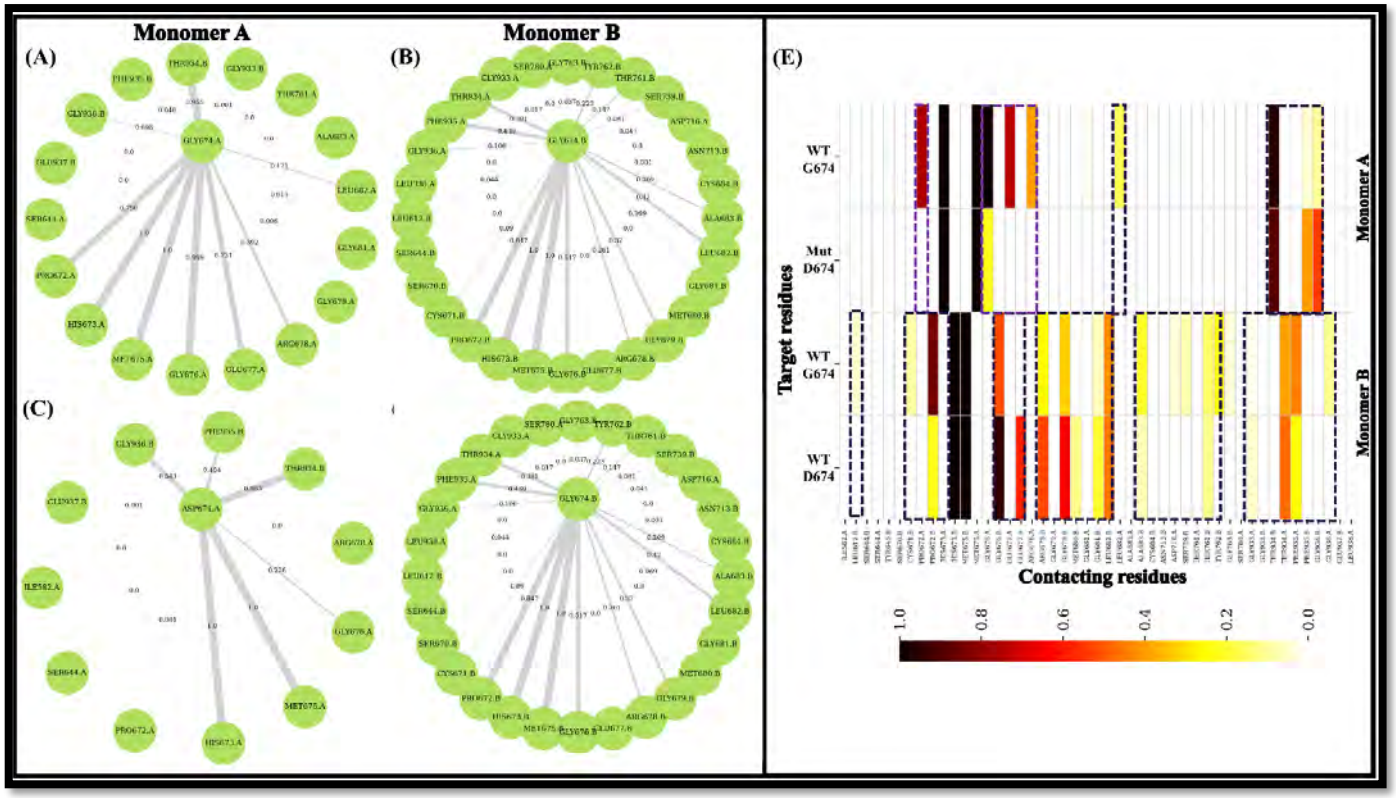


Figure S5.2.3. Contact-map representing frequency (%) of residue-residue interaction around Monomer A and B G674 of the (WT) in comparison to the mutation (G674D). (A) Weighted residues contact map around G674 (WT) of Monomer A, (B) Weighted residue contact map around G674 (WT) of Monomer B, (C) Weighted residue contact map around G674 (mutation) of Monomer A, (D) Weighted residue contact map around G674 (mutation) of Monomer B and (E) Weighted residues contact heatmap showing the inter-residue interaction around G674 (WT) and G674 (mutation) of Monomer A and Monomer B of DPD. The shaded boxes in each frame represent significant differences in the frequency of interaction between the G674 residues and the surrounding contacting residues in the WT versus the mutations. White (0) to dark-red (1.0) represent no-contact (0) to strong-contact (1.0).

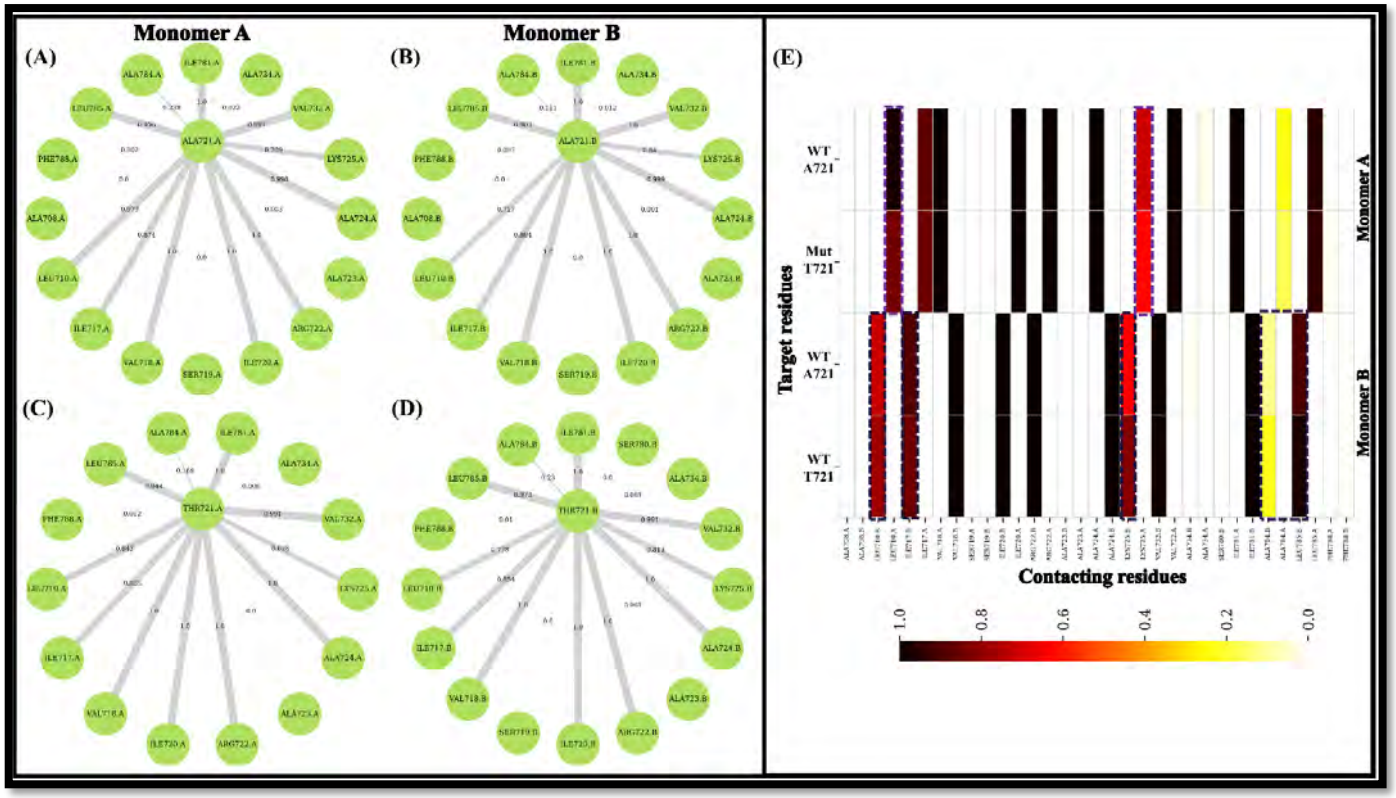


Figure S5.2.4. Contact-map representing frequency (%) of residue-residue interaction around Monomer A and B A721 of the (WT) in comparison to the mutation (A721T). (A) Weighted residues contact map around A721 (WT) of Monomer A, (B) Weighted residue contact map around A721 (WT) of Monomer B, (C) Weighted residue contact map around A721 (mutation) of Monomer A, (D) Weighted residue contact map around A721 (mutation) of Monomer B and (E) Weighted residues contact heatmap showing the inter-residue interaction around G674 (WT) and A7214 (mutation) of Monomer A and Monomer B of DPD. The shaded boxes in each frame represent significant differences in the frequency of interaction between the A721 residues and the surrounding contacting residues in the WT versus the mutations. White (0) to dark-red (1.0) represent no-contact (0) to strong-contact (1.0)

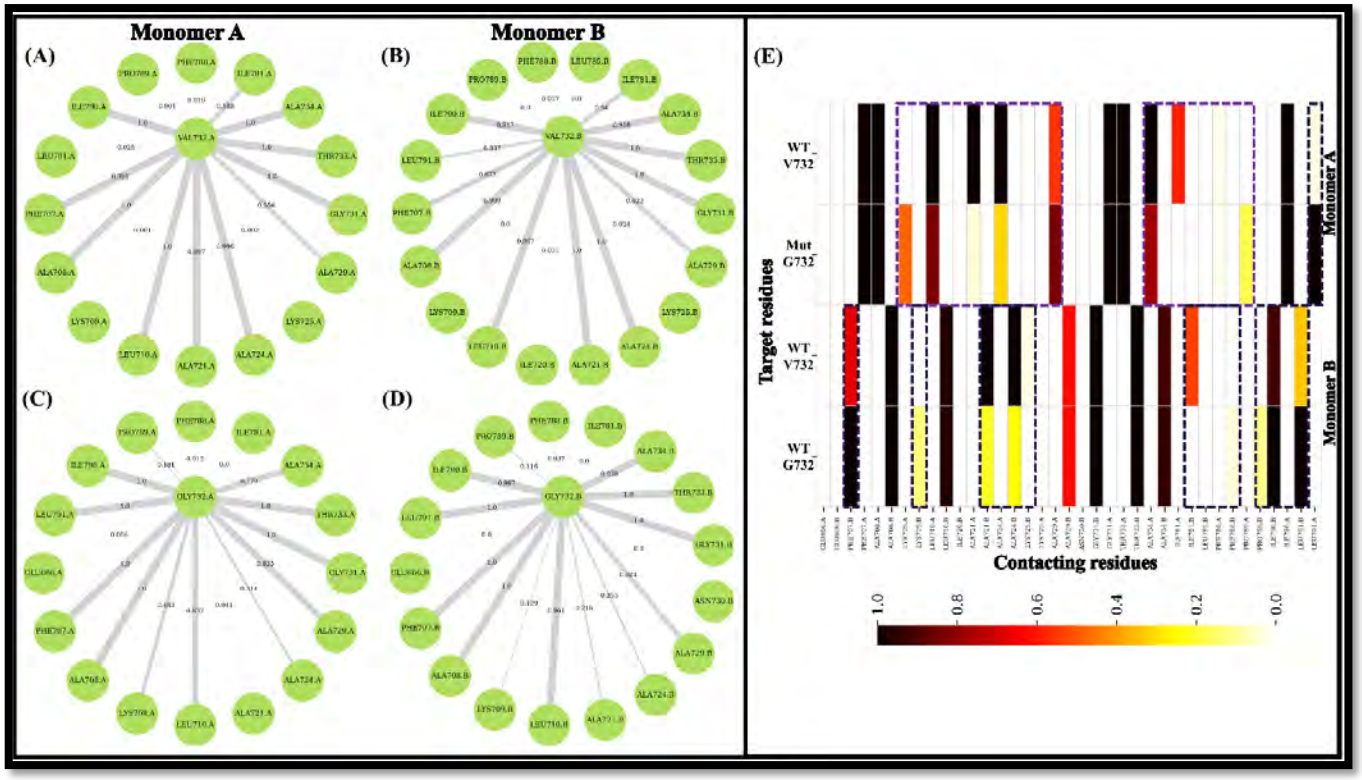


Figure S5.2.5. Contact-map representing frequency (%) of residue-residue interaction around Monomer A and B V732 of the (WT) in comparison to the mutation (V732G). (A) Weighted residues contact map around V732 (WT) of Monomer A, (B) Weighted residue contact map around V732 (WT) of Monomer B, (C) Weighted residue contact map around V732 (mutation) of Monomer A, (D) Weighted residue contact map around V732 (mutation) of Monomer B and (E) Weighted residues contact heatmap showing the inter-residue interaction around V732 (WT) and V732 (mutation) of Monomer A and Monomer B of DPD. The shaded boxes in each frame represent significant differences in the frequency of interaction between the A721 residues and the surrounding contacting residues in the WT versus the mutations. White (0) to dark-red (1.0) represent no-contact (0) to strong-contact (1.0)

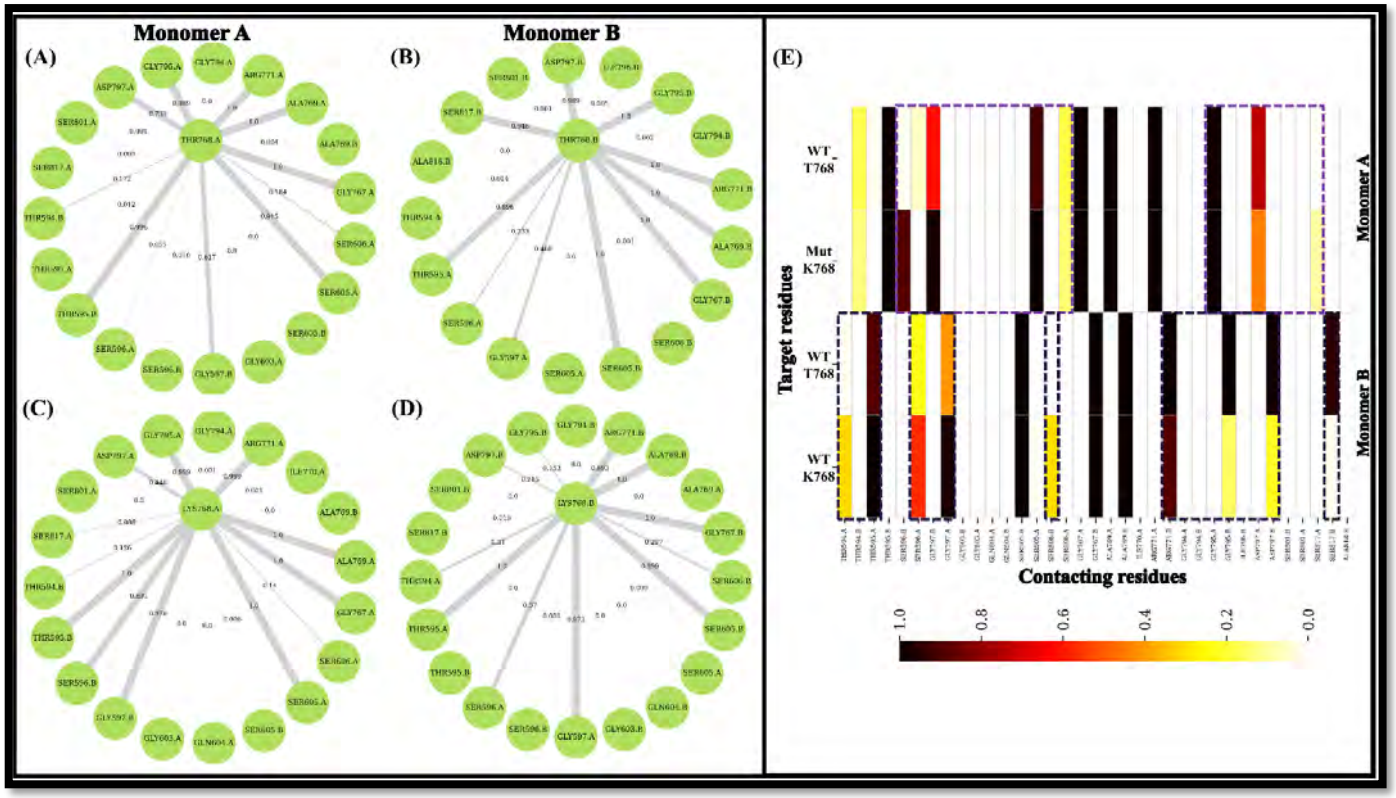


Figure S5.2.6. Contact-map representing frequency (%) of residue-residue interaction around Monomer A and B T768 of the (WT) in comparison to the mutation (T768K). (A) Weighted residues contact map around T768 (WT) of Monomer A, (B) Weighted residue contact map around T768 (WT) of Monomer B, (C) Weighted residue contact map around T768 (mutation) of Monomer A, (D) Weighted residue contact map around V732 (mutation) of Monomer B and (E) Weighted residues contact heatmap showing the inter-residue interaction around T768 (WT) and T768 (mutation) of Monomer A and Monomer B of DPD. The shaded boxes in each frame represent significant differences in the frequency of interaction between the T768 residues and the surrounding contacting residues in the WT versus the mutations. White (0) to dark-red (1.0) represent no-contact (0) to strong-contact (1.0)

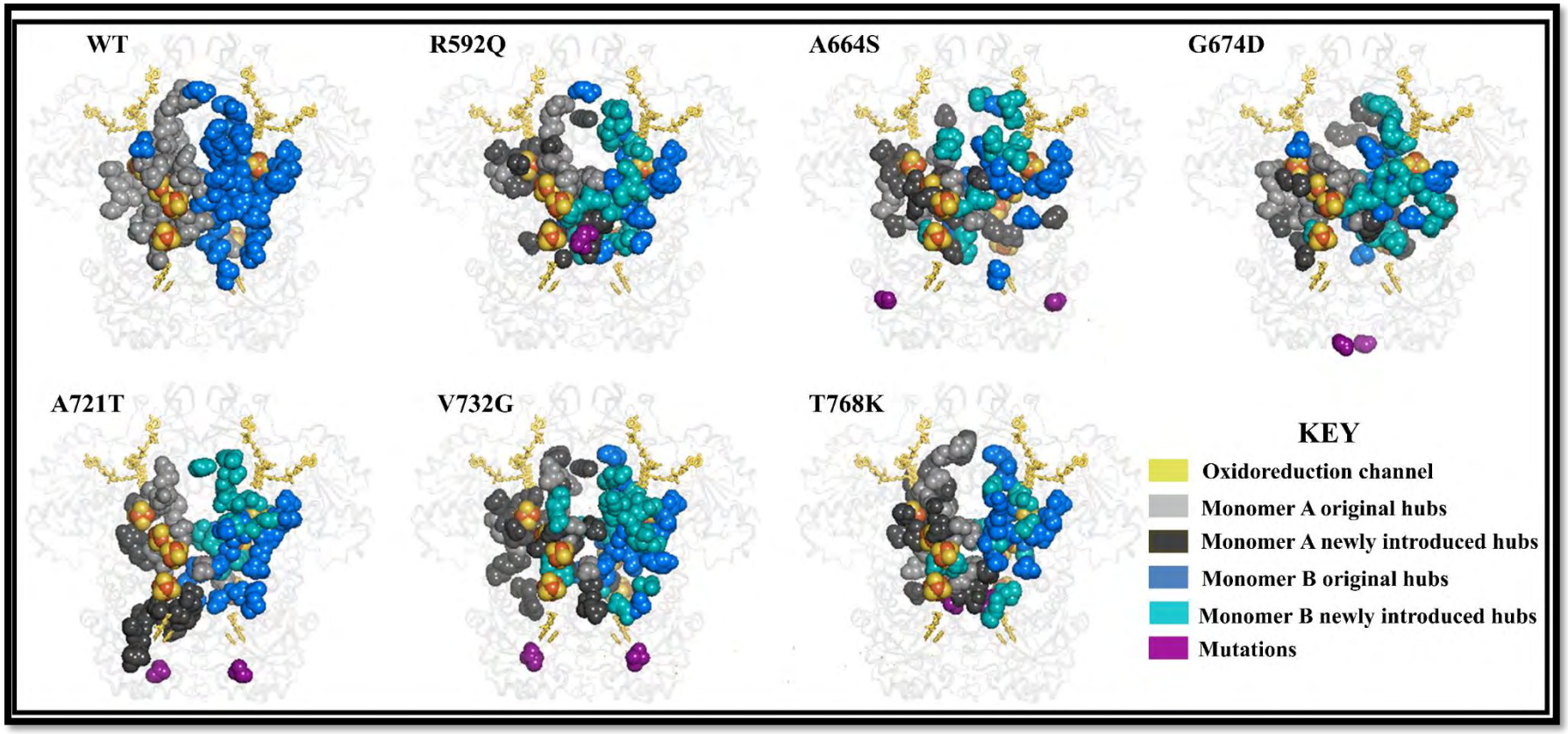


Figure S5.3.1. 3D mapping of the global top 4 % for averaged *BC* metric of the WT and mutations in DPD drug free ensembles. A communication pathway along oxidoreduction (gold) pathway from Monomer A NADPH to FAD to 2 x $Fe^{2+}{}_4S^{2-}{}_4$ clusters, proceeding to Monomer B 2 x $Fe^{2+}{}_4S^{2-}{}_4$ to FMN for the reduction of 5-FU. The hubs in Monomer A are colored light grey (original) and dark grey (newly introduced) and Monomer A are colored sky blue (original) and royal blue (newly introduced)

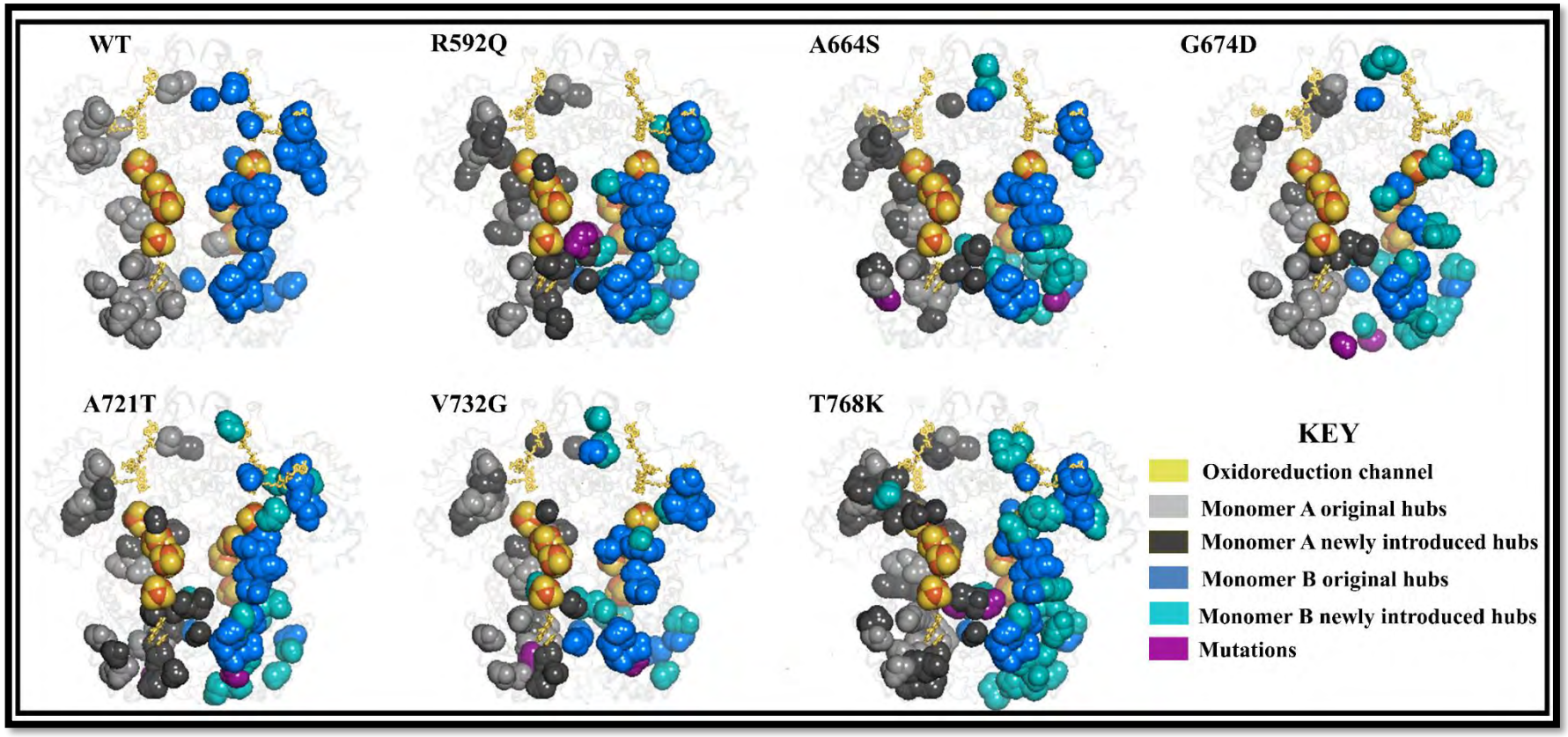


Figure S5.3.2. **3D mapping of the global top 4 % for averaged *DC* metric of the WT and mutations in DPD drug free ensembles.** A communication pathway along oxidoreduction (gold) pathway from Monomer A NADPH to FAD to 2 x $Fe^{2+}{}_4S^{2-}{}_4$ clusters, proceeding to Monomer B 2 x $Fe^{2+}{}_4S^{2-}{}_4$ to FMN for the reduction of 5-FU. The hubs in Monomer A are colored light grey (original) and dark grey (newly introduced) and Monomer A are colored sky blue (original) and royal blue (newly introduced)

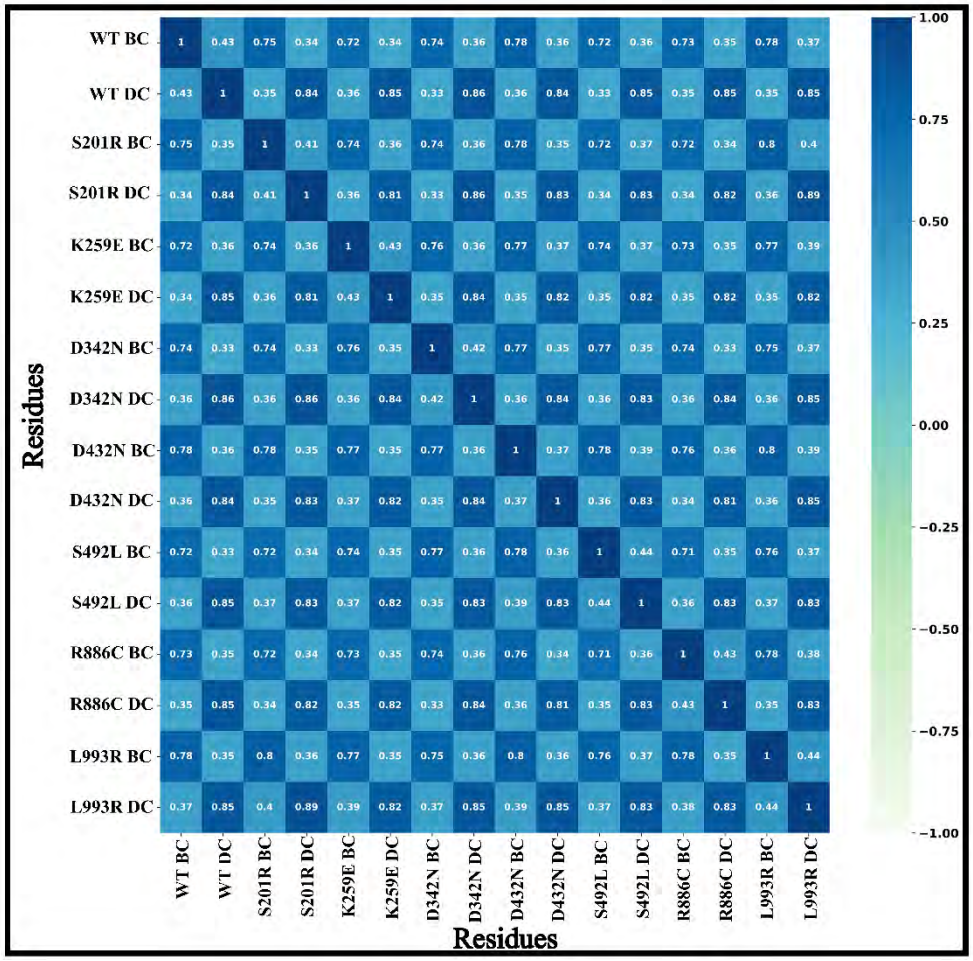


Figure S5.3.3. **A direct cross-correlation (DCC) heatmap showing the relationship between WT and mutations in DPD drug free ensembles.**

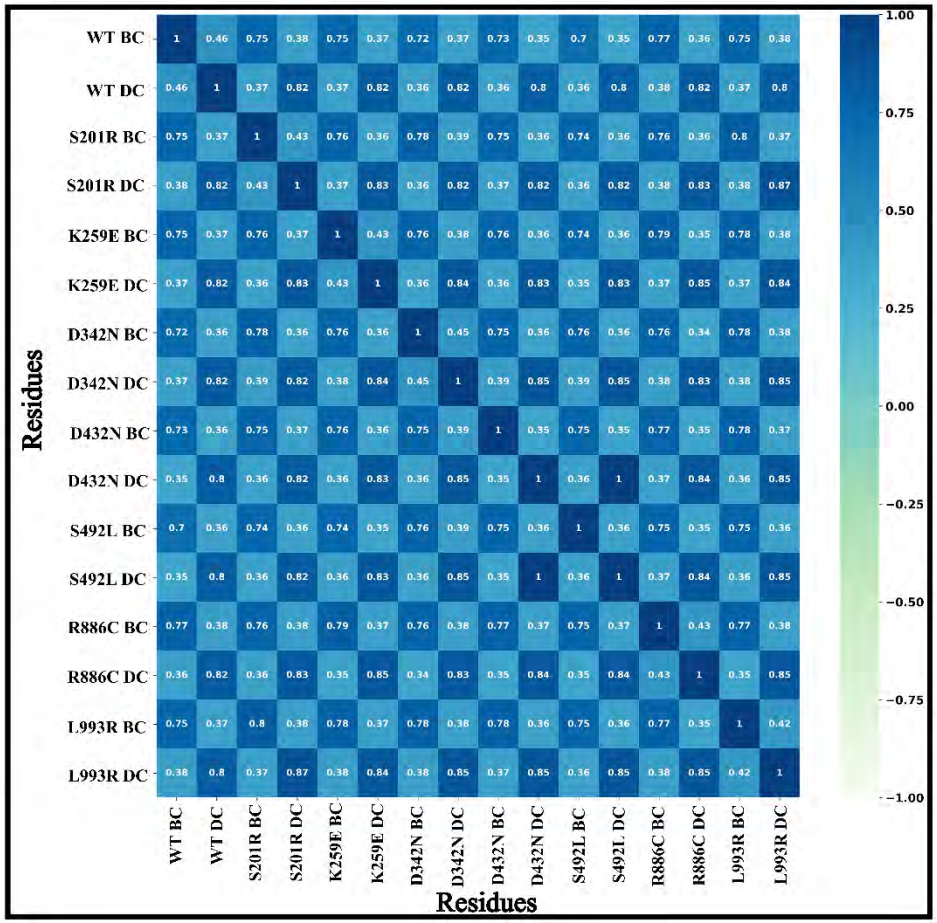


Figure S5.3.4. A direct cross-correlation (DCC) heatmap showing the relationship between WT and mutations in DPD drug bound ensembles.

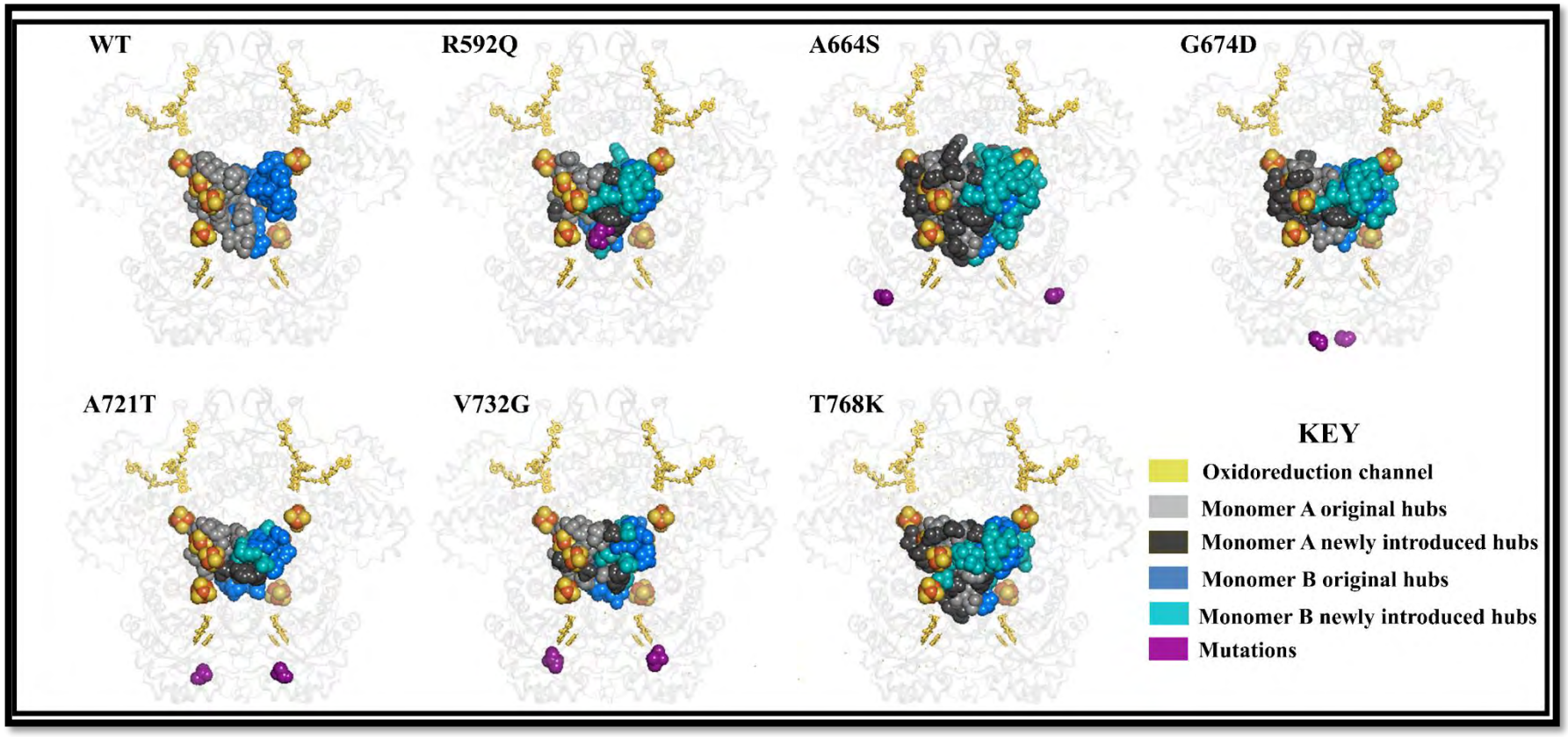


Figure S5.3.5. **3D mapping of the global top 4 % for averaged *CC* metric of the WT and mutations in DPD drug free ensembles.** A communication pathway along oxidoreduction (gold) pathway from Monomer A NADPH to FAD to 2 x $Fe^{2+}{}_4S^{2-}{}_4$ clusters, proceeding to Monomer B 2 x $Fe^{2+}{}_4S^{2-}{}_4$ to FMN for the reduction of 5-FU. The hubs in Monomer A are colored light grey (original) and dark grey (newly introduced) and Monomer A are colored sky blue (original) and royal blue (newly introduced)

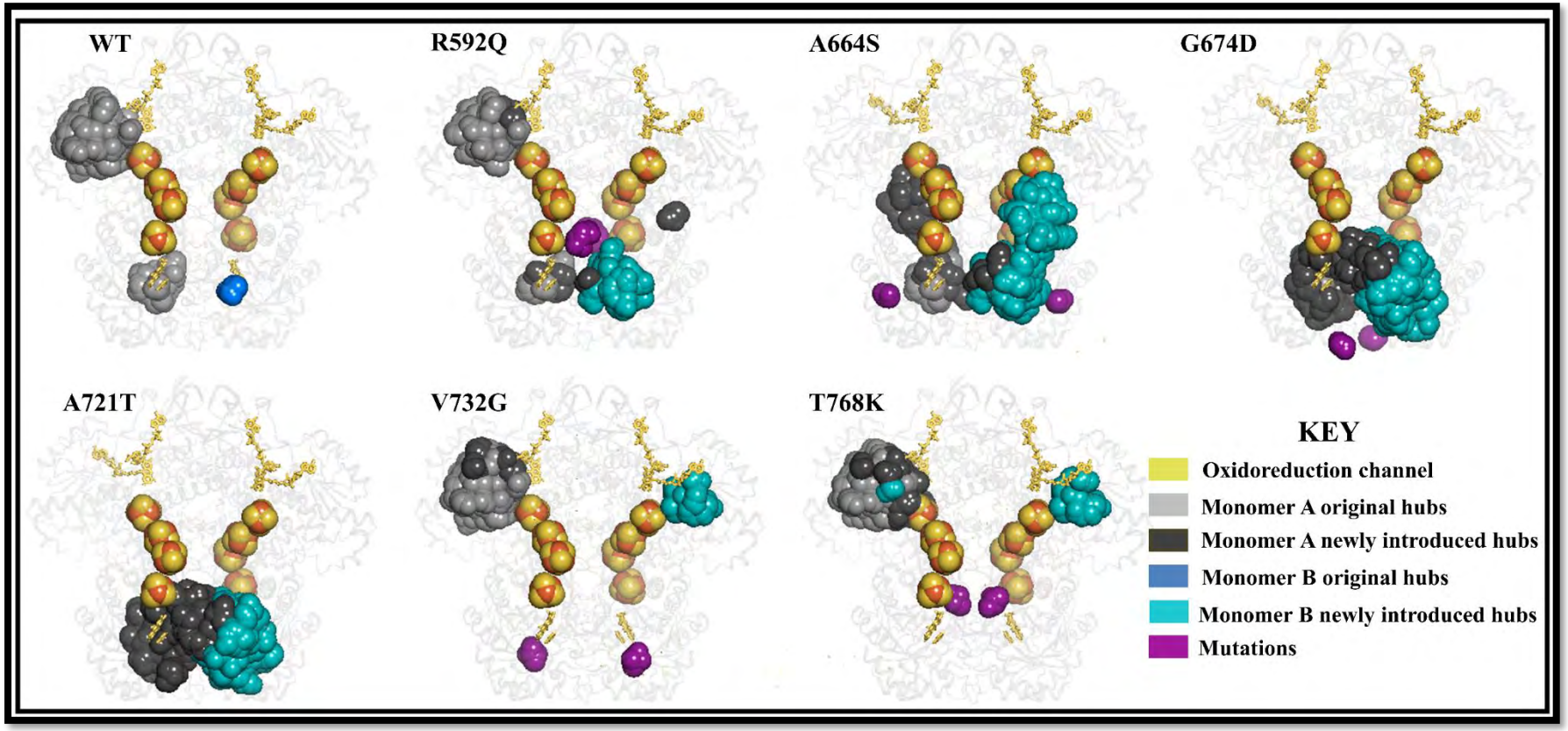


Figure S5.3.6. 3D mapping of the global top 4 % for averaged *EC* metric of the WT and mutations in DPD drug free ensembles. A communication pathway along oxidoreduction (gold) pathway from Monomer A NADPH to FAD to 2 x $Fe^{2+}{}_4S^{2-}{}_4$ clusters, proceeding to Monomer B 2 x $Fe^{2+}{}_4S^{2-}{}_4$ to FMN for the reduction of 5-FU. The hubs in Monomer A are colored light grey (original) and dark grey (newly introduced) and Monomer A are colored sky blue (original) and royal blue (newly introduced).

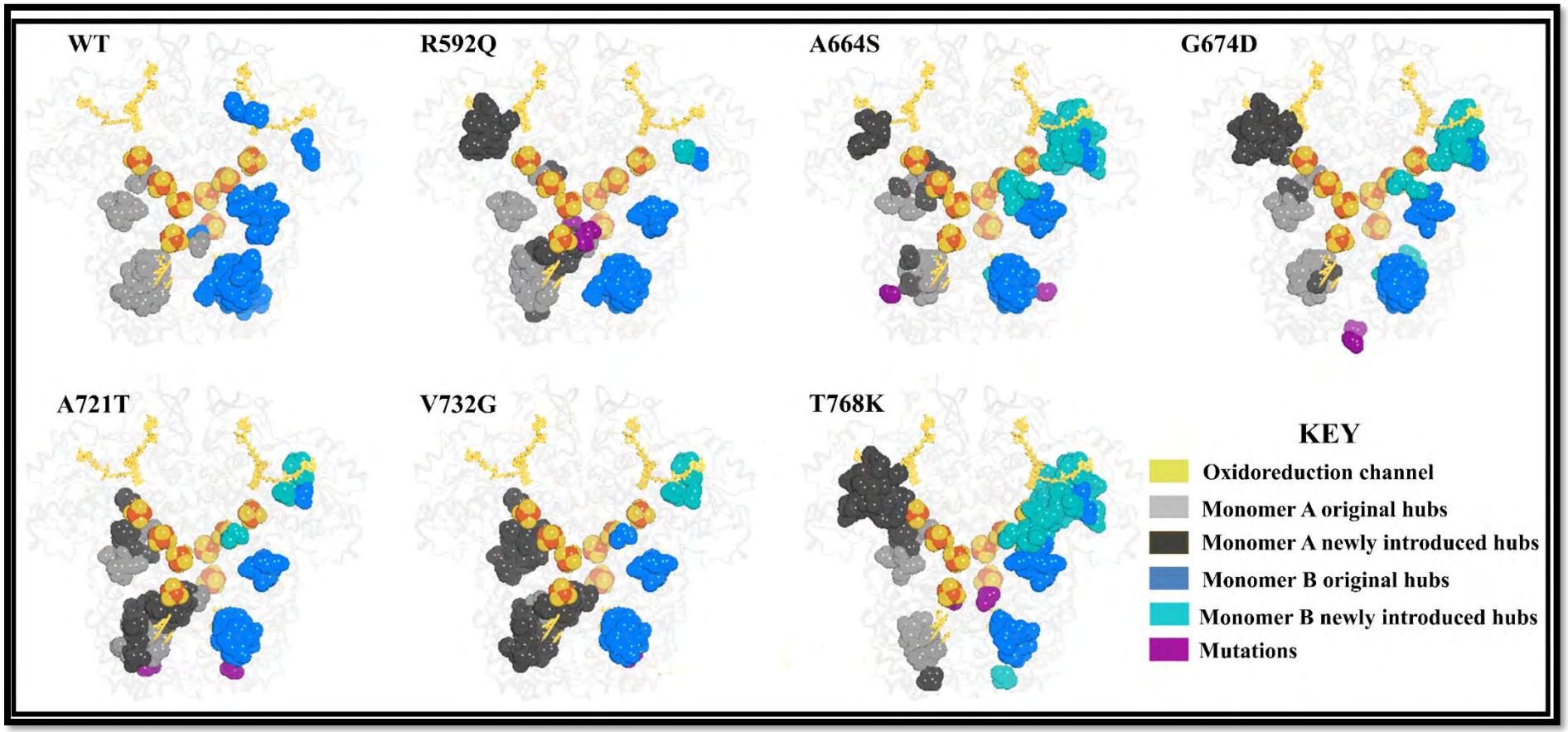


Figure S5.3.7. 3D mapping of the global top 4 % for averaged *KC* metric of the WT and mutations in DPD drug free ensembles. A communication pathway along oxidoreduction (gold) pathway from Monomer A NADPH to FAD to 2 x $Fe^{2+}{}_4S^{2-}{}_4$ clusters, proceeding to Monomer B 2 x $Fe^{2+}{}_4S^{2-}{}_4$ to FMN for the reduction of 5-FU. The hubs in Monomer A are colored light grey (original) and dark grey (newly introduced) and Monomer A are colored sky blue (original) and royal blue (newly introduced)

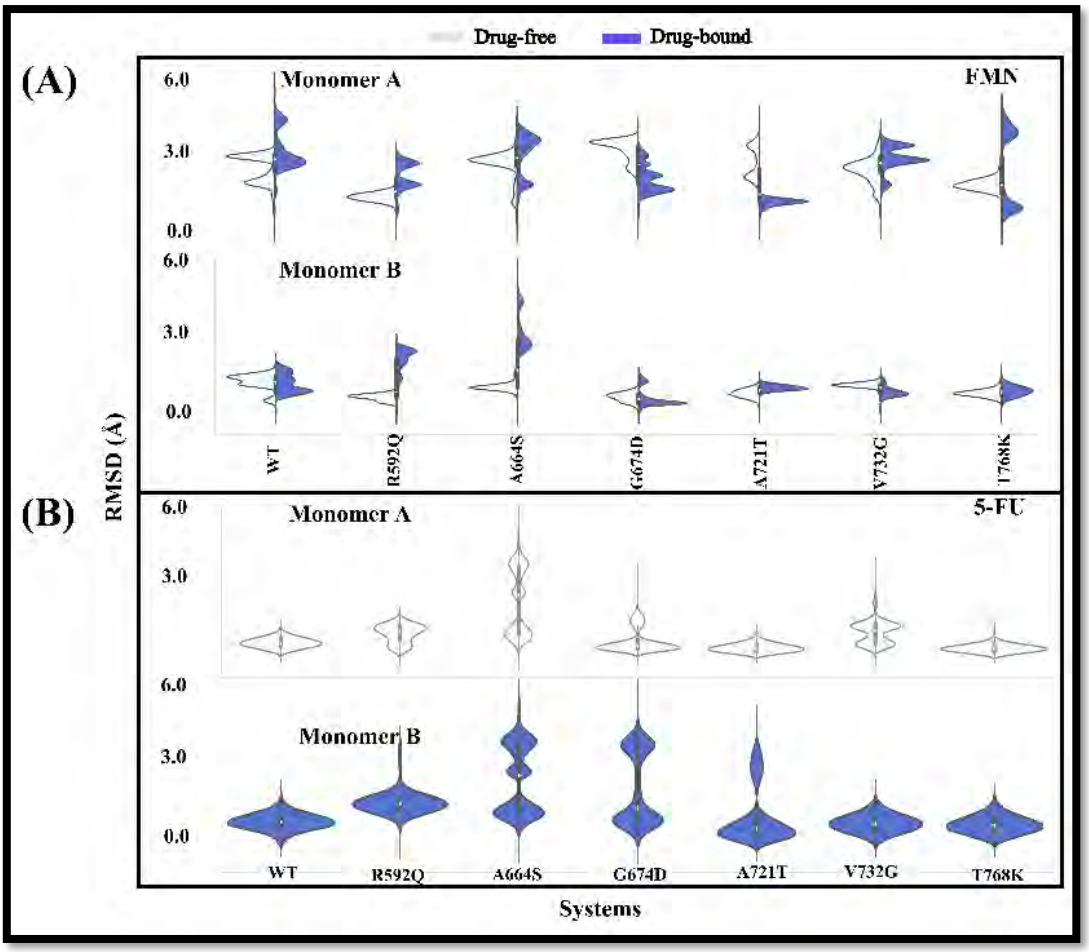


Figure S4.-. split Violin plots depicting 600 ns RMSD simulation for drug free and drug bound systems of cofactor (FMN) and ligand (5-FU)

Table S2.1: Quality validation of Human DPD protein modeled structures. Table adapted from Tendwa et al., (2021)

PROTEINS	VERIFY 3D (%)	QMEAN6	ProSA		PROCHECK (%)		
	3D-ID score	QMEAN score	Monomers z-Score		Ramachandran (residues location)		
			A	B	Favored	Allowed	Disallowed
Template (PDB ID: 1H7X)	85.79	0.91	-13.56	-13.47	93.5	6.4	0.1
DPD Holo	85.01	0.90	-13.41	-13.44	89.4	10.2	0.0
DPD Holo-5-FU drug	85.01	0.89	- 13.42	-13.45	92.2	7.7	0.1

Table S2.2. Titratable residues in the human DPD protein and their respective pKa values. Table adapted from Tendwa et al., (2021)

Residue	pK (int)	pKa_(1/2)	Residue	pK (int)	pKa_(1/2)	Residue	pK (int)	pKa_(1/2)
	Chain-A							
NTALA-2	7.417	7.972	ARG-678	11.484	>12.000	CYS-324	8.990	9.066
LYS-7	10.092	11.678	CYS-684	11.950	>12.000	HIS-325	6.186	6.037
ASP-8	4.515	0.918	ASP-687	3.873	1.358	ARG-332	12.252	>12.000
ASP-10	3.600	3.978	GLU-689	4.008	3.203	ASP-342	2.523	<0.000
ASP-11	3.436	3.435	ARG-692	11.329	>12.000	ASP-346	4.452	<0.000
GLU-13	4.390	4.841	CYS-695	11.905	>12.000	CYS-347	9.618	>12.000
ARG-21	11.752	>12.000	ARG-696	11.889	>12.000	ARG-353	12.006	>12.000
HIS-25	6.649	7.179	ARG-699	12.639	>12.000	CYS-354	12.003	>12.000
CYS-29	9.529	10.217	LYS-709	6.035	11.023	ARG-357	11.430	>12.000
LYS-34	9.536	>12.000	ASP-716	2.681	<0.000	ARG-358	11.391	>12.000
LYS-35	9.563	>12.000	ARG-722	11.424	>12.000	ARG-364	11.276	>12.000
ASP-37	5.142	1.200	LYS-725	10.304	10.679	LYS-365	9.534	9.163
LYS-38	9.329	>12.000	GLU-726	5.068	2.383	ARG-371	11.195	>12.000
LYS-39	10.143	10.937	ASP-730	5.918	2.792	GLU-375	3.112	<0.000
HIS-40	6.461	6.871	LYS-745	9.516	>12.000	GLU-376	4.400	6.860
LYS-42	9.998	>12.000	ASP-747	3.896	<0.000	GLU-378	4.413	<0.000
ARG-43	13.175	>12.000	LYS-758	10.327	11.421	LYS-381	8.938	>12.000
ASP-46	4.116	<0.000	ARG-759	12.228	>12.000	GLU-382	5.675	<0.000
LYS-47	9.719	>12.000	TYR-762	14.594	>12.000	GLU-383	6.836	<0.000
CYS-49	9.248	9.348	ARG-771	10.083	>12.000	LYS-384	10.693	>12.000
CYS-52	10.767	>12.000	ARG-776	10.282	>12.000	CYS-385	11.195	>12.000
GLU-53	5.888	0.730	ARG-783	11.892	>12.000	GLU-386	5.553	<0.000
LYS-54	9.683	10.896	ASP-797	4.647	<0.000	ARG-394	11.524	>12.000
GLU-56	6.258	<0.000	GLU-800	3.816	<0.000	LYS-395	10.171	10.585
ASP-60	5.726	<0.000	HIS-807	5.207	3.368	LYS-399	10.108	10.435
ASP-61	5.047	<0.000	CYS-816	7.570	7.492	ARG-402	12.151	>12.000
LYS-63	9.179	>12.000	ASP-823	2.599	<0.000	ARG-410	11.826	>12.000

HIS-64	5.720	9.050	GLU-828	3.751	5.528	GLU-412	5.136	0.902
GLU-69	3.222	<0.000	ASP-829	1.478	<0.000	ASP-414	3.839	2.274
ARG-70	8.644	>12.000	TYR-830	10.911	>12.000	GLU-415	4.126	3.367
ARG-74	10.915	>12.000	CYS-831	9.358	>12.000	LYS-418	10.218	10.910
GLU-75	4.124	<0.000	LYS-835	9.380	>12.000	GLU-421	4.361	1.315
ARG-78	10.937	>12.000	TYR-839	12.375	>12.000	ASP-422	4.847	5.473
CYS-79	7.155	8.371	LYS-841	9.611	>12.000	GLU-423	5.357	3.329
LYS-81	9.191	>12.000	GLU-844	4.631	3.112	ASP-424	3.742	2.658
CYS-82	6.821	>12.000	GLU-845	5.283	5.206	HID-428	6.245	6.107
ASP-84	5.617	<0.000	ASP-848	4.229	3.553	LYS-430	9.983	>12.000
CYS-87	5.793	>12.000	ASP-850	3.751	<0.000	ASP-432	4.667	2.550
LYS-89	10.120	>12.000	HIS-859	2.394	<0.000	ASP-444	2.973	1.331
CYS-91	6.640	8.917	LYS-861	10.276	>12.000	LYS-446	9.801	10.964
HIS-94	5.713	0.409	LYS-863	10.097	>12.000	LYS-448	9.760	11.020
ASP-96	2.303	<0.000	ARG-867	11.666	>12.000	GLU-449	4.386	4.021
LYS-98	6.466	>12.000	GLU-870	6.530	3.817	LYS-455	10.222	11.077
LYS-107	8.251	>12.000	ASP-873	3.827	3.252	ARG-458	11.797	>12.000
TYR-109	12.293	>12.000	LYS-874	10.222	11.604	GLU-463	4.330	4.038
TYR-110	11.905	>12.000	LYS-875	10.158	>12.000	ASP-465	3.138	2.845
LYS-114	8.864	>12.000	TYR-882	11.326	>12.000	GLU-467	3.882	3.682
ASP-119	5.422	<0.000	GLU-884	4.874	2.698	GLU-473	4.164	3.200
CYS-126	6.601	>12.000	ARG-886	11.628	>12.000	ASP-481	1.702	<0.000
CYS-130	6.202	>12.000	LYS-887	10.151	>12.000	GLU-491	3.011	<0.000
ASP-134	4.533	<0.000	LYS-888	10.369	11.315	ASP-495	2.170	<0.000
CYS-136	6.036	>12.000	GLU-892	5.494	4.079	LYS-497	8.758	>12.000
CYS-140	6.456	5.546	GLU-893	5.290	3.716	TYR-502	10.836	>12.000
TYR-143	10.672	>12.000	LYS-894	9.690	>12.000	HIS-504	5.309	4.651
GLU-146	4.597	<0.000	ARG-896	12.114	>12.000	LYS-505	10.092	10.876
GLU-147	5.224	<0.000	LYS-898	10.086	>12.000	TYR-506	11.696	>12.000
GLU-161	4.307	2.857	GLU-899	4.934	3.926	TYR-511	11.474	>12.000
LYS-164	9.536	>12.000	LYS-908	10.362	10.775	LYS-518	9.969	>12.000

ARG-172	13.132	>12.000	ARG-909	11.520	>12.000	GLU-520	5.117	2.691
GLU-180	4.415	3.790	CYS-911	10.130	11.404	TYR-525	11.106	>12.000
LYS-181	10.258	10.573	LYS-915	10.243	11.792	ASP-529	4.834	<0.000
GLU-184	4.462	3.974	ARG-916	11.939	>12.000	ASP-532	3.498	2.437
TYR-186	11.038	11.742	LYS-922	10.179	11.304	GLU-536	4.535	3.981
LYS-189	10.624	>12.000	ASP-923	4.440	2.744	LYS-541	10.366	11.337
CYS-202	10.875	>12.000	LYS-927	10.571	10.938	ARG-561	11.833	>12.000
ARG-208	11.892	>12.000	TYR-931	11.367	>12.000	ARG-562	9.217	>12.000
TYR-211	12.788	>12.000	GLU-937	4.620	2.157	GLU-565	4.518	3.413
ASP-213	4.946	1.769	GLU-942	4.228	3.649	LYS-574	7.509	>12.000
GLU-218	2.317	<0.000	ASP-949	3.028	0.531	ASP-579	3.747	2.915
LYS-219	8.693	>12.000	GLU-950	4.668	5.855	LYS-580	9.498	>12.000
GLU-221	4.132	3.561	GLU-951	5.702	1.520	ASP-581	4.111	3.531
TYR-222	9.094	>12.000	CYS-953	7.496	>12.000	ARG-589	8.869	>12.000
GLU-230	3.569	<0.000	CYS-956	7.107	1.932	ARG-592	12.314	>12.000
ARG-235	9.353	>12.000	LYS-958	7.074	>12.000	TYR-600	10.708	>12.000
TYR-238	11.091	>12.000	CYS-959	6.653	<0.000	GLU-611	5.412	<0.000
ASP-239	4.261	0.642	TYR-960	12.121	>12.000	GLU-615	4.869	<0.000
GLU-244	2.724	<0.000	CYS-963	5.350	11.686	LYS-616	8.766	>12.000
GLU-246	4.835	3.403	ASP-965	2.428	<0.000	TYR-620	10.801	>12.000
LYS-249	9.374	>12.000	TYR-968	11.555	>12.000	CYS-622	11.760	>12.000
ASP-250	4.804	3.488	ASP-974	2.649	0.129	GLU-627	5.247	3.649
LYS-254	10.056	>12.000	GLU-976	4.583	3.605	LYS-629	10.732	>12.000
CYS-257	9.490	>12.000	HIS-978	5.305	9.984	ASP-631	5.100	2.340
LYS-259	10.095	10.908	ASP-984	4.401	3.121	ASP-634	4.392	3.873
GLU-265	4.494	3.918	CYS-986	7.289	>12.000	CYS-643	10.096	>12.000
LYS-272	9.978	11.037	CYS-989	6.958	8.920	TYR-645	10.610	>12.000
GLU-273	4.767	3.907	CYS-992	5.732	<0.000	LYS-647	10.186	10.183
LYS-274	9.903	10.056	CYS-996	7.455	10.503	ASP-649	4.813	1.835
TYR-276	11.567	>12.000	ASP-1000	3.607	<0.000	GLU-652	4.490	4.296
LYS-277	10.604	11.424	CYS-1001	8.150	>12.000	LYS-655	10.192	11.038

GLU-287	3.240	2.792	LYS-1002	3.228	>12.000	LYS-656	9.621	>12.000
LYS-290	9.997	11.006	ARG-1003	10.588	>12.000	GLU-658	5.018	4.231
ASP-291	3.660	1.830	ARG-1007	11.908	>12.000	ASP-659	6.916	7.080
ASP-300	4.078	3.428	TYR-1011	12.126	>12.000	ASP-663	5.385	1.746
TYR-304	11.589	>12.000	GLU-1012	4.464	3.817	GLU-666	4.003	<0.000
LYS-307	7.682	>12.000	LYS-1014	9.453	11.036	CYS-671	9.081	10.646
ASP-308	2.913	1.506	ARG-1015	12.068	>12.000	HIS-673	4.876	5.868
LYS-315	9.571	>12.000	LEU-1017	3.812	2.839	GLU-677	4.172	3.721
LYS-318	10.523	>12.000	**Chain-B**			ARG-678	11.497	>12.000
CYS-322	8.439	8.884	NTALA-2	7.303	7.567	CYS-684	10.768	>12.000
CYS-324	8.474	9.151	LYS-7	10.092	11.514	ASP-687	3.422	1.126
HIS-325	6.353	6.228	ASP-8	4.485	0.829	GLU-689	3.988	3.859
ARG-332	12.495	>12.000	ASP-11	3.481	3.217	ARG-692	11.261	>12.000
ASP-342	2.251	<0.000	GLU-13	4.381	4.529	CYS-695	11.885	>12.000
ASP-346	3.570	<0.000	ARG-21	11.916	>12.000	ARG-696	11.809	>12.000
CYS-347	9.395	>12.000	HIS-25	6.433	7.127	ARG-699	13.391	>12.000
ARG-353	12.182	>12.000	CYS-29	9.514	10.390	LYS-709	6.259	>12.000
CYS-354	12.357	>12.000	LYS-34	9.319	>12.000	ASP-716	3.023	<0.000
ARG-357	11.291	>12.000	LYS-35	8.741	10.536	ARG-722	11.522	>12.000
ARG-358	10.450	>12.000	ASP-37	4.901	0.807	LYS-725	10.246	10.437
ARG-364	11.344	>12.000	LYS-38	8.850	>12.000	GLU-726	5.090	2.468
LYS-365	9.444	8.767	LYS-39	10.086	11.352	LYS-745	9.406	>12.000
ARG-371	11.444	>12.000	HIS-40	6.881	7.522	ASP-747	4.039	<0.000
GLU-375	3.182	<0.000	LYS-42	9.927	>12.000	LYS-758	10.182	11.247
GLU-376	4.609	7.909	ARG-43	13.227	>12.000	ARG-759	12.513	>12.000
GLU-378	4.472	<0.000	ASP-46	3.757	<0.000	TYR-762	13.151	>12.000
LYS-381	8.961	>12.000	LYS-47	9.620	>12.000	ARG-771	9.652	>12.000
GLU-382	5.890	<0.000	CYS-49	9.602	9.668	ARG-776	10.476	>12.000
GLU-383	7.131	2.331	CYS-52	9.631	10.695	ARG-783	12.031	>12.000
LYS-384	10.273	>12.000	GLU-53	5.615	1.985	ASP-797	4.369	<0.000
CYS-385	11.369	>12.000	LYS-54	10.179	>12.000	GLU-800	3.723	<0.000

GLU-386	5.498	<0.000	GLU-56	6.027	0.376	HIS-807	4.995	2.920
ARG-394	12.049	>12.000	ASP-60	5.580	<0.000	CYS-816	7.583	7.301
LYS-395	10.185	10.533	ASP-61	4.931	<0.000	ASP-823	2.425	<0.000
LYS-399	10.127	10.237	LYS-63	9.065	>12.000	GLU-828	3.787	5.495
ARG-402	12.348	>12.000	HIS-64	5.372	8.066	ASP-829	1.085	<0.000
ARG-410	12.126	>12.000	GLU-69	3.049	<0.000	TYR-830	10.028	>12.000
GLU-412	5.336	3.049	ARG-70	8.610	>12.000	CYS-831	9.409	>12.000
ASP-414	3.957	2.361	ARG-74	10.614	>12.000	LYS-835	9.531	>12.000
GLU-415	4.611	3.124	GLU-75	4.176	<0.000	TYR-839	12.083	>12.000
LYS-418	10.487	11.457	ARG-78	11.185	>12.000	LYS-841	9.345	>12.000
GLU-421	5.204	2.049	CYS-79	7.452	10.201	GLU-844	4.858	3.424
ASP-422	5.009	6.599	LYS-81	9.384	>12.000	GLU-845	4.934	4.976
GLU-423	4.637	2.362	CYS-82	6.829	>12.000	ASP-848	4.196	3.298
ASP-424	3.906	3.001	ASP-84	5.188	<0.000	ASP-850	3.373	<0.000
HID-428	6.190	5.720	CYS-87	6.049	>12.000	HIS-859	4.635	5.008
LYS-430	10.049	11.656	LYS-89	10.015	>12.000	LYS-861	10.211	>12.000
ASP-432	4.726	2.443	CYS-91	6.684	9.325	LYS-863	9.993	>12.000
ASP-444	3.025	1.501	HIS-94	5.457	<0.000	ARG-867	11.714	>12.000
LYS-446	9.783	10.975	ASP-96	1.971	<0.000	GLU-870	5.014	3.574
LYS-448	10.003	11.215	LYS-98	6.142	>12.000	ASP-873	3.845	3.093
GLU-449	4.446	4.084	LYS-107	8.064	>12.000	LYS-874	10.169	11.356
LYS-455	10.237	11.121	TYR-109	11.951	>12.000	LYS-875	10.049	>12.000
ARG-458	11.879	>12.000	TYR-110	12.459	>12.000	TYR-882	11.538	>12.000
GLU-463	4.395	4.226	LYS-114	8.918	>12.000	GLU-884	4.766	2.427
ASP-465	2.478	1.902	ASP-119	5.577	<0.000	ARG-886	11.545	>12.000
GLU-467	3.620	3.475	CYS-126	7.598	>12.000	LYS-887	10.081	>12.000
GLU-473	4.169	3.252	CYS-130	6.616	>12.000	LYS-888	10.359	11.369
ASP-481	1.697	<0.000	ASP-134	4.849	<0.000	GLU-892	5.120	3.584
GLU-491	1.527	<0.000	CYS-136	5.948	>12.000	LYS-894	9.588	>12.000
ASP-495	1.865	<0.000	CYS-140	6.569	6.413	ARG-896	12.028	>12.000
LYS-497	8.910	>12.000	TYR-143	11.062	>12.000	LYS-898	10.313	>12.000

TYR-502	10.721	>12.000	GLU-146	4.899	<0.000	GLU-899	4.713	3.870
HIS-504	5.529	5.057	GLU-147	5.091	<0.000	GLU-908	4.451	4.161
LYS-505	10.020	10.886	GLU-161	3.736	<0.000	ARG-909	12.053	>12.000
TYR-506	11.472	>12.000	LYS-164	8.930	>12.000	ASP-910	4.599	4.392
TYR-511	11.391	>12.000	ARG-172	12.886	>12.000	CYS-911	11.082	>12.000
LYS-518	10.405	>12.000	GLU-180	4.340	3.833	LYS-915	10.368	>12.000
GLU-520	4.694	2.460	LYS-181	10.263	10.540	ARG-916	11.973	>12.000
TYR-525	11.171	>12.000	GLU-184	4.454	3.989	LYS-922	10.215	11.315
ASP-529	4.550	<0.000	TYR-186	10.640	11.045	ASP-923	3.488	1.495
ASP-532	3.367	2.155	LYS-189	10.554	11.622	LYS-927	10.526	10.847
GLU-536	4.520	4.070	CYS-202	11.328	>12.000	TYR-931	11.357	>12.000
LYS-541	10.375	11.439	ARG-208	12.173	>12.000	GLU-937	5.376	3.135
ARG-561	11.682	>12.000	TYR-211	14.406	>12.000	GLU-942	4.193	3.579
ARG-562	9.172	>12.000	ASP-213	5.427	2.277	ASP-949	3.081	0.434
GLU-565	4.588	3.426	GLU-218	2.676	<0.000	GLU-950	4.768	5.720
LYS-574	7.621	>12.000	LYS-219	8.580	>12.000	GLU-951	5.644	0.188
ASP-579	5.227	9.491	GLU-221	4.261	3.937	CYS-953	7.489	>12.000
LYS-580	9.587	>12.000	TYR-222	9.707	>12.000	CYS-956	7.052	<0.000
ASP-581	4.074	3.498	GLU-230	4.551	0.334	LYS-958	7.356	>12.000
ARG-589	8.773	>12.000	ARG-235	9.374	>12.000	CYS-959	7.047	<0.000
ARG-592	12.279	>12.000	TYR-238	11.339	>12.000	TYR-960	12.823	>12.000
TYR-600	10.478	>12.000	ASP-239	3.611	<0.000	CYS-963	5.987	>12.000
GLU-611	4.902	<0.000	GLU-244	3.536	<0.000	ASP-965	2.615	<0.000
GLU-615	4.875	<0.000	GLU-246	5.141	4.394	TYR-968	11.129	>12.000
LYS-616	8.724	>12.000	LYS-249	9.644	>12.000	ASP-974	3.421	1.243
TYR-620	11.923	>12.000	ASP-250	5.261	4.174	GLU-976	4.826	3.810
CYS-622	11.465	>12.000	LYS-254	10.513	>12.000	HIS-978	5.660	10.280
GLU-627	3.900	1.193	CYS-257	9.804	>12.000	ASP-984	4.239	2.667
LYS-629	10.761	>12.000	LYS-259	10.095	10.949	CYS-986	7.143	9.614
ASP-631	4.956	2.740	GLU-265	4.239	3.526	CYS-989	7.164	6.202
ASP-634	4.364	3.885	LYS-272	10.041	11.198	CYS-992	6.714	>12.000

CYS-643	9.773	>12.000	GLU-273	4.637	3.923	CYS-996	7.589	9.756
TYR-645	10.854	11.797	LYS-274	9.928	9.807	ASP-1000	3.752	<0.000
LYS-647	9.652	9.712	TYR-276	11.844	>12.000	CYS-1001	8.191	>12.000
ASP-649	4.764	2.051	LYS-277	10.656	11.399	LYS-1002	4.736	>12.000
GLU-652	4.299	4.117	GLU-287	3.373	3.024	ARG-1003	10.742	>12.000
LYS-655	9.934	10.986	LYS-290	9.694	10.778	ARG-1007	11.792	>12.000
LYS-656	9.203	11.476	ASP-291	3.366	1.735	TYR-1011	12.215	>12.000
GLU-658	4.790	4.743	ASP-300	3.895	3.517	GLU-1012	4.463	3.777
ASP-659	4.923	4.970	TYR-304	11.644	>12.000	LYS-1014	9.489	10.840
ASP-663	5.380	1.857	LYS-307	7.604	>12.000	ARG-1015	11.964	>12.000
GLU-666	3.471	<0.000	ASP-308	2.671	1.112	CTala-2035	4.036	2.880
CYS-671	10.614	>12.000	LYS-315	9.386	>12.000			
HIS-673	5.824	6.323	LYS-318	11.015	>12.000			
GLU-677	4.057	2.963	CYS-322	9.919	11.243			

Table S2.3. Summary statistics of human DPD parameters and coordinate files for Model 1 (B3LYP/6-31G*): AMBER parameter file. Table adapted from Tendwa et al., (2021)

##Model 1 (B3LYP/6-31G*) AMBER parameter file

REMARK GOES HERE, THIS FILE IS GENERATED BY *MCPB.PY*
MASS

FE1	55.85		Fe ion
FE2	55.85		Fe ion
FE3	55.85		Fe ion
FE4	55.85		Fe ion
FE5	55.85		Fe ion
FE6	55.85		Fe ion
FE7	55.85		Fe ion
FE8	55.85		Fe ion
CysSG(78)	32.06	2.900	S in cystine
CysSG(81)	32.06	2.900	S in cystine
CysSG(86)	32.06	2.900	S in cystine
CysSG(90)	32.06	2.900	S in cystine
CysSG(129)	32.06	2.900	S in cystine
CysSG(135)	32.06	2.900	S in cystine
CysSG(139)	32.06	2.900	S in cystine
GnOE(155)	16.00	0.434	carbonyl group oxygen
S1	32.060	2.900	S ion
S2	32.060	2.900	S ion
S3	32.060	2.900	S ion
S4	32.060	2.900	S ion
S5	32.060	2.900	S ion
S6	32.060	2.900	S ion
S7	32.060	2.900	S ion
S8	32.060	2.900	S ion

BOND	(kcal/mol/Å)	(Å)	
CysSG(90)-FE1	44.6	2.3851	Created by Seminario method using MCPB.py
S2-FE1	56.1	2.2759	Created by Seminario method using MCPB.py
S2-FE3	57.3	2.3031	Created by Seminario method using MCPB.py
S2-FE4	54.0	2.2876	Created by Seminario method using MCPB.py
S3-FE1	55.1	2.2007	Created by Seminario method using MCPB.py
S3-FE2	66.5	2.1837	Created by Seminario method using MCPB.py
S3-FE4	55.6	2.1893	Created by Seminario method using MCPB.py
S4-FE1	66.1	2.2277	Created by Seminario method using MCPB.py
S4-FE2	62.0	2.2510	Created by Seminario method using MCPB.py
S4-FE3	76.3	2.2375	Created by Seminario method using MCPB.py
CysSG(135)-FE2	53.4	2.3262	Created by Seminario method using MCPB.py
S1-FE2	54.7	2.2552	Created by Seminario method using MCPB.py
S1-FE3	49.3	2.2649	Created by Seminario method using MCPB.py
S1-FE4	50.5	2.2541	Created by Seminario method using MCPB.py
GnOE(155)-FE3	60.4	1.9176	Created by Seminario method using MCPB.py
CysSG(129)-FE4	36.4	2.4150	Created by Seminario method using MCPB.py
CysSG(139)-FE5	36.4	2.3965	Created by Seminario method using MCPB.py
S6-FE5	45.1	2.2410	Created by Seminario method using MCPB.py
S6-FE7	45.3	2.2325	Created by Seminario method using MCPB.py
S6-FE8	41.5	2.2698	Created by Seminario method using MCPB.py
S7-FE5	64.3	2.2502	Created by Seminario method using MCPB.py
S7-FE6	69.0	2.2387	Created by Seminario method using MCPB.py
S7-FE8	61.8	2.2645	Created by Seminario method using MCPB.py
S8-FE5	58.3	2.2124	Created by Seminario method using MCPB.py
S8-FE6	51.2	2.2250	Created by Seminario method using MCPB.py
S8-FE7	64.7	2.1906	Created by Seminario method using MCPB.py
CysSG(86)-FE6	50.0	2.3515	Created by Seminario method using MCPB.py
S5-FE6	66.3	2.2363	Created by Seminario method using MCPB.py
S5-FE7	60.8	2.2371	Created by Seminario method using MCPB.py
S5-FE8	57.0	2.2642	Created by Seminario method using MCPB.py

CysSG(78)-FE7	40.2	2.3710	Created by Seminario method using MCPB.py
CysSG(81)-FE8	37.1	2.4103	Created by Seminario method using MCPB.py
C -GnOE(155)	570.0	1.229	JCC,7,(1986),230; AA,CYT,GUA,THY,URA
CT-CysSG(90)	237.0	1.810	changed from 222.0 based on methanethiol nmodes
CT-CysSG(135)	237.0	1.810	changed from 222.0 based on methanethiol nmodes
CT-CysSG(129)	237.0	1.810	changed from 222.0 based on methanethiol nmodes
CT-CysSG(139)	237.0	1.810	changed from 222.0 based on methanethiol nmodes
CT-CysSG(86)	237.0	1.810	changed from 222.0 based on methanethiol nmodes
CT-CysSG(78)	237.0	1.810	changed from 222.0 based on methanethiol nmodes
CT-CysSG(81)	237.0	1.810	changed from 222.0 based on methanethiol nmodes

ANGLE	(kcal mol/rad)		(degrees)
C -GnOE(155)-FE3	75.86	130.30	Created by Seminario method using MCPB.py
CT-CysSG(90)-FE1	106.68	103.98	Created by Seminario method using MCPB.py
CT-CysSG(135)-FE2	100.90	107.39	Created by Seminario method using MCPB.py
CT-CysSG(129)-FE4	97.96	96.66	Created by Seminario method using MCPB.py
CT-CysSG(139)-FE5	94.60	105.21	Created by Seminario method using MCPB.py
CT-CysSG(86)-FE6	101.58	107.61	Created by Seminario method using MCPB.py
CT-CysSG(78)-FE7	100.02	108.65	Created by Seminario method using MCPB.py
CT-CysSG(81)-FE8	106.65	103.60	Created by Seminario method using MCPB.py
FE2-S3-FE1	75.11	68.39	Created by Seminario method using MCPB.py
FE2-S4-FE1	80.79	66.76	Created by Seminario method using MCPB.py
FE3-S2-FE1	23.20	69.98	Created by Seminario method using MCPB.py
FE3-S4-FE1	26.76	72.04	Created by Seminario method using MCPB.py
FE3-S4-FE2	40.49	66.60	Created by Seminario method using MCPB.py
FE3-S1-FE2	47.73	66.08	Created by Seminario method using MCPB.py
FE4-S2-FE1	53.67	65.65	Created by Seminario method using MCPB.py
FE4-S2-FE3	20.95	64.15	Created by Seminario method using MCPB.py
FE4-S3-FE1	29.62	68.59	Created by Seminario method using MCPB.py
FE4-S3-FE2	61.07	68.33	Created by Seminario method using MCPB.py
FE4-S1-FE2	78.37	66.00	Created by Seminario method using MCPB.py
FE4-S1-FE3	53.87	65.30	Created by Seminario method using MCPB.py
FE6-S7-FE5	63.39	66.89	Created by Seminario method using MCPB.py
FE6-S8-FE5	77.31	67.77	Created by Seminario method using MCPB.py
FE7-S6-FE5	42.52	67.16	Created by Seminario method using MCPB.py
FE7-S8-FE5	64.00	68.38	Created by Seminario method using MCPB.py
FE7-S8-FE6	42.93	68.59	Created by Seminario method using MCPB.py
FE7-S5-FE6	35.00	67.59	Created by Seminario method using MCPB.py
FE8-S6-FE5	61.00	67.08	Created by Seminario method using MCPB.py
FE8-S6-FE7	44.11	66.60	Created by Seminario method using MCPB.py
FE8-S7-FE5	46.02	67.02	Created by Seminario method using MCPB.py
FE8-S7-FE6	47.60	70.85	Created by Seminario method using MCPB.py
FE8-S5-FE6	36.63	70.89	Created by Seminario method using MCPB.py
FE8-S5-FE7	67.61	66.62	Created by Seminario method using MCPB.py
CysSG(90)-FE1-S2	41.30	111.43	Created by Seminario method using MCPB.py
CysSG(90)-FE1-S3	63.82	115.56	Created by Seminario method using MCPB.py
CysSG(90)-FE1-S4	65.80	106.49	Created by Seminario method using MCPB.py
S2-FE1-S3	43.74	109.45	Created by Seminario method using MCPB.py
S2-FE1-S4	28.56	103.16	Created by Seminario method using MCPB.py
S2-FE3-S4	28.52	101.98	Created by Seminario method using MCPB.py
S2-FE4-S3	60.52	109.42	Created by Seminario method using MCPB.py
S3-FE1-S4	35.95	109.98	Created by Seminario method using MCPB.py
S3-FE2-S4	69.35	109.74	Created by Seminario method using MCPB.py
CysSG(135)-FE2-S3	52.84	115.18	Created by Seminario method using MCPB.py
CysSG(135)-FE2-S4	75.62	109.23	Created by Seminario method using MCPB.py
CysSG(135)-FE2-S1	47.43	105.58	Created by Seminario method using MCPB.py
S1-FE2-S3	62.28	106.66	Created by Seminario method using MCPB.py
S1-FE2-S4	34.47	110.31	Created by Seminario method using MCPB.py
S1-FE3-S2	16.26	111.70	Created by Seminario method using MCPB.py
S1-FE3-S4	22.23	110.45	Created by Seminario method using MCPB.py
S1-FE4-S2	28.28	112.69	Created by Seminario method using MCPB.py
S1-FE4-S3	59.25	106.50	Created by Seminario method using MCPB.py
GnOE(155)-FE3-S2	34.09	115.71	Created by Seminario method using MCPB.py
GnOE(155)-FE3-S4	59.86	109.31	Created by Seminario method using MCPB.py
GnOE(155)-FE3-S1	51.72	107.58	Created by Seminario method using MCPB.py
CysSG(129)-FE4-S2	53.37	112.42	Created by Seminario method using MCPB.py
CysSG(129)-FE4-S3	46.08	111.58	Created by Seminario method using MCPB.py

CysSG(129)-FE4-S1	64.85	104.01	Created by Seminario method using MCPB.py
CysSG(139)-FE5-S6	53.22	109.82	Created by Seminario method using MCPB.py
CysSG(139)-FE5-S7	45.83	111.05	Created by Seminario method using MCPB.py
CysSG(139)-FE5-S8	51.92	109.60	Created by Seminario method using MCPB.py
S6-FE5-S7	28.91	110.49	Created by Seminario method using MCPB.py
S6-FE5-S8	36.64	105.86	Created by Seminario method using MCPB.py
S6-FE7-S8	42.99	106.90	Created by Seminario method using MCPB.py
S6-FE8-S7	20.07	108.94	Created by Seminario method using MCPB.py
S7-FE5-S8	59.08	109.89	Created by Seminario method using MCPB.py
S7-FE6-S8	37.10	109.86	Created by Seminario method using MCPB.py
CysSG(86)-FE6-S7	45.25	109.29	Created by Seminario method using MCPB.py
CysSG(86)-FE6-S8	63.88	116.82	Created by Seminario method using MCPB.py
CysSG(86)-FE6-S5	53.36	107.61	Created by Seminario method using MCPB.py
S5-FE6-S7	29.55	104.82	Created by Seminario method using MCPB.py
S5-FE6-S8	44.49	107.73	Created by Seminario method using MCPB.py
S5-FE7-S6	29.29	111.70	Created by Seminario method using MCPB.py
S5-FE7-S8	76.82	108.93	Created by Seminario method using MCPB.py
S5-FE8-S6	42.92	109.34	Created by Seminario method using MCPB.py
S5-FE8-S7	16.77	103.08	Created by Seminario method using MCPB.py
CysSG(78)-FE7-S6	62.02	113.79	Created by Seminario method using MCPB.py
CysSG(78)-FE7-S8	55.31	113.14	Created by Seminario method using MCPB.py
CysSG(78)-FE7-S5	62.04	102.35	Created by Seminario method using MCPB.py
CysSG(81)-FE8-S6	47.02	116.54	Created by Seminario method using MCPB.py
CysSG(81)-FE8-S7	51.70	110.21	Created by Seminario method using MCPB.py
CysSG(81)-FE8-S5	53.95	107.88	Created by Seminario method using MCPB.py
2C-C -GnOE(155)	80.0	120.40	
CX-CT-CysSG(90)	50.0	108.60	AA cys (was CT-CT-SH)
CX-CT-CysSG(135)	50.0	108.60	AA cys (was CT-CT-SH)
CX-CT-CysSG(129)	50.0	108.60	AA cys (was CT-CT-SH)
CX-CT-CysSG(139)	50.0	108.60	AA cys (was CT-CT-SH)
CX-CT-CysSG(86)	50.0	108.60	AA cys (was CT-CT-SH)
CX-CT-CysSG(78)	50.0	108.60	AA cys (was CT-CT-SH)
CX-CT-CysSG(81)	50.0	108.60	AA cys (was CT-CT-SH)
N -C -GnOE(155)	80.0	122.90	AA general
CysSG(90)-CT-H1	50.0	109.50	AA cyx changed based on NMA nmodes
CysSG(135)-CT-H1	50.0	109.50	AA cyx changed based on NMA nmodes
CysSG(129)-CT-H1	50.0	109.50	AA cyx changed based on NMA nmodes
CysSG(139)-CT-H1	50.0	109.50	AA cyx changed based on NMA nmodes
CysSG(86)-CT-H1	50.0	109.50	AA cyx changed based on NMA nmodes
CysSG(78)-CT-H1	50.0	109.50	AA cyx changed based on NMA nmodes
CysSG(81)-CT-H1	50.0	109.50	AA cyx changed based on NMA nmodes

DIHE

2C-2C-C -GnOE(155)	1	0.0	0.0	1.0	
2C-C -GnOE(155)-FE3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
C -GnOE(155)-FE3-S2	3	0.00	0.00	3.0	Treat as zero by MCPB.py
C -GnOE(155)-FE3-S4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
C -GnOE(155)-FE3-S1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CT-CysSG(90)-FE1-S2	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CT-CysSG(90)-FE1-S3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CT-CysSG(90)-FE1-S4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CT-CysSG(135)-FE2-S3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CT-CysSG(135)-FE2-S4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CT-CysSG(135)-FE2-S1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CT-CysSG(129)-FE4-S2	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CT-CysSG(129)-FE4-S3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CT-CysSG(129)-FE4-S1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CT-CysSG(139)-FE5-S6	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CT-CysSG(139)-FE5-S7	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CT-CysSG(139)-FE5-S8	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CT-CysSG(86)-FE6-S7	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CT-CysSG(86)-FE6-S8	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CT-CysSG(86)-FE6-S5	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CT-CysSG(78)-FE7-S6	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CT-CysSG(78)-FE7-S8	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CT-CysSG(78)-FE7-S5	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CT-CysSG(81)-FE8-S6	3	0.00	0.00	3.0	Treat as zero by MCPB.py

CT-CysSG(81)-FE8-S7	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CT-CysSG(81)-FE8-S5	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CX-CT-CysSG(90)-FE1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CX-CT-CysSG(135)-FE2	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CX-CT-CysSG(129)-FE4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CX-CT-CysSG(139)-FE5	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CX-CT-CysSG(86)-FE6	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CX-CT-CysSG(78)-FE7	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CX-CT-CysSG(81)-FE8	3	0.00	0.00	3.0	Treat as zero by MCPB.py
H -N -C -GnOE(155)	1	2.5	180.0	-2.0	JCC,7,(1986),230
H -N -C -GnOE(155)	1	2.0	0.0	1.0	J.C.cistrans-NMA DE
FE1-CysSG(90)-CT-H1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE2-S3-FE1-S4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE2-CysSG(135)-CT-H1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE3-S2-FE1-S3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE3-S2-FE1-S4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE3-S1-FE2-S3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE3-S1-FE2-S4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE4-S2-FE1-S3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE4-S2-FE1-S4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE4-S2-FE3-S4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE4-S3-FE1-S4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE4-S3-FE2-S4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE4-S1-FE2-S3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE4-S1-FE2-S4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE4-S1-FE3-S2	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE4-S1-FE3-S4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE4-CysSG(129)-CT-H1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE5-CysSG(139)-CT-H1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE6-S7-FE5-S8	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE6-CysSG(86)-CT-H1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE7-S6-FE5-S7	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE7-S6-FE5-S8	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE7-S5-FE6-S7	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE7-S5-FE6-S8	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE7-CysSG(78)-CT-H1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE8-S6-FE5-S7	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE8-S6-FE5-S8	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE8-S6-FE7-S8	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE8-S7-FE5-S8	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE8-S7-FE6-S8	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE8-S5-FE6-S7	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE8-S5-FE6-S8	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE8-S5-FE7-S6	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE8-S5-FE7-S8	3	0.00	0.00	3.0	Treat as zero by MCPB.py
FE8-CysSG(81)-CT-H1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
N -C -GnOE(155)-FE3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CysSG(90)-FE1-S2-FE3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CysSG(90)-FE1-S2-FE4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CysSG(90)-FE1-S3-FE2	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CysSG(90)-FE1-S3-FE4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CysSG(90)-FE1-S4-FE2	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CysSG(90)-FE1-S4-FE3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S2-FE1-S3-FE2	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S2-FE1-S3-FE4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S2-FE1-S4-FE2	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S2-FE1-S4-FE3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S2-FE3-S4-FE1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S2-FE3-S4-FE2	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S2-FE3-S1-FE2	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S2-FE4-S3-FE1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S2-FE4-S3-FE2	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S2-FE4-S1-FE2	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S2-FE4-S1-FE3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S3-FE1-S4-FE2	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S3-FE1-S4-FE3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S3-FE2-S4-FE1	3	0.00	0.00	3.0	Treat as zero by MCPB.py

S3-FE2-S4-FE3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S3-FE4-S2-FE1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S3-FE4-S2-FE3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S3-FE4-S1-FE2	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S3-FE4-S1-FE3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S4-FE2-S3-FE1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S4-FE3-S2-FE1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S4-FE3-S1-FE2	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CysSG(135)-FE2-S3-FE1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CysSG(135)-FE2-S3-FE4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CysSG(135)-FE2-S4-FE1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CysSG(135)-FE2-S4-FE3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CysSG(135)-FE2-S1-FE3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CysSG(135)-FE2-S1-FE4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S1-FE2-S3-FE1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S1-FE2-S3-FE4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S1-FE2-S4-FE1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S1-FE2-S4-FE3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S1-FE3-S2-FE1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S1-FE3-S2-FE4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S1-FE3-S4-FE1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S1-FE3-S4-FE2	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S1-FE4-S2-FE1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S1-FE4-S2-FE3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S1-FE4-S3-FE1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S1-FE4-S3-FE2	3	0.00	0.00	3.0	Treat as zero by MCPB.py
GnOE(155)-C -2C-HC	1	0.08	180.0	-3.0	Junmei et al, 1999 (HC-CT-C -O)
GnOE(155)-C -2C-HC	1	0.0	0.0	-2.0	
GnOE(155)-C -2C-HC	1	0.8	0.0	1.0	
GnOE(155)-FE3-S2-FE1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
GnOE(155)-FE3-S2-FE4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
GnOE(155)-FE3-S4-FE1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
GnOE(155)-FE3-S4-FE2	3	0.00	0.00	3.0	Treat as zero by MCPB.py
GnOE(155)-FE3-S1-FE2	3	0.00	0.00	3.0	Treat as zero by MCPB.py
GnOE(155)-FE3-S1-FE4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CysSG(129)-FE4-S2-FE1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CysSG(129)-FE4-S2-FE3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CysSG(129)-FE4-S3-FE1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CysSG(129)-FE4-S3-FE2	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CysSG(129)-FE4-S1-FE2	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CysSG(129)-FE4-S1-FE3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CysSG(139)-FE5-S6-FE7	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CysSG(139)-FE5-S6-FE8	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CysSG(139)-FE5-S7-FE6	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CysSG(139)-FE5-S7-FE8	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CysSG(139)-FE5-S8-FE6	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CysSG(139)-FE5-S8-FE7	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S6-FE5-S7-FE6	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S6-FE5-S7-FE8	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S6-FE5-S8-FE6	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S6-FE5-S8-FE7	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S6-FE7-S8-FE5	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S6-FE7-S8-FE6	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S6-FE7-S5-FE6	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S6-FE8-S7-FE5	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S6-FE8-S7-FE6	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S6-FE8-S5-FE6	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S6-FE8-S5-FE7	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S7-FE5-S8-FE6	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S7-FE5-S8-FE7	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S7-FE6-S8-FE5	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S7-FE6-S8-FE7	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S7-FE8-S6-FE5	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S7-FE8-S6-FE7	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S7-FE8-S5-FE6	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S7-FE8-S5-FE7	3	0.00	0.00	3.0	Treat as zero by MCPB.py
S8-FE6-S7-FE5	3	0.00	0.00	3.0	Treat as zero by MCPB.py

```
S8-FE7-S6-FE5              3     0.00     0.00  3.0   Treat as zero by MCPB.py
S8-FE7-S5-FE6              3     0.00     0.00  3.0   Treat as zero by MCPB.py
CysSG(86)-FE6-S7-FE5       3     0.00     0.00  3.0   Treat as zero by MCPB.py
CysSG(86)-FE6-S7-FE8       3     0.00     0.00  3.0   Treat as zero by MCPB.py
CysSG(86)-FE6-S8-FE5       3     0.00     0.00  3.0   Treat as zero by MCPB.py
CysSG(86)-FE6-S8-FE7       3     0.00     0.00  3.0   Treat as zero by MCPB.py
CysSG(86)-FE6-S5-FE7       3     0.00     0.00  3.0   Treat as zero by MCPB.py
CysSG(86)-FE6-S5-FE8       3     0.00     0.00  3.0   Treat as zero by MCPB.py
S5-FE6-S7-FE5              3     0.00     0.00  3.0   Treat as zero by MCPB.py
S5-FE6-S7-FE8              3     0.00     0.00  3.0   Treat as zero by MCPB.py
S5-FE6-S8-FE5              3     0.00     0.00  3.0   Treat as zero by MCPB.py
S5-FE6-S8-FE7              3     0.00     0.00  3.0   Treat as zero by MCPB.py
S5-FE7-S6-FE5              3     0.00     0.00  3.0   Treat as zero by MCPB.py
S5-FE7-S6-FE8              3     0.00     0.00  3.0   Treat as zero by MCPB.py
S5-FE7-S8-FE5              3     0.00     0.00  3.0   Treat as zero by MCPB.py
S5-FE7-S8-FE6              3     0.00     0.00  3.0   Treat as zero by MCPB.py
S5-FE8-S6-FE5              3     0.00     0.00  3.0   Treat as zero by MCPB.py
S5-FE8-S6-FE7              3     0.00     0.00  3.0   Treat as zero by MCPB.py
S5-FE8-S7-FE5              3     0.00     0.00  3.0   Treat as zero by MCPB.py
S5-FE8-S7-FE6              3     0.00     0.00  3.0   Treat as zero by MCPB.py
CysSG(78)-FE7-S6-FE5       3     0.00     0.00  3.0   Treat as zero by MCPB.py
CysSG(78)-FE7-S6-FE8       3     0.00     0.00  3.0   Treat as zero by MCPB.py
CysSG(78)-FE7-S8-FE5       3     0.00     0.00  3.0   Treat as zero by MCPB.py
CysSG(78)-FE7-S8-FE6       3     0.00     0.00  3.0   Treat as zero by MCPB.py
CysSG(78)-FE7-S5-FE6       3     0.00     0.00  3.0   Treat as zero by MCPB.py
CysSG(78)-FE7-S5-FE8       3     0.00     0.00  3.0   Treat as zero by MCPB.py
CysSG(81)-FE8-S6-FE5       3     0.00     0.00  3.0   Treat as zero by MCPB.py
CysSG(81)-FE8-S6-FE7       3     0.00     0.00  3.0   Treat as zero by MCPB.py
CysSG(81)-FE8-S7-FE5       3     0.00     0.00  3.0   Treat as zero by MCPB.py
CysSG(81)-FE8-S7-FE6       3     0.00     0.00  3.0   Treat as zero by MCPB.py
CysSG(81)-FE8-S5-FE6       3     0.00     0.00  3.0   Treat as zero by MCPB.py
CysSG(81)-FE8-S5-FE7       3     0.00     0.00  3.0   Treat as zero by MCPB.py

IMPR
X -X -C -GnOE(155)         10.5   180.        2.      JCC,7,(1986),230

NONB
  FE1              1.4090  0.0172100000    IOD set for Fe2+ ion from Li et al. JCTC, 2013, 9, 2733
  FE2              1.4090  0.0172100000    IOD set for Fe2+ ion from Li et al. JCTC, 2013, 9, 2733
  FE3              1.4090  0.0172100000    IOD set for Fe2+ ion from Li et al. JCTC, 2013, 9, 2733
  FE4              1.4090  0.0172100000    IOD set for Fe2+ ion from Li et al. JCTC, 2013, 9, 2733
  FE5              1.4090  0.0172100000    IOD set for Fe2+ ion from Li et al. JCTC, 2013, 9, 2733
  FE6              1.4090  0.0172100000    IOD set for Fe2+ ion from Li et al. JCTC, 2013, 9, 2733
  FE7              1.4090  0.0172100000    IOD set for Fe2+ ion from Li et al. JCTC, 2013, 9, 2733
  FE8              1.4090  0.0172100000    IOD set for Fe2+ ion from Li et al. JCTC, 2013, 9, 2733
  CysSG(78)        2.0000  0.2500          W. Cornell CH3SH and CH3SCH3 FEP's
  CysSG(81)        2.0000  0.2500          W. Cornell CH3SH and CH3SCH3 FEP's
  CysSG(86)        2.0000  0.2500          W. Cornell CH3SH and CH3SCH3 FEP's
  CysSG(90)        2.0000  0.2500          W. Cornell CH3SH and CH3SCH3 FEP's
  CysSG(129)       2.0000  0.2500          W. Cornell CH3SH and CH3SCH3 FEP's
  CysSG(135)       2.0000  0.2500          W. Cornell CH3SH and CH3SCH3 FEP's
  CysSG(139)       2.0000  0.2500          W. Cornell CH3SH and CH3SCH3 FEP's
  GnOE(155)        1.6612  0.2100          OPLS
  S1               2.0000  0.2500          S ion
  S2               2.0000  0.2500          S ion
  S3               2.0000  0.2500          S ion
  S4               2.0000  0.2500          S ion
  S5               2.0000  0.2500          S ion
  S6               2.0000  0.2500          S ion
  S7               2.0000  0.2500          S ion
  S8               2.0000  0.2500          S ion
```

Table S2.4. Summary statistics of human DPD parameters and coordinate files for Model 2 (LSDA/LANL2DZ): AMBER_VFFDT parameter file. Table adapted from Tendwa *et al.* (2021)

#This file is generated by the VFFDT program for AMBER.

```
MASS
S3              32.06500  2.90000         ; S
F3              55.84500  2.05000         ; Fe


BOND            (kcal/mol/Å)      (Å)
S3-F3           89.23755          2.22545  ; Averaged bond related parameters from VFFDT
F3-F3           59.53185          2.43332  ; Averaged bond related parameters from VFFDT
SG-F3            39.7711          2.3294   ; Averaged manually with SD(10.7658) and SD( 1.0804)
O-F3            24.9728           1.9340   ; Averaged manually with SD( 3.5472) and SD( 0.9834)

ANGL            (kcal mol/rad)    (degrees)
S3-F3-S3         39.52095         109.20   ; Averaged angle related parameters from VFFDT
S3-F3-F3        43.33559          73.06    ; Averaged angle related parameters from VFFDT
F3-S3-F3        26.86171          66.28    ; Averaged angle related parameters from VFFDT
F3-F3-F3        52.40760          60.00    ; Averaged angle related parameters from VFFDT
SG-F3-F3         38.8905          145.0629 ; Averaged manually with SD(11.4418) and SD( 9.6176)
SG-F3-S3        36.1370           113.2785 ; Averaged manually with SD( 6.7190) and SD( 9.1432)
H-SG-F3         23.5646           104.3280 ; Averaged manually with SD( 4.0419) and SD( 8.1811)
F3-O-C          33.8107           134.0050 ; Averaged manually with SD( 4.1504) and SD( 8.2116)
S3-F3-O         41.2264           115.3200 ; Averaged manually with SD( 5.1145) and SD( 8.6921)
O-F3-F3         56.4637           150.9850 ; Averaged manually with SD( 5.3204) and SD( 8.7809)
F3-SG-C         60.7114           112.8094 ; Averaged manually with SD(15.3253) and SD( 9.2780)
F3-SG-HS        0.0000            0.0000   ; Not calculated
F3-SG-2C        0.0000            0.0000   ; Not calculated
SG-F3-SG        0.0000            0.0000   ; Not calculated
O-F3-SG         0.0000            0.0000   ; Not calculated

DIHE
F3-S3-F3-S3     1  0.00000   0.00  1.00     ; Please check it manually.
S3-F3-S3-F3     1  0.00000   0.00  1.00     ; Please check it manually.
S3-F3-F3-S3     1  0.00000   0.00  1.00     ; Please check it manually.
S3-F3-F3-F3     1  0.00000   0.00  1.00     ; Please check it manually.
F3-S3-F3-F3     1  0.00000   0.00  1.00     ; Please check it manually.
F3-F3-S3-F3     1  0.00000   0.00  1.00     ; Please check it manually.
F3-F3-F3-F3     1  0.00000   0.00  1.00     ; Please check it manually.
F3-F3-F3-SG     1  0.00000   0.00  1.00     ; Please check it manually.
F3-F3-F3-O      1  0.00000   0.00  1.00     ; Please check it manually.
F3-F3-SG-2C     1  0.00000   0.00  1.00     ; Please check it manually.
F3-S3-F3-O      1  0.00000   0.00  1.00     ; Please check it manually.
F3-F3-SG-HS     1  0.00000   0.00  1.00     ; Please check it manually.
F3-S3-F3-SG     1  0.00000   0.00  1.00     ; Please check it manually.
F3-F3-O-C       1  0.00000   0.00  1.00     ; Please check it manually.
S3-F3-SG-HS     1  0.00000   0.00  1.00     ; Please check it manually.
S3-F3-F3-SG     1  0.00000   0.00  1.00     ; Please check it manually.
S3-F3-F3-O      1  0.00000   0.00  1.00     ; Please check it manually.
S3-F3-SG-2C     1  0.00000   0.00  1.00     ; Please check it manually.
S3-F3-O-C       1  0.00000   0.00  1.00     ; Please check it manually.
SG-F3-F3-O      1  0.00000   0.00  1.00     ; Please check it manually.
SG-F3-F3-SG     1  0.00000   0.00  1.00     ; Please check it manually.
HS-SG-F3-SG     1  0.00000   0.00  1.00     ; Please check it manually.
SG-F3-SG-2C     1  0.00000   0.00  1.00     ; Please check it manually.
O-F3-SG-2C      1  0.00000   0.00  1.00     ; Please check it manually.
O-F3-SG-HS      1  0.00000   0.00  1.00     ; Please check it manually.
C-O-F3-SG       1  0.00000   0.00  1.00     ; Please check it manually.
```

IMPR

NONB
F3 1.4090 0.0172100000 IOD set for Fe2+ ion from Li *et al.* JCTC, 2013, 9, 2733
S3 2.0000 0.2500 W. Cornell CH3SH and CH3SCH3 FEP's

Table S2.5. Charge distribution for all atoms interacting with the metal center in **A)** Model 1 and (B3LYP/6-31G*) **B)** Model 2 (LSDA/LANL2DZ). Table adapted from Tendwa et al., (2021)

A) Model 1

Atom	Atomic charge		Atom (S^{2-}	Atomic charge		Atom	Atomic
Fe1	0.489		S1	- 0.802		GLN(156)	- 0.467
Fe2	0.725		S2	- 0.841		CYS(136)	- 0.726
Fe3	0.622		S3	- 0.866		CYS(130)	- 0.570
Fe4	0.614		S4	- 0.822		CYS(91)	- 0.672
Fe5	0.553		S5	- 0.758		CYS(82)	- 0.752
Fe6	0.705		S6	- 1.025		CYS(87)	- 0.823
Fe7	0.667		S7	- 0.799		CYS(140)	- 0.718
Fe8	0.638		S8	- 0.752		CYS(79)	- 0.722
B) Model 2							
Fe1	0.226		S1	- 0.226		GLN	- 0.229
Fe2	0.226		S2	- 0.226		CYS	- 0.334
Fe3	0.226		S3	- 0.226		CYS	- 0.334
Fe4	0.226		S4	- 0.226		CYS	- 0.334

Table S2.6. Assessment of **A** bond length, **B** internal and **C** external angles (Å) calculated with X-ray, DFT (B3LYP) and (LSDA/LANL2DZ) method for the molecular cluster model ([$Fe^{2+}_4S^{2-}_4$ (S-Cys)$_3$(S-Gln)]) 1026A of Native DPD protein. Table adapted from Tendwa et al., (2021)

A

Geometry	**Bond length** *(Å)*						
Model System	**$Fe^{2+}_4S^{2-}_4$(S-Cys)$_3$(O-Gln): 1026A Clusters**						
Bond	**X-ray**	**QM**				**AFTER-MD**	
Bond description	**1H7X**	**B3LYP (Model 1)**		**LSDA/LANL2DZ (Model 2)**		**Model 1**	**Model 2**
	Bond length *(Å)*	**Equilibrium bond length [req]** *(Å)*	**Force constant [Kr]** ***(kcal mol^{-1} Å^{-2})***	**Equilibrium bond length [req]** *(Å)*	**Force constant [Kr]** *(kcal mol^{-1} Å^{-2})*	**Bond length** *(Å)* **Mean & SD**	**Bond length** *(Å)* **Mean & SD**
FE1-S2	2.51	2.28	56.10	2.22	89.23	2.28±0.16	2.22±0.21
FE1-S3	2.42	2.20	55.10	2.22	89.23	2.20±0.17	2.23±0.14
FE1-S4	2.58	2.22	66.10	2.22	89.23	2.24± 0.24	2.23± 0.25
FE1-SG91	2.32	2.39	44.60	2.33	39.77	2.36±0.03	2.35±0.02
FE2-S1	2.61	2.26	54.70	2.22	89.23	2.28±0.23	2.23±0.27
FE2-S3	2.64	2.18	66.50	2.22	89.23	2.21±0.30	2.22±0.30
FE2-S4	2.73	2.25	62.00	2.22	89.23	2.27±0.33	2.22±0.43
FE2-SG136	2.37	2.33	53.40	2.33	39.77	2.36±0.01	2.34±0.02
FE3-S1	2.43	2.26	49.30	2.22	89.23	2.25±0.13	2.23±0.14
FE3-S2	2.40	2.30	57.30	2.22	89.23	2.30±0.07	2.23±0.12
FE3-S4	2.56	2.24	76.30	2.22	89.23	2.23±0.23	2.23±0.23
FE3-OE156	1.89	1.92	60.40	1.93	24.97	1.98±0.06	1.92±0.02
FE4-S1	2.60	2.25	50.50	2.22	89.23	2.24±0.25	2.23±0.26
FE4-S2	2.60	2.29	54.00	2.22	89.23	2.25±0.25	2.24±0.25
FE4-S3	2.37	2.19	55.60	2.22	89.23	2.18±0.13	2.22±0.11
FE4-SG130	2.36	2.40	36.40	2.33	39.77	2.38±0.01	2.34±0.01

[1]DFT: density functional theory, [2]B3LYP: Becke three-parameter hybrid exchange and Lee Yang Parr, [3]LSDA/LANL2DZ**:** Los Alamos double-zeta basis, [4]SD: standard deviation

B

Geometry	Internal Angle *(⁰)*						
Model System	**$Fe^{2+}{}_4S^{2-}{}_4$(S-Cys)$_3$(O-Gln): 1026A Clusters**						
Angle	**X-ray**	**QM**				**AFTER-MD**	
Angle description	**1H7X**	**B3LYP (Model 1)**		**LSDA/LANL2DZ (Model 2)**		**Model 1**	**Model 2**
	Average angle *(⁰)*	**Average equilibrium Angle [Θ_{eq}] *(⁰)***	**Force constant [KΘ] *(kcal mol^{-1}rad^{-2})***	**Average equilibrium Angle [Θ_{eq}] *(⁰)***	**Force constant [KΘ] *(kcal mol^{-1}rad^{-2})***	**Angle (0) Mean & SD**	**Angle (0) Mean & SD**
FE1-S2-FE3	68.12	69.98	23.20	66.28	26.86	58.20±7.01	69.05±0.66
FE1-S2-FE4	65.87	65.65	53.67	66.28	26.86	64.32±0.29	69.61±2.64
FE1-S3-FE2	64.61	68.39	75.11	66.28	26.86	64.33±0.20	68.93±3.05
FE1-S3-FE4	65.86	68.59	29.62	66.28	26.86	67.08±0.86	67.81±1.37
FE1-S4-FE2	68.72	66.76	80.79	66.28	26.86	60.34±5.93	70.09±0.96
FE1-S4-FE3	67.60	72.04	26.76	66.28	26.86	63.80± 2.69	63.97± 2.56
FE2-S1-FE3	68.15	66.08	47.73	66.28	26.86	67.27±0.62	68.49±0.24
FE2-S3-FE4	71.15	68.33	61.07	66.28	26.86	65.35±4.10	69.07±1.47
FE2-S4-FE3	64.60	66.60	40.49	66.28	26.86	61.15±2.43	67.96±2.38
FE3-S1-FE4	71.92	65.300	53.87	66.28	26.86	62.67±6.54	67.05± 3.44
FE3-S2-FE4	67.59	64.15	20.95	66.28	26.86	58.20±6.64	68.10±0.36
FE4-S1-FE2	71.60	66.00	78.37	66.28	26.86	62.23±6.62	67.09±3.19
S1-FE2-S3	102.67	106.66	62.28	109.21	39.52	114.04±8.04	105.95±2.31
S1-FE2-S4	102.09	110.31	34.47	109.21	39.52	104.29± 1.55	105.23±2.22
S1-FE3-S2	107.87	111.70	16.26	109.21	39.52	105.68±1.54	105.77±1.48
S1-FE3-S4	107.47	110.45	22.23	109.21	39.52	104.41±2.16	108.21±0.52
S1-FE4-S2	113.03	112.69	28.28	109.21	39.52	111.05±1.40	110.56±1.40
S1-FE4-S3	101.97	106.50	59.25	109.21	39.52	105.86±2.75	109.33±5.20
S2-FE1-S3	107.61	109.45	43.74	109.21	39.52	112.63±3.55	105.74±1.32
S2-FE1-S4	102.08	103.16	18.56	109.21	39.52	111.25± 6.48	101.64±0.31
S2-FE3-S4	107.74	101.98	18.52	109.21	39.52	110.49± 1.94	109.93±1.55
S2-FE4-S3	106.71	109.42	60.52	109.21	39.52	108.75±1.44	108.36±1.17
S3-FE1-S4	107.71	109.98	35.95	109.21	39.52	112.97±3.71	107.41±0.21
S3-FE2-S4	105.38	109.74	69.35	109.21	39.52	109.61±2.99	105.77±0.27

[1]DFT: density functional theory, [2]B3LYP: Becke three-parameter hybrid exchange and Lee Yang Parr, [3]LSDA/LANL2DZ**:** Los Alamos double-zeta basis; [4]SD: standard deviation

C

Geometry	External Angle (^{0})
Model System	**Fe_4S_4(S-Cys)$_3$(O-Gln): 1026A**

Bond calculations	X-ray	QM				AFTER-MD	
Bond description	1H7X	B3LYP (Model 1)		GFN1-xTB (Model 2)		Model 1	Model 2
	Angle (^{0})	Equilibrium Angle (^{0})	Force constant (kcal mol^{-1}rad^{-2})	Equilibrium Angle (^{0})	Force constant (kcal mol^{-1}rad^{-2})	Angle (^{0}) Mean & SD	Angle (^{0}) Mean & SD
C -OE155-FE3	117.29	130.30	75.86	115.32	41.23	115.29± 1.41	114.42±2.02
CT-SG90-FE1	99.53	103.98	106.68	107.39	100.90	108.04±6.02	107.92±5.93
CT-SG135-FE2	102.34	107.39	100.90	107.40	100.90	105.06±1.92	107.41±3.58
CT-SG129-FE4	100.25	97.96	96.66	107.39	100.90	99.42±0.58	107.20±4.91
N-C-OE155-H	104.50	122.90	80.0	118.02	44.55	113.34±6.25	116.93±8.78
SG90-CT-H	107.85	109.50	50.0	104.33	23.56	100.25±5.37	103.85± 2.83
SG135-CT-H	108.36	109.50	50.0	104.33	23.56	109.45±0.77	103.96±3.11
SG129-CT-H	105.31	109.50	50.0	104.33	23.56	98.83±4.58	101.12±2.96
SG90-FE1-S2	106.08	111.30	41.30	113.28	36.14	108.03±1.38	112.68±4.66
SG90-FE1-S3	111.94	115.56	63.82	113.28	36.14	112.26±0.23	113.07±0.80
SG90-FE1-S4	116.45	106.49	65.80	113.28	36.14	106.80±6.82	113.79±1.88
SG135-FE2-S3	112.62	115.18	52.84	113.28	36.14	105.02±5.37	112.46±0.11
SG135-FE2-S4	107.32	109.23	75.62	113.28	36.14	105.13±1.54	113.63± 4.46
SG135-FE2-S1	100.34	105.58	47.43	113.28	36.14	108.76±5.95	113.69± 9.43
OE155-FE3-S2	109.87	111.71	34.09	113.08	40.55	109.75±0.08	111.78±1.35
OE155-FE3-S4	106.39	109.31	59.86	113.08	40.55	114.93±6.03	112.69±4.45
OE155-FE3-S1	105.27	107.58	51.72	113.08	40.55	108.63±2.37	111.81±4.62
SG129-FE4-S2	110.83	112.42	53.37	113.28	36.14	106.42±3.12	113.16±1.64
SG129-FE4-S3	113.52	111.58	46.08	113.28	36.14	110.62±2.05	113.01±0.36
SG129-FE4-S1	99.46	104.01	64.85	113.28	36.14	106.51±4.99	100.49±0.72

[1]DFT: density functional theory, [2]B3LYP: Becke three-parameter hybrid exchange and Lee Yang Parr, [3]GFN1-xTB DA/LANL2DZ**:** Los Alamos double-zeta basis; [4]SD: standard deviation

Table S2.7. Assessment of **A** bond length, **B** internal and **C** external angles (Å) calculated with X-ray, DFT (B3LYP) and (LSDA/LANL2DZ) method for the molecular cluster model ([Fe_4S_4(S-Cys)$_4$]) 1027A of native DPD protein. Table adapted from Tendwa et al., (2021)

A

Geometry	**Bond length (Å)**						
Model System	**([Fe_4S_4(S-Cys)$_4$]): 1027A**						
Bond calculations	**X-ray**	**QM**				**AFTER-MD**	
Bond description	**1H7X**	**B3LYP (Model 1)**		**LSDA/LANL2DZ (Model 2)**		**Model 1**	**Model 2**
	Bond length *(Å)*	**Equilibrium Bond length** *(Å)*	**Force constant [Kr]** ***(kcal mol^{-1} Å^{-2})***	**Equilibrium Bond length** *(Å)*	**Force constant [Kr]** ***(kcal mol^{-1} Å^{-2})***	**Bond length** *(Å)* **Mean & SD**	**Bond length** *(Å)* **Mean & SD**
FE1_S2	2.63	2.24	45.10	2.22	89.23	2.23±0.28	2.22±0.29
FE1_S3	2.43	2.25	64.30	2.22	89.23	2.24± 0.13	2.23±0.14
FE1_S4	2.60	2.21	58.30	2.22	89.23	2.23±0.26	2.23±0.26
FE1_SG140	2.30	2.40	36.10	2.33	39.77	2.39± 0.06	2.35±0.035
FE2_S1	2.61	2.24	66.30	2.22	89.23	2.25±0.25	2.22± 0.27
FE2_S3	2.55	2.24	69.00	2.22	89.23	2.23±0.22	2.22±0.23
FE2_S4	2.25	2.23	51.20	2.22	89.23	2.24±0.01	2.23±0,01
FE2_SG87	2.24	2.35	50.00	2.33	39.77	2.40±0.11	2.34±0.07
FE3_S1	2.44	2.24	60.80	2.22	89.23	2.24± 0.14	2.23±0.15
FE3_S2	2.38	2.23	45.30	2.22	89.23	2.32±0.04	2.23±0.11
FE3_S4	2.25	2.19	64.70	2.22	89.23	2.19± 0.13	2.23±0.11
FE3_SG79	2.30	2.37	40.20	2.33	39.77	2.40±0.07	2.35±0.04
FE4_S1	2.55	2.27	41.50	2.20	89.23	2.26±0.21	2.22±0.23
FE4_S2	2.25	2.27	61.80	2.22	89.23	2.28±0.02	2.23±0.01
FE4_S3	2.60	2.26	57.00	2.22	89.23	2.27± 0.23	2.22±0.27
FE4_SG82	2.40	2.41	37.10	2.33	39.77	2.37±0.02	2.33±0.05

[1]DFT: density functional theory, [2]B3LYP: Becke three-parameter hybrid exchange and Lee Yang Parr, [3]LSDA/LANL2DZ**:** Los Alamos double-zeta basis; [4]SD: standard deviation

B

Geometry	Internal Angle (^{0})						
Model System	**([Fe_4S_4(S-Cys)$_4$]): 1027A**						
Bond calculations	**X-ray**	**QM**				**MD**	
Bond angle description	**1H7X**	**B3LYP (Model 1)**		**LSDA/LANL2DZ (Model 2)**		**Model 1**	**Model 2**
	Angle *(^{0})*	**Equilibrium Angle** *(^{0})*	**Force constant** *(kcal mol^{-1}rad^{-2})*	**Equilibrium Angle** *(^{0})*	**Force constant** *(kcal mol^{-1}rad^{-2})*	**Angle** *(^{0})* **Mean & SD**	**Angle** *(^{0})* **Mean & SD**
FE1-S2-FE3	67.72	67.16	42.52	66.28	26.86	66.86±0.61	68.54±0.58
FE1-S2-FE4	67.87	67.08	61.00	66.28	26.86	65.27±1.84	68.40±0.12
FE1-S3-FE2	68.36	66.89	27.43	66.28	26.86	67.35±0.71	69.70±0.94
FE1-S3-FE4	64.95	70.89	36.63	66.28	26.86	63.30±1.17	67.45±1.77
FE1-S4-FE2	66.89	67.77	77.31	66.28	26.86	62.02±0.61	69.32±1.72
FE1-S4-FE3	70.03	67.02	46.02	66.28	26.86	60.89±6.46	67.74±1.62
FE2-S1-FE3	68.46	68.38	64.00	66.28	26.86	67.33±0.80	66.97±1.05
FE2-S3-FE4	72.01	67.59	35.00	66.28	26.86	64.79±5.11	67.77±3.00
FE2-S4-FE3	69.43	68.59	42.93	66.28	26.86	62.23±5.09	67.67±1.24
FE3-S1-FE4	67.82	66.06	54.32	66.28	26.86	64.75±2.17	68.51±0.49
FE3-S2-FE4	67.71	67.32	36.84	66.28	26.86	66.98±0.52	69.31±1.13
FE4-S1-FE2	69.42	66.62	67.61	66.28	26.86	62.78±4.07	67.45±1.40
S1-FE2-S3	107.69	104.82	49.55	109.21	39.52	108.31±0.43	109.66±1.39
S1-FE2-S4	104.28	107.73	29.29	109.21	39.52	107.45±2.25	103.33±0.67
S1-FE3-S2	117.43	111.70	44.49	109.21	39.52	117.93±0.35	109.85±5.36
S1-FE3-S4	107.59	108.93	76.82	109.21	39.52	109.23±1.16	109.66±1.46
S1-FE4-S2	106.71	109.34	42.92	109.21	39.52	108.31±1.13	110.03±2.35
S1-FE4-S3	101.60	103.08	16.77	109.21	39.52	112.49±7.70	105.50±2.76
S2-FE1-S3	107.70	110.60	28.91	109.21	39.52	109.76± 1.46	109.65±1.38
S2-FE1-S4	105.51	105.86	36.64	109.21	39.52	111.75±4.41	108.32±1.99
S2-FE3-S4	101.61	106.90	42.99	109.21	39.52	108.4±4.80	107.76±4.35
S2-FE4-S3	106.67	108.94	20.07	109.21	39.52	110.93±2.24	108.45±0.49
S3-FE1-S4	112.33	109.89	59.08	109.13	39.52	111.41±0.65	105.10±5.11
S3-FE2-S4	107.37	109.86	37.10	109.21	39.52	104.05±2.35	109.54±1.53

[1]DFT: density functional theory, [2]B3LYP: Becke three-parameter hybrid exchange and Lee Yang Parr, [3]LSDA/LANL2DZ**:** Los Alamos double-zeta basis; [4]SD: standard deviation

C

Geometry	External Angle (^{0})						
Model System	**([Fe_4S_4(S-Cys)$_4$]): 1027A**						
Bond calculations	**X-ray**	**QM**				**AFTER-MD**	
Bond description	**1H7X**	**B3LYP (Model 1)**		**GFN1-xTB (Model 2)**		**Model 1**	**Model 2**
	Angle *(^{0})*	**Equilibrium Angle** *(^{0})*	**Force constant** *(kcal mol^{-1}rad^{-2})*	**Equilibrium Angle** *(^{0})*	**Force constant** *(kcal mol^{-1}rad^{-2})*	**Angle (^{0}) Mean & SD**	**Angle (^{0}) Mean & SD**
CT-SG78-FE3	123.43	108.65	100.02	115.32	41.23	108.76±10.37	114.21±6.52
CT-SG139-FE1	101.35	105.21	94.60	107.39	100.90	105.13±2.67	107.30±4.21
CT-SG86-FE2	102.43	107.61	107.61	107.39	100.90	110.52±6.43	107.14±4.04
CT-SG81-FE4	100.25	103.60	106.65	107.39	100.90	105.82±3.94	107.56±5.17
SG139-CT-H	114.50	122.90	80.0	118.02	44.55	114.93 ± 0.30	116.43±1.36
SG86-CT-H	107.58	109.50	50.0	104.33	23.56	106,63±0.67	100.91±4.71
SG78-CT-H	108.61	109.50	50.0	104.33	23.56	112.28±2.60	103.65±3.51
SG81-CT-H	105.01	109.50	50.0	104.33	23.56	110.72±4.03	103.12±1.34
SG139-FE1-S2	107.08	109.82	53.22	113.28	36.14	105.60±1.05	113.10±4.25
SG139-FE1-S3	115.94	111.05	45.22	113.28	36.14	110.80± 3.63	113.16±1.97
SG139-FE1-S4	112.45	109.60	51.92	113.28	36.14	112.52±0.05	112.64±0.13
SG86-FE2-S3	112.62	109.29	45.25	113.28	36.14	106.83±4.09	111.18±1.01
SG86-FE2-S4	117.32	116.82	63.88	113.28	36.14	112.53±3.38	113.37±2.80
SG86-FE2-S1	100.34	107.61	53.36	113.28	36.14	99.83±0.36	111.84±8.13
SG78-FE3-S2	109.78	113.79	62.02	113.28	36.14	110.21±0.30	112.49± 1.91
SG78-FE3-S4	106.39	113.14	55.31	113.28	36.14	109.30±2.05	112.38± 4.24
SG78-FE3-S1	105.27	102.35	62.04	113.28	36.14	108.77±2.47	113.18±5.59
SG81-FE4-S2	110.83	116.54	47.02	113.28	36.14	112.78±1.38	113.33±1.77
SG81-FE4-S3	113.26	110.21	51.70	113.28	36.14	107.39±4.15	112.98±0.20
SG81-FE4-S4	110.66	107.95	53.93	113.28	36.14	109.50±0.82	111.56±0.64

[1]DFT: density functional theory, [2]B3LYP: Becke three-parameter hybrid exchange and Lee Yang Parr, [3]GFN1-xTB: eXtended TB; [4]SD: standard deviation

Table S2.8. Dihedral associated force constants for X-ray and post-MD simulation for both models' clusters ([Fe4S4(S-Cys)3(S-Gln)]) and ([Fe4S4(S-Cys)4]) of Native DPD protein. Table adapted from Tendwa et al., (2021).

Fe_4S_4 Clusters name	Geometry	Dihedral angles (^{0})						
	Model System	Fe_4S_4(S-Cys)$_3$(O-Gln) and ([Fe_4S_4(S-Cys)$_4$]) Clusters						
	Bond	X-ray	QM				AFTER-MD	
	Bond Description	X-ray	B3YLP (Model 1)		LSDA/LANL2DZ (Model 2)		Model 1	Model 2
		Average dihedral angle (^{0})	Equilibrium Bond angle (^{0})	Force constant (kcal/mol rad^2)	Equilibrium Bond angle (^{0})	Force constant (kcal/mol rad^2)	Bond angle (^{0}) Mean & SD	Bond angle (^{0}) Mean & SD
Cluster 1026_A	S-FE-SG-CT	67.43					74.87± 5.26	76.64±6.51
	S-FE-OE-CT	58.67					62.56± 2.75	63.42±3.35
	CT-CT-FE-S	-112.78					-105.36±5.25	-105.79±4.94
	CT-CT-FE-SG	-54.79					-50.87±2.77	-49.93±3.43
Cluster 1027_A	S-FE-SG-CT	63.16					76.92±76.92	75.48± 8.71
	S-FE-SG-CT	56.23					63.54±5.17	63.74± 5.31
	CT-CT-FE-S	-111.87					-103.33±6.04	-105.67±4.38
	CT-CT-FE-SG	-56.42					-51.76±3.30	-51.93±3.1
Cluster 1028_B	S-FE-SG-CT	63.01					76.75±9.72	76.86±9.79
	S-FE-SG-CT	59.99					64.40±3.12	69.87±6.99
	CT-CT-FE-S	-112.61					-104.75±5.56	-105.23±5.22
	CT-CT-FE-SG	-55.28					-50.87±3.12	-50.11±3.66
Cluster 1028_B	S-FE-SG-CT	65.14					77.56±8.78	76.69±8.17
	S-FE-SG-CT	60.22					67.41±5.08	68.87±6.12
	CT-CT-FE-S	-110.94					-103.57± 5.21	-105.38± 3.93
	CT-CT-FE-SG	-58.23					-54.08±2.93	-52.39±4.13

[1]DFT: density functional theory, [2]B3LYP: Becke three-parameter hybrid exchange and Lee Yang Parr, [3]LSDA/LANL2DZ**:** Los Alamos double-zeta basis; [4]SD: standard deviation

CHAPTER 3

Table S3.1. Changing aspects of 5-FU retention within the active site and frame range selection from equilibrated section for successive dynamic residue network calculation (DRN) as showed by MD simulations of drug bound system.

System	Monomer A				Monomer B			
	Drug exit (Y/N)	**Exit time (ns)**	**Exit frame**	**DRN calculation Monomer A frame range**	**Drug exit (Y/N)**	**Exit time (ns)**	**Exit frame**	**DRN calculation monomer B frame range**
WT	N	N/A	N/A	550000 - 580000	N	N/A	N/A	550000 - 580000
C29R	N	N/A	N/A	470000-500000	N	N/A	N/A	470000-500000
M166V	N	N/A	N/A	270000-300000	Y	420	42000	270000-300000
Y186C	N	N/A	N/A	250000-280000	N	N/A	N/A	250000-280000
S534N	Y	148	14800	190000-220000	N	N/A	N/A	190000-220000
I543V	N	N/A	N/A	250000-280000	Y	350	35000	250000-280000
D949V	N	N/A	N/A	30000-60000	N	N/A	N/A	30000-60000

Table S3.2. Frame range selection from equilibrated area for successive dynamic residue network calculation (DRN) as shown by MD simulations of drug free systems.

System	Monomer A	Monomer B
	DRN calculation Monomer A frame range	**DRN calculation monomer B frame range**
WT	550000 -580000	550000 -580000
C29R	470000-500000	470000-500000
M166V	270000-300000	270000-300000
Y186C	250000-280000	250000-280000
S534N	190000-220000	190000-220000
I543V	250000-280000	250000-280000
D949V	30000-60000	30000-60000

Table S3.3.1 DPD missense mutations identified in the African population by the 1000 genomes and Exom sequencing projects (ESP MAF). Variants were classified as rare or common based on their minor allele frequency (MAF) of 0.01. Using VAPOR web servers, we predicted the effect of six clinically significant mutations on DPD structure and function.

Variant	SNP ID	DPD activity ± SD	P value	Exom sequencing project (ESP MAF > 0.01)				Functional impact			Stability change	Reference
				Population allele frequency count				HUMA-VAPOR				
				Adenine (A) or cytosine (C)	Guanine (G) or thymine (T)	African American	African	Polyphen-2	Proven	PhD-SNP	MUpro	PHARMGKB
C29R	rs1801265	n/a	n/a	A:0.558 (738)	G:0.442 (584)	0.60032	0.4512	Benign	Tolerated	Neutral	Decreases	Doi: 10.1007/BF02698048
M166V	rs2297595	1.20± 0.07	0.025	**C**: 0.031 (41)	T:0.969 (1281)	0.04085	0.0488	Probably damaging	Disease	Neutral	Decreases	Doi:10.1158/0008-5472.CAN-13-2482
Y186C	rs115232898	0.85± 0.03	0.027	C:0.023 (30)	T:0.977 (1292)	0.022202	0.0028	Probably damaging	Disease	Disease	Decreases	Doi:10.1038/clpt.2013.69
S534N	rs1801158	na	na	C:0.998 (1319)	T:0.002 (3)	0.00454	0.0061	Probably damaging	Tolerated	Neutral	Decreases	Doi: 10.1158/0008-5472.CAN-13-2482
I543V	rs1601159	n/a	n/a	A:0.474 (626)	G:0.526 (696)	0.155524	0.2030	Benign	Tolerated	Neutral	Decreases	Doi:10:1038/clpt.2013.69
D949V	rs67376798	0.78± 0.17	0.11	A: 0.001 (1)	T: 0.999 (1321	0.00091	0.0020	Probably damaging	Disease	Disease	Increases	Doi:org/10.1016/j.bbadis

Table S3.3.2 DPD missense mutations identified in the African population by the 1000 genomes and Exom sequencing projects (ESP MAF). Variants were classified as rare or common based on their minor allele frequency (MAF) of 0.01. Allele frequency in West and East African as well as African American populations are tabulated

Variant	SNP ID	African-Region		Allele frequency per population	
		Sub-region	Country	Adenine (A) or cytosine (C)	Guanine (G) or thymine (T)
C29R	rs1801265	West Africa	Nigeria	A:1000(198)	
			Gambia	A: 0.982 (222)	
		East Africa	Kenya	A: 0.975 (193)	
		African America		A: 0.982 (234)	
M166V	rs2297595	West Africa	Nigeria	C: 0.005 (1)	T: 0.995 (197)
			Gambia	C: 0.027 (6)	T: 0.973 (220)
		East Africa	Kenya	C: 0.101 (20)	T: 0899 (178)
		African America		C: 0.049 (6)	T: 0.951 (116)
Y186C	rs115232898	West Africa	Nigeria	C: 0.010 (2)	T: 0.990 (190)
			Gambia	C: 0.040 (9)	T: 0.960 (217)
		East Africa	Kenya	C: 0.015 (3)	T: 0.985 (195)
		African America		C: 0.016 (7)	T: 0984 (120)
S534N	rs1801158	West Africa	Nigeria	C: 1.000 (198)	
			Gambia	C: 1.000 (226)	
		East Africa	Kenya	C: 1.000 (198)	
		African America		C: 0.975 (119)	T: 0.025 (3)
I543V	rs1601159	West Africa	Nigeria	A: 0.449 (89)	G: 0.551 (109)
			Gambia	A: 0.509 (115)	G: 0.491 (11)
		East Africa	Kenya	A: 0.505 (100)	G: 0.495 (98)
		African America		A: 0.508 (62)	G: 0.492 (60)
D949V	rs67376798	West Africa	Nigeria		T: 0.992 (121)
			Gambia		T: 1.000 (198)
		East Africa	Kenya		T: 1.000 (226)
		African America			T: 1000 (198)

Table S3.4. Common residues identified as being involved in interface formation by consensus from several (PDBePISA, HSPred, Hotregion, PPCheck, Robetta and PyMol interfaceResidues) programs and all the cofactors' binding residues.

Interphase residues	A2, V4, L5, S6, K7, D8, L16, L18, 19N, P20, R21, T22, Q23, S24, H25, A26, A27, L28, H29, T30, L31, K33, K34, K37, W41, R43, N44, K46, H64, T65, L68, G69, E70, R71, G73, L74, R75, E76, M77, R78, L80, C82, A83, D84, A85, P86, L88, K89, S90, C91, P92, T93, H94, S118, D119, S133, D134, L135, G138, G139, C140, N141, A144, T145, E146, E147, N151, F205, G366, F367, V368, N369, R371, A372, V373, K384, G386, F387, L388, F390, R410, Q413, Q425, I426, V427, H428, L429, P466, E467, A486, E491, N494, D495, Q498, W501, P519, E520, L521, P522, L523, A552, P553, S557, S558, M559, R561, R562, A566, D579, V583, T584, N585, V586, S587, P588, R589, I590, V591, R592, T594, T595, S596, G597, P598,M599, Y600, Y601, G602, S605, F607, E615, K617, A619, Y620, Q623, S624, E627, L628, E677, G679, M680, G681, N713, V714, T715, D716, S718, V738, S739, G740, L741, M742, L744, G743, L745, K746, D747, G748, T748, T749, V753, G754, A756, K758, R759, T760, T761, V765, T768, A769, L770, R771, P772, I773, L775, R776, T779, T780, Q804, H807, S808, V819, Q820, N821, Q822, D823, F824, T825, Q825, Q829, D850, G851, Q860, K861, G862, A928, Y931, L932, G933, T934, F935, E937, L938, S939, N940, I954, G943, N955, C956, G957, M961, T962, N964, D965, S966, G967, Y968, G969, P975, H978, G988, C989, T990, L991, S994, V995, C996, P997, I999, P1013, K1014, R1015, G1016 and L1017
5-fluorouracils	N609, N668, G764, T737 and N736 (E611, L612, I613)
FMN	N609, A549, S550, A551, A552, K574, T575, I590, A609, E611, I613, N668, K709, T735, T737, S766, T767, I770, T793, G794, I796, G794, V815, C816, S817, A818 and Q820,
NDP	L226, A291, L289, G339, A340, G341, N342, T343, R364, K365, R371, P393, G439 S440, A437, D481, G483, M485 and A486, N487
FAD	V129, C130, P131, L193, A195, A198, S199, F217, Q218, K219, Q220,T222, L226, S227, E230, I231, R235, K259, L261, G282, I283,G284, L285, P286, L310, D342, T343, D346, V447, G480, N481, L482, T489, V490 and S492.
Domain I ($Fe^{2+}_4S^{2-}_4$ clusters)	Residue (27-172)

Domain II (FAD)	Residues 173-286 and 442-524
Domain III (NADPH)	Residues 287-441
Domain IV (FMN & 5-FU)	Residues 525-847
Domain V (Fe2+4S2-4 clusters)	Residues 1-26 and 848-1025
$Fe^{2+}_4S^{2-}_4$ cluster residues.	C79, C82, C87, C91, C130, C136, C140, Q156, C953, C956, C959, C963, C986, C989, C992 AND C996

Table S3.5.1. Averaged *betweenness centrality (BC)* of potential hubs found in drug free and drug bound WT systems, hubs gained due to mutation, and hubs that were maintained despite the presence of mutation.

BC		Drug free		Drug bound	
Systems		Monomer A	Monomer B	Monomer A	Monomer B
WT	Hubs	38, 61, 64, 66, 67, 71, 73, 74, 75, 76, 77, 80, 95, 97, 101, 102, 104, 108, 116, 117, 120, 121, 132, 133, 136, 137, 142, 145, 146, 152, 153, 155, 235, 350, 370, 372, 373, 386, 493, 597, 598, 599, 603, 768, 775, 799, 802, 803, 826, 862 = **50**	62, 63, 65, 66, 86, 90, 101, 102, 105, 108, 114, 117, 121, 127, 130, 131, 132, 133, 137, 139, 142, 143, 144, 147, 149, 155, 160, 361, 367, 370, 372, 373, 376, 599, 601, 603, 604, 816, 818, 826, 955, 991 = **42**	61, 65, 80, 85, 97, 101, 102, 108, 114, 116, 117, 121, 122, 123, 125, 126, 130, 132, 133, 137, 140, 142, 155, 200, 227, 231, 234, 310, 372, 373, 374, 375, 597, 598, 599, 604, 605, 768, 775, 776, 797, 799, 803, 804, 816, 825, 958, 962, 996, 997 = **50**	22, 44, 62, 69, 76, 77, 92, 95, 97, 101, 102, 113, 116, 120, 124, 125, 127, 131, 133, 135, 137, 138, 141, 142, 145, 149, 155, 228, 231, 234, 370, 372, 373, 374, 380, 383, 595, 602, 603, 605, 775, 799, 822, 853, 856 = **45**
C29R	Retained hubs	66, 71, 76, 77, 108, 116, 120, 132, 142, 152, 155, 373, 597, 598, 603, 775, 799, 803, 826 = 19	62, 86, 108, 127, 131, 142, 147, 155, 367, 370, 372, 373, 599 = 13	97, 108, 116, 123, 130, 140, 142, 155, 372, 373, 598, 797, 825, 958 = 14	22, 62, 77, 101, 102, 116, 120, 124, 127, 131, 141, 142, 149, 155, 370, 372, 373, 603, 775, 799, 822 = 21
	Gained hubs	62, 63, 68, 70, 72, 85, 124, 131, 147, 149, 150, 157, 158, 387, 604, 859, 958, 995, 997 = 19	22, 60, 67, 70, 71, 73, 81, 87, 95, 116, 120, 124, 156, 158, 159, 493, 598, 605, 772, 799, 802, 863, 997 = 23	59, 60, 67, 88, 95, 120, 135, 138, 144, 145, 146, 150, 157, 235, 493, 595 = 16	63, 66, 67, 68, 70, 71, 81, 82, 130, 146, 147, 150, 157, 159, 160, 349, 366, 493, 597, 598, 604, 765, 862, 962, 995, 997 = 26
M166V	Retained hubs	66, 67, 73, 75, 95, 97, 101, 116, 117, 132, 142, 145, 155, 372, 373, 493, 597, 598, 599, 775, 799, 803 = 22	66, 108, 114, 117, 131, 132, 133, 137, 142, 147, 149, 155, 372, 373, 376, 599, 816 = 17	97, 108, 116, 123, 133, 137, 140, 142, 155, 372, 373, 598, 599, 604, 605, 797, 958, 962, 997 = 19	22, 116, 137, 142, 155, 234, 373, 799 = 8
	Gained hubs	22, 23, 45, 65, 72, 105, 128, 130, 131, 138, 140, 141, 144, 150, 159, 375, 381, 605 = 18	68, 70, 76, 82, 83, 85, 116, 145, 150, 153, 157, 158, 159, 236, 344, 345, 374, 437, 597, 598, 765, 832, 958, 962, 994, 995, 997 = 27	62, 68, 70, 77, 81, 82, 91, 98, 104, 105, 106, 120, 138, 141, 144, 146, 147, 149, 150, 152, 159, 160, 493, 589, 772, 861 = 26	42, 61, 68, 70, 75, 79, 80, 81, 96, 104, 105, 114, 117, 121, 144, 147, 150, 159, 236, 368, 375, 597, 598, 599, 803, 863, 997 = 27
Y186C	Retained hubs	67, 80, 108, 116, 117, 133, 137, 142, 155, 235, 373,	63, 65, 101, 102, 130, 132, 143, 144, 155, 160, 372, 373,	65, 101, 102, 108, 114, 116, 117, 130, 142, 155, 372, 373, 375,	44, 76, 77, 95, 97, 101, 116, 120, 124, 133, 137, 138, 142, 155,

		597, 598, 599 = 14	376, 599 =14	597, 599, 605, 775, 997 = 18	370, 372, 373, 380 = 18
	Gained hubs	92, 128, 130, 135, 138, 141, 150, 232, 374, 595, 596, 602, 605, 863, 956, 958, 962 = 17	63, 65, 101, 102, 130, 132, 143, 144, 155, 160, 372, 373, 376, 599 = 14	62, 63, 66, 68, 70, 73, 82, 141, 144, 147, 157, 158, 349, 493, 553, 595, 596, 832 = 18	43, 45, 63, 64, 68, 82, 108, 130, 144, 147, 150, 153, 156, 343, 376, 486, 487, 488, 598, 958, 962 = 21
S534N	Retained hubs	67, 77, 95, 97, 101, 102, 132, 146, 152, 155, 373, 599 = 12	101, 102, 117, 121, 130, 131, 132, 133, 137, 142, 149, 155, 160, 367, 370, 372, 373, 818, 955 = 19	65, 80, 97, 101, 102, 108, 114, 116, 117, 121, 155, 775, 797, 799, 803, 804, 962 = 17	76, 77, 113, 116, 124, 131, 137, 138, 142, 155, 372, 373, 380, 595, 603 = 15
	Gained hubs	63, 70, 78, 79, 82, 83, 91, 93, 127, 150, 158, 160, 384, 490, 528, 593, 596, 604, 605, 822, 853, 997 = 22	44, 46, 47, 60, 72, 82, 85, 103, 116, 120, 122, 123, 125, 129, 141, 157, 374, 528, 553, 598, 765, 775, 797, 853, 863, 958, 995 = 26	22, 62, 63, 67, 74, 76, 77, 82, 91, 106, 135, 138, 146, 150, 157, 341, 349, 376, 379, 380, 381, 486, 487, 553, 591, 826, 832 = 27	43, 64, 65, 68, 70, 82, 96, 108, 117, 121, 130, 132, 147, 150, 151, 152, 157, 158, 160, 345, 375, 376, 381, 488, 598, 599, 771, 772, 825, 859, 958, 962, 966, 996, 997 = 35
I543V	Retained hubs	61, 67, 75, 77, 80, 101, 102, 104, 116, 120, 132, 137, 142, 145, 146, 155, 373, 493, 599, 775, 799, 803 = 22	66, 102, 108, 132, 142, 144, 147, 155, 370, 372, 373, 599, 604, 816 = 14	65, 108, 114, 116, 117, 123, 142, 155, 200, 372, 598, 599, 797, 958, 997 = 15	22, 44, 76, 101, 102, 116, 120, 131, 141, 142, 155, 370, 372, 373, 603 = 15
	Gained hubs	72, 83, 91, 130, 138, 140, 143, 150, 368, 381, 591, 595, 601, 604, 605, 804, 863, 958, 962 = 19	44, 46, 47, 60, 72, 82, 85, 103, 116, 120, 122, 123, 125, 129, 141, 157, 374, 528, 553, 598, 765, 775, 797, 853, 863, 958, 995 = 27	63, 70, 82, 124, 138, 141, 146, 147, 150, 157, 236, 346, 386, 493, 591, 832 = 16	45, 65, 80, 81, 82, 85, 88, 108, 132, 144, 147, 150, 151, 363, 367, 379, 488, 598, 821, 958, 962, 998 = 22
D949V	Retained hubs	76, 77, 101, 102, 116, 117, 121, 133, 137, 142, 145, 146, 155, 370, 372, 373, 599, 799, 803 = 19	62, 63, 65, 108, 117, 127, 142, 143, 155, 160 = 10	61, 114, 116, 117, 137, 140, 142, 155, 234, 372, 373, 597, 598, 599, 799, 803, 958, 962 = 18	22, 62, 95, 113, 116, 125, 141, 155, 370, 595, 602, 775, 799 = 13
	Gained hubs	65, 70, 79, 81, 91, 122, 150, 151, 157, 158, 159, 367, 375, 376, 596, 817, 994, 995, 999 = 19	61, 73, 76, 77, 95, 110, 113, 124, 128, 129, 140, 141, 146, 150, 151, 153, 157, 235, 346, 363, 380, 386, 493, 596, 597, 602, 606, 771, 775, 799, 801, 803, 804, 995, 998, 999, 1001 = 37	22, 23, 67, 70, 73, 82, 120, 127, 129, 138, 141, 143, 145, 147, 150, 153, 157, 158, 370, 376, 493, 832, 861, 863, 995 = 25	63, 64, 73, 75, 85, 88, 91, 108, 144, 160, 232, 235, 566, 596, 597, 599, 768, 772, 803, 816, 825, 862, 956, 958, 962 = 25

Table S3.5.2. Averaged *degree centrality (DC)* of potential hubs found in drug free and drug bound WT systems, hubs gained due to mutation, and hubs that were maintained despite the presence of mutation.

DC		Drug free		Drug bound	
Systems		Monomer A	Monomer B	Monomer A	Monomer B
WT	Hubs	72, 90, 108, 110, 113, 116, 138, 155, 159, 198, 199, 200, 201, 202, 203, 207, 218, 282, 314, 356, 435, 482, 493, 499, 548, 576, 625, 710, 740, 778, 779, 781, 790, 807, 808, 813, 832, 833, 837, 928 = 40	108, 110, 155, 192, 193, 200, 201, 202, 203, 214, 225, 282, 335, 337, 482, 496, 499, 554, 625, 653, 654, 694, 707, 711, 729, 732, 775, 778, 779, 781, 790, 808, 813, 833, 990 = 35	72, 108, 110, 113, 155, 157, 199, 200, 203, 206, 207, 218, 232, 313, 314, 335, 350, 351, 478, 480, 542, 554, 605, 625, 628, 641, 653, 654, 665, 666, 707, 714, 729, 732, 740, 778, 779, 781, 790, 807, 808, 813, 833, 837, 928 = 45	72, 108, 110, 113, 138, 139, 153, 155, 159, 203, 207, 225, 241, 337, 356, 478, 568, 595, 603, 605, 625, 653, 654, 694, 707, 729, 732, 778, 779, 781, 790, 796, 807, 808, 832, 833, 928, 962 = 38
C29R	Retained hubs	108, 110, 113, 116, 155, 198, 199, 200, 201, 202, 203, 207, 218, 282, 482, 493, 499, 625, 740, 778, 779, 781, 790, 807, 808, 832, 833, 928 = 28	108, 110, 155, 201, 202, 203, 214, 282, 482, 496, 499, 554, 625, 653, 654, 729, 732, 778, 779, 781, 790, 808, 813, 833 = 24	72, 108, 110, 113, 155, 157, 199, 203, 207, 218, 480, 554, 625, 628, 707, 729, 732, 740, 778, 779, 781, 790, 807, 808, 813, 833, 837, 928 = 28	72, 108, 113, 155, 159, 203, 225, 337, 568, 595, 625, 653, 778, 779, 781, 790, 807, 808, 832, 928 = 20
	Gained hubs	67, 197, 335, 337, 431, 478, 479, 490, 495, 496, 554, 568, 628, 653, 729, 732, 775, 835, 990, 995 = 20	72, 154, 159, 190, 198, 199, 207, 318, 469, 478, 479, 480, 486, 493, 495, 542, 641, 657, 710, 740, 807, 816 = 22	67, 90, 116, 138, 139, 159, 194, 198, 201, 202, 282, 337, 344, 356, 479, 482, 486, 492, 493, 495, 496, 499, 568, 595, 734, 765, 810, 835, 990 = 29	218, 314, 318, 359, 482, 554, 618, 710, 721, 740, 767, 813, 835, 837 = 14
M166V	Retained hubs	108, 110, 113, 138, 155, 159, 198, 199, 201, 202, 203, 207, 282, 356, 482, 493, 499, 625, 740, 778, 781, 790, 808, 832, 833, 837, 928 = 27	108, 110, 200, 201, 202, 203, 282, 482, 554, 653, 707, 732, 778, 779, 781, 790, 808, 813, 833 = 19	72, 108, 110, 113, 155, 199, 203, 206, 314, 554, 605, 625, 628, 654, 729, 732, 740, 778, 779, 781, 790, 807, 808, 813, 833, 837, 928 = 27	72, 108, 110, 113, 155, 568, 595, 625, 653, 732, 778, 781, 790, 808, 832 = 15
	Gained hubs	67, 206, 337, 344, 347, 431, 437, 469, 471, 478, 479, 495, 496, 554, 568, 605, 653, 676, 707, 732, 775, 776, 835 = 23	72, 113, 154, 194, 195, 198, 199, 206, 218, 344, 480, 576, 595, 605, 740, 807, 832, 835, 837 = 19	116, 138, 198, 201, 202, 282, 462, 499, 503, 548, 568, 589, 595, 832, 835 = 15	314, 500, 710, 721, 770, 813, 829 = 7

Y186C	Retained hubs	72, 90, 108, 110, 113, 116, 138, 155, 159, 198, 199, 201, 202, 203, 218, 282, 435, 482, 493, 499, 548, 625, 740, 778, 779, 781, 790, 807, 808, 813, 832, 833, 837, 928 = 34	108, 110, 155, 202, 203, 282, 482, 496, 499, 554, 625, 653, 707, 732, 775, 778, 779, 781, 790, 808, 813, 833 = 22	72, 108, 110, 113, 155, 199, 203, 207, 218, 478, 480, 554, 625, 653, 654, 665, 729, 732, 740, 778, 781, 790, 807, 808, 813, 833 = 26	72, 108, 110, 113, 138, 153, 155, 159, 203, 207, 337, 356, 595, 625, 653, 654, 694, 707, 729, 732, 778, 779, 781, 790, 807, 808, 833 = 27
	Gained hubs	104, 125, 127, 139, 194, 206, 469, 479, 480, 486, 492, 495, 496, 542, 554, 568, 595, 628, 653, 721, 732, 775, 962 = 23	72, 113, 116, 138, 159, 160, 198, 199, 218, 314, 356, 469, 478, 479, 480, 486, 493, 495, 568, 628, 710, 740, 832, 837, 955 = 25	90, 138, 139, 154, 159, 198, 201, 282, 482, 486, 499, 508, 568, 589, 595, 657, 694, 710, 734, 770 = 20	90, 100, 116, 122, 140, 199, 218, 282, 314, 554, 608, 710, 740, 765, 775, 813, 835 = 17
S534N	Retained hubs	72, 108, 110, 155, 203, 218, 548, 625, 710, 740, 778, 779, 781, 790, 808, 813, 833, 837, 928 = 19	108, 110, 155, 201, 203, 282, 482, 554, 625, 653, 729, 778, 779, 781, 790, 808, 833 = 17	72, 108, 110, 113, 155, 199, 203, 218, 478, 625, 732, 740, 778, 781, 790, 808, 813, 833 = 18	72, 108, 110, 113, 138, 155, 203, 356, 625, 653, 729, 778, 779, 781, 790, 808, 833 = 17
	Gained hubs	139, 225, 238, 313, 335, 431, 554, 641, 653, 707, 732, 772, 775 = 13	72, 75, 90, 113, 138, 153, 194, 198, 199, 206, 218, 313, 314, 356, 478, 480, 568, 606, 628, 765, 767, 810, 928 = 22	67, 90, 116, 122, 138, 159, 198, 201, 202, 282, 337, 344, 361, 437, 479, 482, 493, 495, 496, 499, 568, 718, 721, 832, 835 = 25	282, 482, 496, 499, 542, 548, 554, 691, 740, 775, 813 = 11
I543V	Retained hubs	108, 110, 113, 116, 138, 155, 198, 199, 203, 282, 356, 482, 499, 778, 779, 781, 790, 808, 813, 832, 833, 928 = 22	108, 110, 155, 203, 225, 337, 482, 499, 554, 625, 653, 654, 732, 775, 778, 779, 781, 790, 808, 833 = 20	108, 110, 113, 155, 199, 200, 203, 207, 218, 478, 480, 554, 625, 628, 653, 654, 666, 729, 732, 740, 778, 779, 781, 790, 808, 813, 833, 928 = 28	72, 108, 110, 113, 138, 139, 155, 159, 203, 207, 625, 653, 778, 779, 781, 790, 808, 833 = 18
	Gained hubs	122, 139, 335, 337, 344, 437, 469, 478, 479, 480, 486, 492, 495, 554, 662, 775 = 16	72, 116, 125, 194, 199, 218, 227, 238, 241, 344, 356, 595, 628, 641, 667, 710, 734, 767, 810, 837 = 20	31, 116, 122, 138, 159, 194, 198, 201, 202, 282, 344, 347, 437, 482, 486, 496, 499, 589, 595, 691, 710, 832 = 22	90, 122, 127, 197, 198, 199, 201, 202, 218, 282, 314, 479, 480, 482, 486, 493, 495, 499, 554, 740, 767, 813, 816 = 23
D949V	Retained hubs	72, 108, 110, 155, 159, 203, 218, 710, 740, 778, 779, 781, 790, 808,	108, 110, 155, 201, 202, 282, 482, 496, 499, 554, 625, 653, 654, 729, 732, 775, 778, 779,	108, 110, 113, 199, 203, 207, 218, 554, 605, 625, 628, 665, 740, 778, 779, 781, 790, 808,	108, 110, 138, 155, 159, 203, 478, 595, 625, 653, 778, 779, 781, 790, 808,

		832, 833, 928 = 17	781, 790, 808, 813 = 22	813, 928 = 20	833, 962 = 17
	Gained hubs	65, 335, 431, 480, 554, 568, 628, 653, 654, 666, 707, 732, 775, 938 = 14	138, 153, 194, 198, 199, 218, 356, 479, 486, 493, 495, 572, 575, 734, 740, 814, 837 = 17	31, 90, 138, 139, 160, 194, 198, 201, 202, 282, 337, 493, 499, 589, 595, 662, 672, 681, 775, 810, 829, 832, 835 = 23	75, 90, 202, 228, 238, 282, 479, 480, 482, 486, 495, 499, 554, 606, 721, 740, 810, 816, 835, 837 = 20

Table S3.5.3. Averaged *closeness centrality (CC)* of potential hubs found in drug free and drug bound WT systems, hubs gained due to mutation, and hubs that were maintained despite the presence of mutation.

CC		Drug free		Drug bound	
Systems		Monomer A	Monomer B	Monomer A	Monomer B
WT	Hubs	67, 71, 72, 73, 74, 75, 76, 77, 78, 79, 101, 104, 105, 150, 151, 152, 155, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 768 = 28	73, 76, 77, 101, 104, 105, 146, 147, 150, 151, 155, 596, 597, 598, 599, 600, 601, 603, 605, 768 = 20	71, 72, 73, 75, 76, 79, 80, 81, 97, 98, 101, 102, 103, 104, 105, 137, 140, 141, 142, 145, 146, 147, 150, 151, 152, 155, 596, 597, 598, 599, 600, 601, 602, 604, 605, 768 = 36	67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 87, 97, 101, 104, 105, 137, 140, 141, 142, 145, 146, 147, 149, 150, 151, 152, 153, 155, 595, 596, 599, 602, 603, 605 = 38
C29R	Retained hubs	67, 71, 72, 73, 74, 75, 76, 77, 78, 79, 101, 104, 105, 150, 151, 152, 155, 596, 597, 598, 599, 600, 601, 603, 604, 605 = 26	73, 76, 77, 101, 104, 146, 147, 150, 151, 155, 596, 597, 598, 599, 600, 601, 603, 605, 768 = 19	71, 72, 73, 75, 76, 79, 80, 81, 97, 101, 102, 104, 105, 140, 141, 142, 145, 146, 147, 150, 151, 152, 155, 596, 597, 598, 599, 600, 601, 602, 604, 605, 768 = 33	67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 101, 104, 105, 140, 141, 142, 145, 146, 147, 149, 150, 151, 152, 155, 595, 596, 599, 602, 603, 605 = 34
	Gained hubs	68, 69, 70, 80, 140, 142, 145, 146, 147, 149, 153, 154, 158, 859, 862, 997 = 16	67, 68, 69, 70, 71, 72, 74, 75, 78, 79, 80, 81, 140, 141, 142, 145, 152, 154, 158, 602, 604, 997, 998 = 23	67, 68, 74, 77, 78, 149, 154, 158, 595 = 9	66, 81, 102, 154, 158, 159, 593, 597, 598, 604, 606, 862, 995, 997 = 15
M166V	Retained hubs	67, 72, 73, 75, 76, 78, 79, 101, 105, 150, 151, 152, 155, 597, 598, 599, 601, 605 = 18	73, 76, 77, 101, 104, 105, 147, 150, 151, 155, 597, 598, 599, 601 = 14	71, 72, 73, 75, 76, 79, 80, 81, 97, 98, 101, 104, 105, 140, 141, 142, 145, 146, 147, 150, 151, 152, 155, 596, 597, 598, 599, 600, 601, 602, 604, 605 = 32	67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 101, 104, 105, 140, 142, 146, 147, 149, 150, 151, 152, 155, 596, 599, 602, 605 = 30
	Gained hubs	80, 81, 140, 142, 146, 147 = 6	68, 70, 71, 72, 74, 75, 78, 79, 80, 140, 142, 149, 152, 997 = 14	67, 68, 70, 74, 77, 78, 148, 149, 603, 997 = 10	81, 597, 598, 600, 997, 999 = 6
Y186C	Retained	67, 71, 72, 73, 74, 75, 76, 77,	73, 76, 77, 101, 104, 105, 146,	71, 72, 73, 75, 76, 79, 80, 81,	67, 68, 69, 70, 71, 72, 73, 74,

	hubs	78, 79, 101, 104, 105, 150, 151, 152, 155, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 768 = 28	147, 150, 151, 155, 596, 597, 598, 599, 600, 601, 603, 605, 768 = 20	97, 98, 101, 102, 104, 105, 137, 140, 141, 142, 145, 146, 147, 150, 151, 152, 155, 596, 597, 598, 599, 600, 601, 602, 604, 605, 768 = 35	75, 76, 77, 78, 79, 80, 87, 97, 101, 104, 105, 137, 140, 141, 142, 145, 146, 147, 149, 150, 151, 152, 153, 155, 595, 596, 599, 605 = 36
	Gained hubs	68, 69, 70, 80, 81, 82, 97, 98, 102, 106, 132, 137, 138, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 153, 154, 156, 158, 594, 595, 606, 767, 769, 821, 997 = 34	67, 68, 69, 70, 71, 72, 74, 75, 78, 79, 80, 81, 97, 98, 102, 106, 137, 140, 141, 142, 144, 145, 149, 152, 153, 154, 156, 158, 159, 595, 602, 606, 769, 821, 822, 956, 958, 995, 997 = 39	65, 67, 68, 69, 70, 74, 77, 78, 82, 87, 143, 144, 148, 149, 153, 154, 156, 158, 595, 603, 606, 767, 769, 997 = 24	81, 82, 100, 102, 143, 144, 148, 154, 156, 158, 159, 597, 598, 600, 601, 604, 768, 769, 997 = 19
S534N	Retained hubs	67, 71, 72, 73, 74, 75, 76, 77, 78, 79, 101, 105, 150, 151, 152, 155, 596, 599, 601, 603, 604, 605 = 22	76, 101, 146, 147, 150, 151, 155, 598, 600 = 9	71, 72, 73, 75, 76, 79, 80, 101, 104, 105, 140, 142, 150, 151, 152, 155, 596, 598, 599, 600, 601 = 21	68, 70, 72, 73, 75, 76, 77, 78, 79, 101, 105, 142, 147, 150, 151, 152, 155, 596, 599, 603 = 20
	Gained hubs	68, 70, 142, 146, 147, 158, 595, 858, 997 = 9	72, 74, 75, 78, 79, 80, 81, 137, 140, 142, 145, 148, 149, 152, 153 = 15	67, 74, 77, 78, 82, 149 = 6	597, 598, 601, 997 = 4
I543V	Retained hubs	67, 71, 72, 73, 74, 75, 76, 77, 78, 79, 101, 104, 105, 150, 151, 152, 155, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 768 = 28	73, 76, 77, 101, 104, 105, 146, 147, 150, 151, 155, 596, 597, 598, 599, 600, 601, 603, 605 = 19	71, 72, 73, 75, 76, 79, 80, 81, 97, 98, 101, 102, 103, 104, 105, 137, 140, 141, 142, 145, 146, 147, 150, 151, 152, 155, 596, 597, 598, 599, 600, 601, 602, 604, 605, 768 = 36	67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 97, 101, 104, 105, 140, 141, 142, 145, 146, 147, 149, 150, 151, 152, 153, 155, 595, 596, 599, 602, 603, 605 = 36
	Gained hubs	68, 69, 70, 80, 81, 82, 85, 87, 98, 100, 102, 103, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 149, 153, 154, 156, 158, 595, 606, 769, 821, 997 = 33	67, 68, 69, 70, 71, 72, 74, 75, 78, 79, 80, 81, 82, 98, 102, 103, 106, 140, 141, 142, 145, 152, 153, 154, 156, 158, 595, 602, 604, 821, 859, 860, 862, 954, 956, 997, 998, 999 = 38	67, 68, 69, 70, 74, 77, 78, 82, 87, 100, 106, 149, 153, 154, 156, 158, 159, 595, 603, 769, 997 = 21	65, 81, 82, 98, 102, 154, 156, 158, 597, 598, 600, 601, 604, 821, 997, 998 = 16
D949V	Retained hubs	67, 71, 72, 73, 74, 75, 76, 77, 78, 79, 101, 104, 105, 150, 151, 152, 155, 596, 597, 598,	73, 76, 77, 101, 104, 105, 146, 147, 150, 151, 155, 596, 597, 598,	71, 72, 73, 75, 76, 79, 80, 81, 101, 105, 140, 142, 145, 146, 147, 150, 151, 155, 596, 597,	71, 72, 73, 75, 76, 80, 101, 104, 105, 142, 146, 155, 596,

		599, 600, 601, 602, 603, 604, 605 = 27	599, 600, 601, 605 = 18	598, 599, 601, 602, 605 = 25	599, 602, 605 = 16
	Gained hubs	65, 68, 70, 80, 81, 97, 98, 102, 137, 140, 141, 142, 145, 146, 147, 148, 149, 153, 154, 156, 158, 159, 595, 821 = 24	65, 67, 70, 71, 72, 74, 75, 78, 79, 102, 140, 142, 152, 153, 154, 158, 595, 602, 604, 997, 998, 999 = 22	67, 70, 74, 77, 78, 82, 149 = 7	81, 102, 597, 600, 768, 769 = 6

Table S3.5.4. Averaged *eigen centrality (EC)* of potential hubs found in drug free and drug bound WT systems, hubs gained due to mutation, and hubs that were maintained despite the presence of mutation.

EC		Drug free		Drug bound	
Systems		Monomer A	Monomer B	Monomer A	Monomer B
WT	Hubs	778, 808 = 2	121, 122, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 211, 214, 215, 216, 217, 218, 224, 225, 227, 248, 253, 279, 280, 281, 282, 283, 284, 469, 477, 478, 479, 480, 482, 483, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 502, 503, 732, 734, 774, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810 = 71	100, 103, 104, 106, 108, 109, 110, 111, 112, 113, 114, 115, 116, 159, 665, 666, 667, 707, 708, 710, 717, 718, 721, 724, 729, 732, 733, 734, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 790, 791, 792, 794, 796, 799, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 813, 829, 830, 832, 833, 834, 835, 836, 837, 852, 926, 928, 929 = 69	595, 707, 708, 710, 721, 732, 734, 740, 741, 744, 775, 776, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 813 = 24
C29R	Retained hubs	0	121, 122, 190, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 214, 216, 253, 279, 280, 281, 282, 283, 284, 469, 477, 478, 479, 480, 482, 483, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 502, 503 = 47	100, 103, 104, 108, 109, 110, 111, 112, 113, 115, 116, 159, 732, 734, 774, 775, 776, 777, 778, 779, 781, 782, 790, 791, 792, 796, 799, 802, 803, 804, 805, 806, 807, 808, 809, 810, 813, 829, 832, 833, 836, 837, 926, 928 = 44	741 = 1
	Gained hubs	31, 197, 198, 199, 201, 202, 282, 479, 480, 492, 493, 495, 496, 499 = 14	92, 93, 120, 124, 125, 126, 462, 464, 470, 471, 481, 484, 485, 486, 488, 489, 490, 501, 504 = 19	93, 120, 122, 155, 156, 160, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 216, 280, 281, 282, 283, 469, 477, 478, 479, 480, 481, 482, 483, 484, 486, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 502, 503 = 48	31 = 1

M166V	Retained hubs	0	0	778 = 1	734, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810 = 13
	Gained hubs	92, 93, 120, 121, 122, 126, 190, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 209, 211, 214, 216, 217, 248, 253, 280, 281, 282, 283, 284, 285, 462, 464, 469, 470, 471, 476, 477, 478, 479, 480, 481, 482, 483, 484, 486, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504 = 68	31 = 1	121, 122, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 214, 216, 280, 281, 282, 462, 464, 469, 477, 478, 479, 480, 482, 483, 486, 489, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504 = 47	774 = 1
Y186C	Retained hubs	0	198, 199, 201, 202, 280, 282, 469, 478, 479, 480, 482, 483, 492, 493, 494, 495, 496, 498, 499	707, 708, 710, 717, 718, 721, 724, 729, 732, 733, 734, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 790, 791, 792, 794, 796, 802, 804, 805, 806, 807, 808, 809, 810, 813, 926, 928 = 39	595, 710, 732, 734, 740, 741, 775, 776, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 813 = 20
	Gained hubs	92, 93, 120, 121, 122, 123, 125, 126, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 216, 280, 281, 282, 283, 284, 462, 464, 469, 471, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 502, 503, 504 = 62	31, 486 = 2	198, 595, 605, 735, 738, 740, 765, 767, 768, 769, 770, 793, 795 = 13	605, 606, 608, 738, 763, 764, 765, 766, 767, 769, 770, 771, 772, 774 = 14

S534N	Retained hubs	778, 808 = 2	734, 774, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810 = 14	0	0
	Gained hubs	593, 594, 595, 606, 738, 739, 740, 741, 742, 743, 744, 748, 750, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 776 = 24	545, 547, 549, 595, 606, 707, 708, 709, 710, 712, 717, 718, 720, 721, 724, 729, 731, 732, 733, 734, 735, 738, 740, 741, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 779, 780, 781, 782, 783, 784, 785, 790, 791, 792, 793, 794, 795, 796, 799, 800, 801, 802, 803, 804, 805, 806, 807, 809, 810, 811, 812, 813, 814, 815, 816, 833, 837, 925, 926, 927, 928, 929, 932 = 76	92, 93, 120, 121, 122, 126, 190, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 209, 214, 216, 217, 218, 248, 253, 280, 281, 282, 283, 284, 462, 464, 469, 470, 471, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504 = 67	31 = 1
I543V	Retained hubs	778 = 1	732, 734, 774, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810 = 15	0	0
	Gained hubs	192, 193, 194, 196, 197, 198, 199, 201, 202, 203, 280, 281, 282, 283, 284, 462, 464, 469, 471, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 502, 503, 741 = 42	31, 666, 667, 668, 707, 708, 709, 710, 717, 721, 724, 729, 731, 733, 735, 773, 776, 780, 791, 793, 794, 804, 806, 807, 812, 813, 814, 926, 928 = 29	31 = 1	92, 93, 120, 121, 122, 123, 125, 126, 127, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 214, 216, 217, 218, 225, 226, 280, 281, 282, 283, 284, 462, 464, 469, 470, 471, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504 = 69
D949V	Retained hubs	778 = 1	192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 214, 216, 217, 218, 280, 281, 282, 283, 284, 469, 477, 478,	734, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 790, 792, 794, 804, 805, 806, 807, 808, 809, 810, 813,	595, 734, 740, 741, 775, 776, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810 = 17

			479, 480, 482, 483, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 502, 503 = 44	926, 928 = 26	
	Gained hubs	31 = 1	92, 120, 125, 126, 462, 464, 470, 471, 481, 484, 485, 486, 488, 489, 490, 501, 504 = 17	192, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 216, 280, 282, 479, 492, 493, 495, 496, 497, 499, 500, 593, 595, 605, 606, 735, 738, 740, 765, 767, 768, 769, 770, 795 = 35	596, 606, 738, 742, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 794, 795, 816, 817 = 19

Table S3.5.5. Averaged *katz centrality (KC)* of potential hubs found in drug free and drug bound WT systems, hubs gained due to mutation, and hubs that were maintained despite the presence of mutation.

KC		**Drug free**		**Drug bound**	
Systems		**Monomer A**	**Monomer B**	**Monomer A**	**Monomer B**
WT	**Hubs**	103, 104, 108, 110, 111, 112, 113, 116, 155, 159, 198, 199, 200, 201, 202, 203, 282, 479, 496, 499, 710, 732, 734, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 832, 833	108, 110, 111, 155, 192, 193, 194, 195, 196, 198, 199, 200, 201, 202, 203, 206, 280, 282, 479, 480, 492, 493, 495, 496, 499, 500, 707, 710, 729, 732, 734, 774, 775, 776, 777, 778, 779, 781, 782, 790, 792, 805, 808, 810, 832, 833	103, 104, 108, 109, 110, 111, 112, 113, 199, 203, 313, 350, 605, 640, 641, 665, 666, 667, 707, 710, 732, 734, 775, 776, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 813, 832, 833, 836, 837, 928 = **40**	108, 110, 111, 155, 159, 595, 707, 708, 710, 732, 734, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 833 = **24**
C29R	**Retained hubs**	108, 110, 111, 112, 113, 116, 155, 198, 199, 200, 201, 202, 203, 282, 479, 496, 499, 775, 778, 779, 790, 792, 805, 808, 810, 832, 833 = **27**	108, 110, 111, 155, 198, 199, 201, 202, 203, 206, 280, 282, 479, 480, 492, 493, 495, 496, 499, 500, 710, 732, 734, 775, 777, 778, 779, 781, 782, 790, 805, 808, 810, 833 = **33**	103, 104, 108, 109, 110, 111, 112, 113, 199, 203, 732, 734, 775, 776, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 813, 832, 833, 836, 837, 928 = **30**	108, 111, 155, 159, 732, 734, 775, 777, 778, 779, 781, 782, 790, 805, 808, 810, 833 = **17**
	Gained hubs	197, 280, 469, 478, 480, 482, 492, 493, 495, 497, 500 = **11**	72, 104, 109, 112, 113, 116, 120, 122, 153, 154, 156, 158, 159, 197, 205, 469, 478, 482, 483, 486, 494, 497, 498 = **23**	100, 116, 153, 155, 156, 159, 194, 197, 198, 201, 202, 282, 478, 479, 480, 482, 486, 492, 493, 495, 496, 499, 554, 740, 774, 777, 807 = **26**	112, 740, 774, 776 = **4**
M166V	**Retained hubs**	108, 110, 111, 113, 155, 198, 199, 200, 201, 202, 203, 282, 479, 496, 499, 732, 734, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 810, 832, 833 = **30**	108, 110, 194, 198, 199, 200, 201, 202, 203, 282, 479, 480, 496, 732, 734, 774, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 810, 832, 833 = **28**	108, 110, 111, 112, 113, 199, 203, 605, 732, 734, 775, 776, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 832, 833 = **24**	108, 110, 111, 155, 595, 710, 732, 734, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 833 = **21**
	Gained hubs	197, 206, 280, 469, 478, 480, 482, 483, 492, 493, 495, 500, 503, 772, 774, 776 = **16**	595, 740, 741, 809, 813 = **5**	194, 196, 198, 200, 201, 202, 205, 206, 280, 282, 469, 478, 479, 480, 492, 493, 495, 496, 499, 500, 503, 595, 740, 769, 777 = **25**	113, 605, 767, 769, 770, 771, 774, 776, 796, 813, 829, 832 = **12**
Y186C	**Retained hubs**	103, 104, 108, 110, 111, 112, 113, 116, 155, 159, 198, 199, 200, 201, 202, 203, 282, 479, 496, 499, 775, 777, 778, 779,	108, 110, 111, 155, 198, 199, 201, 202, 280, 282, 479, 480, 492, 493, 495, 496, 499, 710, 732, 734, 774, 775, 776, 777,	104, 108, 109, 110, 111, 112, 113, 199, 203, 710, 732, 734, 775, 776, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810,	108, 110, 111, 155, 159, 595, 707, 708, 710, 732, 734, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 833 =

		781, 790, 808, 810, 832, 833 = **30**	778, 779, 781, 782, 790, 792, 805, 808, 810, 832, 833 = **35**	813, 832, 833 = **26**	**24**
	Gained hubs	100, 109, 120, 122, 123, 125, 127, 160, 194, 197, 206, 280, 469, 478, 480, 482, 483, 486, 492, 493, 494, 495, 497, 498, 500, 836 = **24**	103, 112, 113, 159, 469, 478, 482, 483, 486 = **9**	154, 155, 159, 198, 201, 282, 478, 479, 480, 482, 486, 492, 495, 496, 499, 740, 770, 772, 774, 777 = **20**	100, 103, 104, 112, 113, 116, 138, 153, 156, 606, 608, 740, 765, 767, 770, 813, 832 = **17**
S534N	**Retained hubs**	108, 110, 155, 710, 732, 734, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 833 = **19**	108, 110, 111, 155, 194, 198, 199, 200, 201, 202, 203, 282, 479, 480, 734, 775, 776, 777, 778, 779, 781, 790, 805, 808, 810, 833 = **26**	103, 108, 110, 111, 112, 113, 199, 203, 732, 734, 775, 776, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 832, 833, 836 = **25**	108, 110, 111, 732, 734, 775, 778, 779, 781, 782, 790, 792, 805, 808, 810, 833 = **16**
	Gained hubs	740, 770, 771, 772, 773, 776, 804, 807, 813, 928 = **10**	112, 113, 138, 153, 159, 478, 482, 606, 740, 765, 769, 809 = **12**	116, 122, 155, 159, 194, 197, 198, 200, 201, 202, 205, 206, 280, 282, 469, 477, 478, 479, 480, 482, 483, 492, 493, 494, 495, 496, 498, 499, 500, 721, 777 = **31**	113, 482, 499, 813, 832 = **5**
I543V	**Retained hubs**	108, 110, 111, 112, 198, 199, 202, 203, 282, 479, 496, 499, 734, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 810, 832, 833 = **26**	108, 110, 111, 194, 195, 196, 199, 707, 710, 732, 734, 774, 775, 776, 777, 778, 779, 781, 782, 790, 792, 805, 808, 810, 832, 833 = **26**	108, 110, 111, 112, 113, 199, 203, 666, 710, 732, 734, 775, 778, 779, 781, 782, 790, 808, 810, 833 = **20**	108, 110, 111, 155, 159, 734, 775, 777, 778, 779, 781, 782, 790, 805, 808, 810, 833 = **17**
	Gained hubs	469, 478, 480, 482, 486, 492, 495, 498, 772, 776, 813 = **11**	103, 112, 225, 640, 641, 666, 667, 708, 709, 733, 809 = **11**	122, 155, 159, 194, 197, 198, 200, 201, 202, 282, 479, 480, 482, 492, 493, 495, 496, 499, 500, 554, 777 = **20**	112, 113, 122, 127, 196, 197, 198, 199, 200, 201, 202, 203, 280, 282, 469, 478, 479, 480, 482, 483, 486, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 774, 832 = **33**
D949V	**Retained hubs**	108, 110, 111, 112, 155, 710, 732, 734, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 832, 833 = **22**	108, 155, 192, 194, 198, 199, 201, 202, 203, 280, 282, 479, 480, 492, 493, 495, 496, 499, 500, 732, 734, 775, 778, 779, 781, 782, 790, 792, 805, 808, 810 = **31**	108, 110, 111, 199, 203, 605, 734, 775, 776, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 832, 833, 928 = **22**	108, 110, 111, 155, 595, 732, 734, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 833 = **20**
	Gained hubs	554, 666, 740, 774, 776 = **5**	197, 469, 478, 482, 483, 486, 549, 554, 572, 793, 809, 813, 814 = **13**	194, 196, 198, 200, 201, 202, 282, 479, 493, 495, 496, 499, 595, 606, 740, 765, 769, 770, 772, 773, 774, 777 = **22**	202, 282, 478, 479, 480, 482, 486, 492, 495, 496, 499, 596, 606, 740, 741, 765, 766, 767, 769, 770, 772, 774, 776, 795, 816, 832 = **26**

Table S3.6.1 A representation of residue interaction within the NADPH in Monomer A and Monomer B of activated (drug bound) WT and mutants as extracted from contact-heat-map. In bold are the target residue and in brackets are all contacting residues.

Systems	**Drug bound**			
Subunits	**Monomer A**		**Monomer B**	
WT	**A340:A-**(G339A, G341A, A344A, I348A, F363A, I370A) **T343:A-**(A340A, G341A, D342A, A344A, D346A, C347) **K365:A-**(F363A,NR364A, G366A, N369A, P389A, F390A, T411A, T419A) **R371:A-**(N48B, G341A, I370A, A372A) **G484:A-**(L461A, G480A, D481A, V482A, V483A, L485A, A486A) **A486:A-**(G480A, G481A, G484A, L485A, N488A, G491A, S492A)		**A340:B-**(G339B, G341B, A344B, I348B, A437B, F438B, G439B) **T343:B-**(A340B, G341B, D342B, A337B) **K365:B-**(F363B, R364B, G366B, F390B, L391B, L418B) **R371:B-**(G341B, D345B, I370B) **G484:B-**(S443B, L461B, G480B, D481B, V482B, V483B, L485B, A486B) **A486:B-**(G480A, G481A, G484A, L485A, N488A, G491A, S492A)	
	Gained residues connections	**Lost residues connections**	**Gained residues connections**	**Lost residues connections**
C29R	**A340-(**D342, R371, E421) **T343-**(S306, A437) **L365-**(I370) **R371-**(A340, I370) **G484-**(N487) **A486-**(T31, G479, N494, D495) = **13**	**A340-(**A344, I341, F363, I370) **T343-**(R235) **R371-**(N48) **G484-**(I461) **A486-**(D481) = **8**	**A340-(**V337, F345, A348, I361, F363, A348, A437) **R371-**(F390, T411, W419, A344) **G484-**(N487) **A486-**(D481, N494, D495) **= 15**	**T343-**(D346, C347) **R364-**(T419) **L365-**(I370) **R371-**(D134) **A486-**(G479) = **6**
M166V	**A340-**(R371, A372) **T343-**(A437) **R364-**(P393) **L365-**(I370) **R371-**(A340, I370) **A486-**(T31, D495) = **9**	**A340-**(I361,F363) **T343-**(R235) **R371-**(N48) **G484-**(I461) = **5**	**A340-**(F345, I361, F363, I370, A372) **T343-**(N487) **R371-**(F390, T411, W419) **A486-**(T31, D481, N494, D495) = **13**	**A340-**(F438, G439) **T343-**(D346, C347) **L365-**(I370) **R371-**(D134, G341) **A486-**(G479) **= 8**
Y186C	**T343-**(G399, A340, A347) **R364-**(V362) **R371-**(L47) **G484-**(N487) **A486-**(G479) **= 7**	**T343-**(R235) **R371-**(N48, G341) **= 3**	**A340-**(L338, I361, F363) **T343-**(F345, N481) **R371-**(F390, T411, W419, V373, E376) **G484-**(L285, G439, S440) **A486-**(D481) **= 14**	**A340-**(A344, S436, F438, G439) **T343-**(C347) **L365-**(I370) **R371-**(I370, D134, G341) **G484-**(G480, V482) **A486-**(G480) **= 13**
S534N	**A340-**(T343, A372, A437)	**T343-**(R235)	**A340-**(I361, F363, I370, A372)	**A340-**(F438, G439)

	T343-(S306, L340, A437) **R371-**(340) **A486-**(T31, G479, N494, D495) = **11**	= **1**	**T343-**(F345, D346, C347, A372, V373, G376) **R371-**(L461) **A486-**(D481, N494, D495) = **14**	**T343-**(C347) **L365-**(I370) **R371-**(I370, D134) = **6**
I543V	**A340-**(T343, I361,M377) **T343-**(A437) **R364-**(F340) **L365-**(I370) **R371-**(D46, L47, I370) **G484-**(N487) = **10**	**A340-**(I361,I370) **T343-**(R235) **L365-**(F390) **R371-**(G341) **G484-**(I461) = **6**	**A340-**(I361, I370, A372) **T343-**(E320, D346) **R371-**(F390, T411, W419) **G484-**(L461, N487) **A486-**(T31, D481, N494, D495) = **14**	**A340-**(F438, G439) **L365-**(I370) **R371-**(I370, D134, G341, D342) = **7**
D949V	**T343-**(A437) **R371-**(A340) **A486-**(T31, D495) = **4**	**T343-**(R235) **G484-**(I461) = **2**	**A340-**(V337, L338, T342, T343 A372) **T343-**(S306, D346, C347) **R371-**(A340) **G484-**(S440, L461, N487) **A486-**(T31, D481, N494, D495) = **16**	**A340-**(F438, G439) **L365-**(I370) **R371-**(I370, D134) **A486-**(G479) = **6**

Table S3.6.2 A representation of residue interaction within the FAD in Monomer A and Monomer B of activated (drug bound) WT and mutants as extracted from contact-heat-map. In bold are the target residue and in brackets are all contacting residues.

Systems	**Drug bound**			
Subunits	**Monomer A**		**Monomer B**	
WT	**A198:A-**(L196, P197, S199, S201, C202, G828, G479, S492, V493, G496) **V129:A-**(T125, C126, G127, V129, C130, L236) **L219:A-**(E218, Q220, G258, S260, L261) **L226:A-**(G225, S227, E230, I231) **R235:A-**(T231, P232, F234, I236, T343, D346, C347) **L261:A-**(F193, L219, S260, S262, M266, T267,V447, A450) **D481:A-**(I283, G479, G480, V483, G484, A486, T489, V493) **R489:A-**(D481, T488, V490, S492, V493)		**A198:B-**(P197, G198, S199, S201, C202, S492, G496, L497, A499) **V129:B-**(C126, G127, M128, C130, I231, R235, L236, T488, V490, E491) **L219:B-**(K218, Q220, G258, L259, S260, L285) **L226:B-**(A195, G196, P197, G224, G225, S227, S229, E230) **R235:B-**(D61, M128, V129, I231, P232, F234, L234) **L261:B-**(F193, S260, S262, M266, T267, L268, I281, I283, V447, A450, K451) **D481:B-**(G282, I283, G284, P286, V441, S443, D444, G480, V482, G484) **T489:B-**(L122, C126, G127, M128, P197, T488, V490, S492, D493)	
	Gained residues connections	**Lost residues connections**	**Gained residues connections**	**Lost residues connections**
C29R	**A198-**(L192, F280) **K219-**(I255, I256) **R235-**(G127) **L261-**(F217, G218, L268, I283) **D481-**(N487) **T486-**(C126, V493) **= 12**	**A198-**(G479, S492, V493) **V129-**(T125) **K219-**(G258, L261) **R235-**(T343, C347) **L261-**(L219, T267) **D481-**(G479, T489) **= 12**	**A198-**(L192, F193, G194, F280, G282, I283, E491, G480, E491) **K219-**(G194) **L226-**(C126, I231) **R235-**(I62, T65, D346) **L261-**(F217) **D481-**(K285, A486) **= 18**	**A198-**(S201, C202, V493, L497) **K219-**(L259) **L226-**(A195) **R235-**(D61) **L261-**(R267) **D481-**(G282, I283) **T489-**(G127) **= 11**
M166V	**K219-**(L259) **R235-**(G127) **L261-**(L268, L451) **D481-**(L461) **T486-**(L122, T125, C126, V493) **= 9**	**V129-**(T125, C126) **K219-**(L261) **R235-**(T343, D346, C343) **L261-**(L219) **D481-**(G479, A486, T489) **= 10**	**A198-**(G282, G479, G480, E491) **V129-**(T125) **K219-**(G225) **L226-**(I231) **R235-**(I62, T65, D346) **D481-**(K285, A486) **= 12**	**A198-**(S201, C202, V493, L497) **R235-**(D61, M128, V129) **L261-**(I283) **481-**(G282, I283) **D T489-**(G127) **= 11**
Y186C	**V129-**(A235) **K219-**(I194, S229) **L226-**(A195, G196, S229) **R235-**(I62)	**V129-**(C126, L236) **K219-**(G258, L261) **R235-**(T343, C347) **L261-**(L219)	**A198-**(G194, G282, G479, G480, E491) **K219-**(G194) **L226-**(P197)	**A198-**(V493, L497) **V129-**(L226, S227) **L226-**(G244) **R235-**(D61, M128)

	L261-(L268, L451) **D481-**(N487) **T486-**(C126, V493) **=12**	**D481-**(G479, T489, S492) **= 10**	**D481-**(K285, V483, L485, A486) **T486-**(L226, S227) **= 13**	**L261-**(R267) **D481-**(G282, I283, N487) **D T489-**(L122, V129) **= 13**
S534N	**A198-**(L192, G225) **V129-**(A235, T488) **K219-**(L259) **L226-**(S229) **R235-**(I62, V129) **L261-**(I283, L451) **D481-**(L461) **T486-**(D342) **=12**	**A198-**(G479, S492, V493) **V129-**(T125, C126) **R235-**(T343, C347) **L261-**(L219) **D481-**(G479) **T489-**(P197) **= 10**	**A198-**(G194, A195 G282, G479, G480, E491) **K219-**(G194) **L226-**(C126, G194, P197, I231) **R235-**(I62, T65, D346) **L261-**(F217) **D481-**(K285, A486) **= 17**	**A198-**(L494, A499) **K219-**(L259) **L226-**(G244, S229) **R235-**(V129) **D481-**(G282, I283) **D T489-**(L122, G127, V129) **= 11**
I543V	**A198-**(L192, F280) **V129-**(T489) **K219-**(I255, I256) **R235-**(I62, V129) **L261-**(F217, E218, I283) **D481-**(N487) **T486-**(C126, G127, V129, C130, V493) = **N16**	**A198-**(G479, S492, V493) **V129-**(T125) **K219-**(G258, L261) **R235-**(T343, C347) **L261-**(L219, T267) **D481-**(G479, T489) **= 12**	**A198-**(G194, G282, G479, G480, E491) **L226-**(C126, I231) **R235-**(T65, D346) **L261-**(F217) **D481-**(K285, A486) **T486-**(G196) **= 13**	**A198-**(V493, L497) **V129-**(G196) **K219-**(L259) **D481-**(G282, I283) **D T489-**(L122, V493) **= 8**
D949V	**A198-**(G194, I200, G225) **V129-**(A235, T490) **K219-**(L259) **L226-**(G196, S229) **R235-**(I62, V129) **L261-**(I283, L451) **D481-**(L461) **T486-**(C126, D342) **= 15**	**A198-**(S492, V493) **V129-**(T125, C126) **R235-**(T343, C347) **L261-**(L219) **D481-**(G479) **T489-**(P197) **= 9**	**A198-**(G282, G479, G480, E491, D495) **V129-**(I62, K63) **K219-**(L261) **L226-**(C126, P197, I231, T489) **R235-**(I62, T65, D346) **D481-**(K285, A486) **T486-**(L226) **= 18**	**A198-**(L494, A499) **V129-**(L226) **L226-**(A195, G244, S229) **R235-**(D61, V129) **D481-**(G282, I283) **= 10**

Table S3.6.3 A representation of residue interaction within the FMN in Monomer A and Monomer B of activated (drug bound) WT and mutants as extracted from contact-heat-map. In bold are the target residue and in brackets are all contacting residues.

Systems	**Drug bound**			
Subunits	**Monomer A**		**Monomer B**	
WT	**K574:A-**(S550, A551, A554, T555, T573, T575, F567, I613, S614, S640) **T575:A-**(S550, L574, F576, I613, S614, S640, I641, M642, N668) **G767:A-**(E666, I667, N668, I669, F707, A708, I710, T733, N736) **T793:A-**(A549, T733, T735, A792, G794, I796, L813, Q814, C816) **C816:A-**(A549, A551, G767, T768, I770, R771, A774, G794, G795, I796, G814, V815, S817) **C817:A-**(A551, F607, S766, T768, C816, A818, E820, N821)		**K574:B-**(S550, A554, T555, T573, T575, S577, S614, K616, S640) **T575:B-**(S550, K574, F576, S614, S640, I641, M642, N668) **G767:B-**(S606, F607, S766, T768, A769, I770, G794, G796) **T793:B-**(A549, T733, A792, G794, I797, G814, C816) **C816:B-**(A549, A551, S766, G767, T768, I770, T793, G794, G795, I796, Q814, V815, S817) **C817:B-**(A551, F607, S766, G767, C816, A818, I819, K820)	
	Gained residues connections	**Lost residues connections**	**Gained residues connections**	**Lost residues connections**
C29R	**G767-**(R771) **= 1**	**G767-**(S817) **S817-**(F607, S766, G767) **= 4**	**L574-**(A551) **T575-**(613) **G767-**(S605) **C816-**(S801) **=3**	**K709-**(S670) **= 1**
M166V	**G767-**(R771) **T793-**(A734) **G795-**(S817) **C817-**(F607) **= 4**	**T575-**(I613) **G767-**(C816, S817) **T793-**(L813) **S817-**(F607, S766, G767) **= 7**	**T575-**(I613) **G767-**(S605) **C816-**(S801) **= 3**	**L574-**(K616) **K709-**(L669, S670) **= 3**
Y186C	**G767-**(R771) **T793-**(A734) **G795-**(S817) **C817-**(Q604) **= 4**	**T575-**(I613) **G767-**(C816) **793-**(I796, L813) **C816-**(G676) **S817-**(F607, S766) **=7**	**T575-**(I613) **G767-**(S605) **C816-**(S801) **= 3**	**L574-**(K616) **K709-**(S670) **S817-**(G767) **= 3**
S534N	**K709-**(S670) **T793-**(S817) **G795-**(S817) **C817-**(F607**) = 4**	**T575-**(I613) **G767-**(C816) **793-**(I796, L813) **= 4**	**T575-**(I613) **G767-**(S605) **C816-**(S801) **= 3**	**K709-**(L669, S670) **= 2**
I543V	**K574-**(K616) **K709-**(T793) **G767-**(R771)	**K574-**(S614) **G767-**(C816) **793-**(I796, L813)	**T575-**(I613) **G767-**(S605, S817) **= 4**	**K709-**(S670) **K709-**(L669, S670) **G767-**(F607, I796)

	T793-(K709, A734) **G795-**(S817) = **6**	**C816-**(G676) **= 5**		**= 5**
D949V	**K709-**(S670, F707) **T793-**(V815) **G795-**(S817) **C817-**(F607) **= 5**	**G767-**(C816) **T793-**(L813) **C816-**(S766, G676) **S817-**(V821) **= 5**	**L574-**(A551) **T575-**(I613) **G767-**(S605) **G C816-**(S801) **= 4**	**L574-**(K616) **C816-**(S766, A511) **= 3**

Table S3.6.4 A representation of residue interaction within the 5-FU in Monomer A and Monomer B of activated (drug bound) WT and mutants as extracted from contact-heat-map. In bold are the target residue and in brackets are all contacting residues

Systems	**Drug bound**			
Subunits	**Monomer A**		**Monomer B**	
WT	**N609:A-**(R589, I590, F607, L608, I610, E611, G763, G764) **E611:A-**(T552, R589, I590, N609, I610, L612) **L612:A-**(N685, K589, I590, A609, I610, E611, P672, H673, F935) **N668:A-**(T575, S640, I641, M642, E666, L667, L669, K709) **S670:A-**(L669, C671, A683, C684, G685, Q686) **S671:A-**(S670, P672, H673, G674, A683) **N736:A-**(S670, L709, L710, T711, P712, T735, T737) **T737:A-**(A736, V738, G764, V765, S766)		**N609:B-**(R589, I590, F607, L608, I610, E611, G736, G764) **E611:B-**(T552, R589, I590, A609, I610, L612) **L612:B-**(N685, R589, I590, A609, I610, E611, I613, P672, H673, F935) **N668:B-**(T575, S640, I641, M642, E666, L667, L669, K709) **S670:B-**(L669, C671, A683, C684, G685, Q686) **S671:B-**(S670, P672, H673, G674, A683) **N736:B-**(S670, K709, L710, T711, P712, T735, T737) **T737:B-**(N736, V738, G764, V765, S766)	
	Gained residues connections	Lost residues connections	Gained residues connections	Lost residues connections
C29R	**N609-**(T552) **L612-**(I610, F935) **N668-**(S670) **S670-**(M642, N668)	**L612-**(H673) **S670-**(G685)	**N609-**(T552) **E611-**(N585) **L612-**(I610, F935) **L612-**(I610, F935) **S670-**(M642, N668) **C671-**(M642, N668, T711)	**L612-**(H673) **N609-**(S670)
M166V	**E611-**(N585) **N668-**(S670) **S670-**(M642, N668)	**E611-**(T552, C671) **L612-**(H673)	**N668-**(S670) **S670-**(M642, N668) **C671-**(M642, N668, N736)	**E611-**(T552) **L612-**(V583, P672, H673) **N609-**(S670)
Y186C	**E611-**(N585) **L612-**(I610, F935) **N668-**(S670) **S670-**(N668)	**L612-**(H673) **S670-**(C684, G686)	**E611-**(C989) **L612-**(I610, F935) **N668-**(S670) **C671-**(N668, T711)	**L612-**(V583, N585, H673)
S534N	**E611-**(N585) **L612-**(F935) **N668-**(S670) **S670-**(N668)	**E611-**(T552, C671) **L612-**(H673) **S670-**(C684, G686)	**E611-**(N585) **L612-**(I610, F935) **N668-**(S550, S670, C671) **S670-**(N668, K709, L710) **C671-**(N668, K709, L710, T711)	**E611-**(I590) **L612-**(P672, H673)
I543V	**E611-**(N585)	**N609-**(S670)	**E611-**(N585, C989)	**N609-**(F607)

	L612-(F935) **N668-**(S670) **S670-**(N668)	**E611-**(C671) **L612-**(H673) **S670-**(A683, G685)	**N668-**(S670) **S670-**(M642, N668) **C671-**(M642, N668, T711, N736)	**E611-**(T552) **L612-**(N585, P672, H673)
D949V	**E611-**(N585) **L612-**(I610, F935) **N668-**(S670) **S670-**(M642, N668)	**N609-**(S670) **E611-**(T552, C671) **L612-**(H673) **S670-**(A683, G685)	**E611-**(N585) **N668-**(S670, C671) **S670-**(N668, K709, L710) **C671-**(M642, N668, K709, L710, T711, N736)	**N609-**(F607) **L612-**(P672, H673)

Table S4.1. DPD missense mutation associated with people of African population as identified by 1000 genomes and Exom sequencing project (ESP MAF), Variants were classified as rare or common based on their minor allele frequency (MAF) of 0.001. Prediction of the effect of 12 missense mutations on DPD structure and function using VAPOR web servers

Variant	SNP ID	DPD activity ± SD	P value	Exom sequencing project (ESP MAF)		Functional impact			Stability change	Reference
						HUMA-VAPOR				
				African American	African	Polyphen-2	Provean	PhD-SNP	MUpro	PHARMGKB
S201R	rs72549308	0.01 ±0.04	3.0×10^{-7}	na	na	Probably damaging	Disease	Neutral	Decreases	Doi: 10.1158/0008-5472.CAN-13-2482
K259E	rs45589337	0.93 ±0.19	0.63	0.00114	0	Benign	Disease	Neutral	Decreases	Doi:10.1016/j.ejca.2018.07.009
D342N	rs183385770	0.17 ±0.03	1.9 × 10-6	0.0020	na	probably damaging	Disease	Neutral	Decreases	Doi: 10.1158/0008-5472.CAN-13-2482
D432N	rs142512579	1.22 ±0.11	0.26	0.00023	na	Benign	Disease	Neutral	Decreases	Doi: 10.1158/0008-5472.CAN-13-2482
S492L	rs72549304	0.18 ±0.04	2.3×10^{-4}	na	na	Probably damaging	Tolerated	Disease	Increases	Doi: 10.1158/0008-5472.CAN-13-2482
R886C	rs147545709	0.99 ±0.12	0.95	0.00045	0	Probably damaging	Tolerated	Disease	Decreases	Doi: 10.1158/0008-5472.CAN-13-2482
L993R	rs139459586	0.94 ±0.11	0.51	0.00113	0.0041	Probably damaging	Tolerated	Disease	Decreases	Doi: 10.1158/0008-5472.CAN-13-2482

Table S4.2. Estimated center of mass (COM) distances (Å) from the location of the missense mutant to either the nearest cofactors or the substrate. The calculations were based on the modeled native human DPD 3D structure.

Variant	Location (Domain)	COM distance to the nearest co-factor (Å)		COM distance to the substrate (Å)	
		Monomer A	Monomer B	Monomer A	Monomer B
S201R	FAD binding domain (II)	17.7	15.2	64.3	64.7
K259E	FAD binding domain (II)	15.5	14.4	76.4	78.6
D342N	NADPH binding domain (III)	5.8	5.4	78.3	.65.4
D432N	NADPH binding domain (III)	19.0	21.6	72.7	63.7
S492L	FAD binding domain (II)	9.3	9.8	60.5	62.5
R886C	4^{th} Fe-S cluster binding domain (V)	31.7	32.9	47.8	48.3
L993R	4^{th} Fe-S cluster binding domain (V)	10.2	8.4	19.5	20.0

Table S4.3. Validation scores from the ERRAT, ProSA, and PROCHECK webservers for 13 drug free and drug bound modeled human DPD 3D structures. Both the drug free and drug bound models received the same score.

Protein	**ERRAT**	**ProSA**		**PROCHECK Model 2**		
	Overall quality factor (%)	**Monomers z-Score**		**Ramachandran (residues location) (%)**		
		A	B	Favored	Allowed	Disallowed
Template (PDB ID: 1H7X)	85.79	-13.56	-13.47	93.5	6.4	0.1
Wild type	88.86	-14.53	-14.27	89.1	10.7	0.2
S201R	88.86	-14.53	-14.27	89.1	10.7	0.2
K259E	88.86	-14.53	-14.27	89.1	10.7	0.2
D342N	88.86	-14.53	-14.27	89.1	10.7	0.2
D432N	88.86	-14.53	-14.27	89.1	10.7	0.2
S492L	88.86	-14.53	-14.27	89.1	10.7	0.2
R886C	88.86	-14.53	-14.27	89.1	10.7	0.2
L993R	88.86	-14.53	-14.27	89.1	10.7	0.2

Table S4.4.1. Monomer A drug free Rg calculation including mean, standard deviation (std), median, upper quartile, lower quartile and inter-quartile range (IQR)

	WT	S201R	K259E	D342N	D432N	S492L	R886C	L993R
Count (nm)	60001	60001	60001	60001	60001	60001	60001	60001
Mean	3.458547	3.48827	3.481433	3.472513	3.458602	3.470903	3.503307	3.490343
std	0.009235	0.022534	0.021647	0.020118	0.008827	0.016636	0.019154	0.025978
Min	3.37021	3.37629	3.35675	3.37447	3.38976	3.3792	3.37597	3.38306
25%	3.45335	3.47715	3.47457	3.46323	3.45335	3.46115	3.49774	3.48154
50%	3.45873	3.48856	3.48564	3.47569	3.45873	3.4692	3.50677	3.49293
75%	3.46415	3.50332	3.49549	3.48587	3.46415	3.48149	3.51431	3.50466
IQR	0.01080	0.02617	0.02092	0.02354	0.0108	0.02034	0.01657	0.02312
Max	3.49143	3.54589	3.53139	3.52905	3.49143	3.52824	3.54925	3.54988

Table S4.4.2. Monomer B drug free Rg calculation including mean, standard deviation (std), median, upper quartile, lower quartile and inter-quartile range (IQR)

	WT	S201R	K259E	D342N	D432N	S492L	R886C	L993R
Count (nm)	60001	60001	60001	60001	60001	60001	60001	60001
Mean	3.47683	3.494551	3.46336	3.505296	3.481653	3.457184	3.49508	3.498773
std	0.015793	0.026024	0.011669	0.029508	0.018163	0.008834	0.020954	0.016073
Min	3.37285	3.37724	3.38499	3.37537	3.36735	3.38463	3.38093	3.39459
25%	3.47321	3.48144	3.45777	3.50072	3.47044	3.45263	3.48343	3.48938
50%	3.47957	3.49629	3.46535	3.51167	3.48151	3.45752	3.49657	3.49896
75%	3.48548	3.51403	3.47111	3.52105	3.49532	3.46253	3.5088	3.50972
IQR	0.01227	0.03259	0.01334	0.02033	0.02488	0.0099	0.05617	0.02034
Max	3.5139	3.5568	3.49657	3.58046	3.54037	3.4891	3.55566	3.54355

Table S4.4.3. Monomer A drug bound Rg calculation including mean, standard deviation (std), median, upper quartile, lower quartile and inter-quartile range (IQR)

	WT	S201R	K259E	D342N	D432N	S492L	R886C	L993R
Count (nm)	60001	60001	60001	60001	60001	60001	60001	60001
Mean	3.437466	3.488114	3.47683	3.48063	3.460421	3.471856	3.459176	3.479314
std	0.011474	0.022117	0.015793	0.0228	0.01201	0.012913	0.013518	0.020331
min	3.37241	3.37181	3.37285	3.37046	3.39378	3.37634	3.38707	3.37055
25%	3.43024	3.48413	3.47321	3.47627	3.45187	3.46413	3.45184	3.47065
50%	3.43688	3.49531	3.47957	3.48673	3.46077	3.47217	3.46038	3.48328
75%	3.44516	3.50175	3.48548	3.49472	3.46917	3.48012	3.46862	3.49219
IQR	0.01492	0.01762	0.01222	0.01845	0.0173	0.01599	0.01678	0.02154
Max	3.48157	3.53729	3.5139	3.52555	3.49918	3.52106	3.50161	3.53062

Table S4.4.4. Monomer B drug bound Rg calculation including mean, standard deviation (std), median, upper quartile, lower quartile and inter-quartile range (IQR)

	WT	S201R	K259E	D342N	D432N	S492L	R886C	L993R
Count (nm)	60001	60001	60001	60001	60001	60001	60001	60001
Mean	3.440881	3.507418	3.485343	3.484837	3.494781	3.498371	3.458602	3.498773
std	0.011206	0.021664	0.019188	0.022969	0.014771	0.01537	0.008827	0.016073
Min	3.37899	3.38327	3.37254	3.37912	3.38786	3.38092	3.38976	3.39459
25%	3.4345	3.50262	3.47583	3.47412	3.48982	3.4886	3.45335	3.48938
50%	3.44169	3.51265	3.48773	3.49038	3.49655	3.49973	3.45873	3.49896
75%	3.44829	3.5201	3.49926	3.50045	3.50326	3.50927	3.46415	3.50972
IQR	0.01379	0.01748	0.02343	0.02633	0.00353	0.02067	0.0108	0.02034
Max	3.48117	3.54815	3.52733	3.54707	3.53442	3.54451	3.49143	3.54355

Table S4.5.1. Representation of Drug bound (Complex) systems indicating equilibrated areas where samples were extracted.

System	Monomer A				Monomer B			
	Drug exit (Y/N)	**Exit time (ns)**	**Exit frame**	**DRN calculation Monomer A frame range**	**Drug exit (Y/N)**	**Exit time (ns)**	**Exit frame**	**DRN calculation monomer B frame range**
WT	N	N/A	N/A	55000-58000	N	N/A	N/A	55000-58000
S201R	N	N/A	N/A	44000-47000	N	N/A	N/A	44000-47000
K259E	N	N/A	N/A	53000-56000	N	N/A	N/A	53000-56000
D342N	N	N/A	N/A	48000-51000	N	N/A	N/A	48000-51000
D432N	N	N/A	N/A	45000-48000	N	N/A	N/A	45000-48000
S492L	N	N/A	N/A	25000-28000	N	N/A	N/A	25000-28000
R886C	N	N/A	N/A	53000-56000	N	N/A	N/A	53000-56000
L993R	N	N/A	N/A	480000-51000	N	N/A	N/A	480000-51000

Table S4.5.2. Illustration of equilibrated areas of Drug free (Holo) systems where samples were obtained for follow-up calculations

System	Monomer A	Monomer B
	DRN calculation Monomer A frame range	**DRN calculation monomer B frame range**
WT	55000-58000	55000-58000
S201R	44000-47000	44000-47000
K259E	53000-56000	53000-56000
D342N	48000-51000	48000-51000
D432N	45000-48000	45000-48000
S492L	25000-28000	25000-28000
R886C	53000-56000	53000-56000
L993R	480000-51000	480000-51000

Table S4.6.1. *Betweenness centrality (BC)* hubs found in drug free and drug bound systems; hubs gained due to mutation, and hubs that were maintained despite the presence of mutation.

BC		**Drug free**		**Drug bound**	
Systems		**Monomer A**	**Monomer B**	**Monomer A**	**Monomer B**
WT	**Hubs**	38, **61**, 64, 66, 67, 71, 73, 74, 75, 76, 77, **80,** 95, 97, **101, 102,** 104, **108**, **116**, **117,** 120, **121**, **132, 133**, 136, 137, **142,** 145, 146, 152, 153, **155**, 235, 350, 370, **372, 373,** 386, 493, **597, 598, 599**, 603, **768, 775**, **799**, 802, 803, 826, 862 **= 50**	**62,** 63, 65, 66, 86, 90, **101, 102**, 105, 108, 114, 117, 121, **127**, 130, **131, 132**, 133, **137**, 139, **142,** 143, 144, 147, **149, 155**, 160, 361, 367, **370, 372, 373**, 376, 599, 601, **603,** 604, 816, 818, 826, 955, 991 **= 42**	**61,** 65, **80**, 85, 97, **101, 102**, **108**, 114**, 116, 117**, **121**, 122, 123, 125, 126, 130, **132, 133**, 137, 140, **142**, **155**, 200, 227, 231, 234, 310, **372**, **373**, 374, 375, **597**, **598, 599**, 604, 605, **768**, **775**, 776, 797, **799**, 803, 804, 816, 825, 958, 962, 996, 997 **= 50**	22, 44, **62**, 69, 76, 77, 92, 95, 97, **101, 102**, 113, 116, 120, 124, 125, **127**, **131, 133,** 135, **137,** 138, 141, **142,** 145, **149, 155,** 228, 231, 234, **370, 372, 373**, 374, 380, 383, 595, 602, **603,** 605, 775, 799, 822, 853, 856 **= 45**
S201R	**Retained hubs**	64, 67, 73, 77, 101, 102, 108, 116, 120, 132, 137, 146, 153, **155**, 373, 599, 775, 894 = **18**	108, 114, 117, 130, 131, 133, 137, 142, 147, **155**, 372, 373, 599 = **13**	61, 85, 108, 123, 132, 137, 140, 142, **155,** 310, 366, 372, 373, 375, 599, 604, 996, 997 = **18**	76, 101, 124, 125, 138, 142, 145, **155,** 372, 373, 383, 599, 605 = **13**
	Gained hubs	70, 123, 125, 138, 140, 147, 156, 157, 158, 201, 374, 375, 591, 595, 596, 771, 772, 797, 853, 995, 997 = **21**	74, 75, 76, 78, 80, 123, 125, 146, 150, 153, 156, 380, 381, 596, 767, 770, 817, 822, 958, 962, 994 = **21**	44, 45, 67, 70, 71, 74, 75, 131, 145, 146, 147, 370, 379, 382, 601, 606, 765, 767, 829 = **19**	65, 68, 70, 74, 75, 79, 80, 81, 108, 113, 117, 121, 123, 132, 147, 150, 151, 157, 158, 345, 375, 376, 382, 596, 599, 604, 768, 771, 772, 819, 826, 832, 958, 995 = **34**
K259E	**Retained hubs**	75, 77, 97, 108, 116, 121, 132, 153, 155, 373, 597, 605, 799, 802, 803 = **15**	65, 66, 102, 108, 113, 127, 131, 132, 142, 155, 160, 370, 373, 376 = **14**	80, 108, 116, 120, 123, 125, 130, 133, 140, **155**, 372, 373, 375, 599, 604, 768, 797, 799, 958, 962, 997 = **21**	76, 77, 93, 97, 101, 113, 116, 124, 125, 133, 135, 137, 138, 141, 142, 155, 372, 373, 374, 803 = **20**
	Gained hubs	63, 82, 85, 128, 130, 138, 141, 157, 375, 381, 384, 493, 596, 829, 995 = **15**	43, 67, 76, 77, 81, 116, 120, 122, 124, 141, 146, 150, 152, 156, 157, 159, 164, 235, 236, 243, 346, 371, 375, 386, 493, 553, 591, 598, 819, 855, 958, 962 = **32**	25, 63, 66, 68, 70, 74, 75, 77, 82, 91, 100, 112, 124, 131, 135, 138, 141, 145, 153, 157, 158, 232, 345, 367, 379, 380, 381, 493, 595, 998 = **30**	63, 65, 68, 70, 76, 91, 93, 110, 117, 126, 128, 129, 146, 147, 150, 157, 160, 375, 376, 493, 528, 599, 604, 803, 818, 958, 962, 997 = **28**
D342N	**Retained hubs**	61, 67, 75, 77, 80, 101, 102, 104, 116, 120, 132, 137, 142, 145, 146, **155**, 373, 493, 599, 605, 775, 799, 803 = **23**	66, 102, 108, 142, 144, 147, **155,** 370, 372, 373, 599, 604, 816 = **13**	101, 108, 116, 130, 132, 140, 141, 142, 153, **155**, 375, 597, 598, 599, 958 = **15**	97, 101, 102, 113, 116, 125, 135, 138, **155**, 372, 775, 799, 962 = **13**

	Gained hubs	72, 83, 91, 130, 138, 140, 143, 150, 368, 381, 591, 595, 601, 604, 804, 863, 958, 962 = **18**	44, 46, 47, 60, 72, 82, 85, 103, 116, 120, 122, 123, 125, 129, 132, 141, 150, 157, 374, 528, 553, 598, 765, 775, 797, 853, 863, 958, 995 = **29**	44, 45, 46, 68, 70, 76, 77, 82, 85, 95, 104, 120, 138, 146, 150, 159, 382, 493, 553, 797 = **20**	22, 25, 38, 43, 44, 45, 63, 65, 74, 75, 122, 140, 146, 151, 164, 528, 596, 598, 600, 767, 803, 804, 826, 958 = **24**
D432N	**Retained hubs**	61, 67, 97, 101, 102, 108, 116, 117, 120, 132, 142, 146, 153, **155,** 372, 373, 386, 493, 768, 799, 826 = **21**	63, 65, 117, 130, 131, 132, 133, 137, 142, 147, **155**, 370, 372, 373, 599, 604 = **16**	61, 97, 108, 121, 130, 133, 137, **155**, 200, 234, 597, 598, 599, 799, 958 = **15**	95, 116, 133, 138, 141, 142, **155**, 595 = **8**
	Gained hubs	44, 68, 70, 79, 82, 91, 114, 124, 125, 130, 138, 141, 150, 157, 234, 375, 384, 595, 596, 598, 604, 797, 817, 863, 958, 962 = **26**	68, 70, 76, 82, 85, 95, 97, 116, 120, 125, 150, 151, 153, 158, 159, 366, 490, 495, 600, 958, 962, 997 = **22**	44, 63, 66, 73, 85, 114, 116, 117, 120, 134, 138, 146, 147, 353, 374, 596, 602, 820, 826, 829, 832, 859, 956, 962, 966 = **25**	47, 48, 61, 63, 64, 65, 66, 67, 68, 70, 71, 75, 82, 102, 117, 121, 128, 140, 143, 144, 150, 151, 157, 158, 375, 376, 377, 381, 384, 385, 596, 597, 598, 772, 825, 995, 998 = **37**
S492L	**Retained hubs**	75, 76, 77, 97, 101, 102, 104, 108, 116, 117, 121, 132, 133, 137, 142, 153, **155**, 372, 373, 598, 799, 803 = **22**	108, 117, 121, 130, 131, 142, 147, **155,** 372, 599, 603, 604 = **12**	65, 97, 116, 117, 121, 130, 151, 153, **155**, 200, 372, 373, 597, 598, 599, 803, 832 = **17**	62, 77, 113, 116, 135, **155**, 372, 373, 605, 799, 962 = **11**
	Gained hubs	58, 59, 60, 85, 91, 127, 138, 140, 141, 147, 150, 156, 344, 375, 553, 554, 595, 597, 604, 804, 863, 962, 998 = **23**	45, 60, 61, 67, 70, 73, 75, 82, 85, 87, 91, 95, 97, 103, 106, 116, 120, 150, 153, 157, 342, 493, 595, 598, 768, 797, 859, 860, 862, 954, 962, 996, 997, 998 = **34**	24, 28, 44, 45, 63, 64, 73, 77, 81, 82, 83, 91, 104, 113, 128, 131, 141, 144, 150, 346, 376, 602, 772, 861, 962, 998 = **26**	62, 69, 76, 77, 92, 95, 97, 101, 102, 113, 116, 120, 124, 125, 127, 131, 133, 135, 137, 138, 141, 142, 145, 149, 155, 228, 231, 234, 370, 372, 373, 374, 380, 383, 595, 602, 603, 605, 775, 799, 822, 853, 856, 962 = **44**
R886C	**Retained hubs**	66, 67, 71, 73, 76, 77, 95, 108, 117, 121, 142, 145, 146, **155,** 235, 386, 598, 599, 775, 799 = **20**	62, 63, 65, 101, 102, 108, 113, 117, 121, 131, 142, 147, **155**, 160, 367, 370, 372, 373, 599, 818, 821 = **21**	66, 67, 71, 73, 76, 77, 95, 108, 117, 121, 142, 145, 146, **155**, 235, 386, 599, 799 = **18**	62, 63, 65, 101, 102, 108, 117, 121, 131, 142, 147, **155**, 160, 367, 370, 372, 373, 599, 818 = **19**
	Gained hubs	22, 43, 62, 63, 68, 70, 81, 93, 113, 124, 141, 144, 150, 160, 232, 240, 346, 367, 375, 591, 595, 596, 605, 879, 997, 1001 = **26**	60, 73, 80, 82, 95, 97, 110, 124, 125, 134, 138, 140, 145, 146, 597, 598, 602, 767, 799, 802, 822, 859, 863, 958, 962, 966 = **26**	22, 43, 62, 63, 70, 81, 93, 113, 124, 141, 144, 150, 160, 232, 240, 346, 375, 591, 595, 596, 598, 605, 879, 997, 1001 = **26**	60, 73, 80, 82, 95, 97, 110, 113, 124, 125, 134, 138, 140, 145, 146, 597, 598, 602, 767, 799, 802, 821, 822, 859, 863, 958, 962, 966 = **28**
L993R	**Retained hubs**	64, 75, 77, 97, 101, 102, 108, 116, 117, 120, 146, **155,** 373, 493, 595, 598, 599 = **17**	62, 63, 66, 101, 102, 108, 117, 121, 130, 131, 132, 147, 149, **155,** 372, 373, 376, 599, 601, 818, 826, 995 = **22**	64, 75, 97, 101, 102, 108, 116, 117, 120, 153, **155,** 373, 493, 599 = **14**	63, 66, 101, 102, 108, 117, 121, 130, 131, 132, 137, 138, 147, 149, 155, 372, 373, 376, 599, 601, 818, 826, 995 = **23**

	Gained hubs	65, 70, 78, 105, 130, 131, 147, 151, 153, 346, 375, 596, 797, 832, 999 = **15**	67, 74, 75, 79, 91, 95, 97, 98, 137, 138, 146, 150, 153, 598, 799, 825, 883, 891, 958 = **19**	65, 70, 78, 105, 130, 131, 147, 151, 346, 375, 595, 596, 598, 797, 832, 999 = **16**	67, 74, 75, 79, 91, 95, 97, 98, 146, 150, 153, 598, 799, 825, 883, 891, 958 = **17**

Table S4.6.2. *Degree centrality (DC*) hubs found in drug free and drug bound systems; hubs gained due to mutation, and hubs that were maintained despite the presence of mutation.

DC		**Drug free**		**Drug bound**	
Systems		**Monomer A**	**Monomer B**	**Monomer A**	**Monomer B**
WT	**Hubs**	72, 90, **108**, 110, 113, 116, 138, **155***, 159, 198, 199, 200, 201, 202, 203, 207, 218, 282, 314, 356, 435, 482, 499, 548, 576, 625, 710, 740, **778, 779, 781, 790,** 807, **808**, 813, 832, 833, 837, 928 = **39**	**108**, 110, **155***, 192, 193, 200, 202, 203, 225, 282, 335, 337, 482, 496, 499, 554, 625, 653, 654, 694, 707, 711, 729, 732, **778, 779**, 781, **790, 808**, 813, 833, 990 = **32**	72, **108,** 110, 113, 116, **155*,** 157, 194, 199, 200, 203, 206, 207, 218, 232, 313, 314, 335, 350, 351, 478, 480, 483, 542, 554, 595, 605, 625, 628, 641, 653, 654, 665, 666, 707, 710, 714, 729, 732, 740, 776, **778, 779, 781, 790**, 807, **808,** 813, 816, 833, 837, 928 = **52**	31, 65, 67, 72, **108**, 110, 113, 116, 139, 153, 154, **155*,** 157, 159, 198, 200, 202, 203, 282, 314, 469, 479, 480, 495, 496, 499, 542, 554, 568, 589, 625, 653, 654, 710, 732, 740, 767, **778**, **781, 790, 808**, 813, 833, 837 = **44**
S201R	**Retained hubs**	72, 90, **108, 110**, 138, **155***, 198, 199, **202, 203**, 218, 282, 314, 356, 499, 625, 778, 779, 781, 790, 808, 832, 837 = **23**	**108, 110, 155***, 200, **202, 203**, 282, 337, 482, 548, **625**, 653, 729, 732, **778, 779, 781, 790, 808**, 833 = **20**	**108, 110**, 113, **155***, 157, 199, **203**, 206, 542, 554, 595, **625**, 628, 654, 729, 732, 740, 765, **778, 779, 781, 790**, 807, **808**, 813, 837 = **26**	72, **108, 110,** 113, 138, **155***, 159, **203**, 337, 356, 605, **625**, 628, 641, 653, 654, 710, 729, 732, **778, 779, 781, 790**, 807, **808**, 813, 832, 833 = **28**
	Gained hubs	190, 194, 195, 206, 313, 337, 347, 348, 359, 437, 469, 479, 480, 482, 495, 496, 503, 554, 568, 595, 653, 691, 695, 729, 732, 740, 835 = **27**	72, 113, 125, 127, 138, 140, 153, 157, 159, 194, 198, 199, 201, 218, 344, 347, 356, 403, 437, 469, 479, 480, 495, 496, 499, 554, 589, 595, 691, 695, 740, 770, 807, 832, 835, 928 = **36**	90, 127, 153, 201, 202, 216, 282, 310, 314, 318, 337, 347, 356, 359, 437, 493, 495, 499, 503, 568, 569, 691, 695, 716, 724, 835 = **26**	198, 199, 201, 202, 216, 282, 347, 403, 431, 437, 479, 493, 495, 496, 499, 554, 595, 607, 667, 669, 691, 695, 724, 740, 835, 837 = **26**
K259E	**Retained hubs**	90, **108**, 110, 113, 116, 138, **155***, 159, 206, 548, 710, 740, **778**, 779, 781, 790, 808, 832, 833, 928 = **20**	108, 110, 155, 203, 337, 482, 625, 653, 694, 729, 732, **778**, 779, 781, 790, 808, 813, 833 = **18**	72, **108**, 110, **155***, 199, 200, 207, 554, 625, 628, 641, 653, 654, 710, 729, 732, 740, **778**, 779, 781, 790, 807, 808, 813, 833, 837, 928 = **27**	72, 108, 110, 138, **155***, 159, 203, 337, 356, 478, 625, 653, 707, 710, **778**, 779, 781, 790, 807, 808, 816, 818, 833 = **23**
	Gained hubs	199, 203, 207, 218, 282, 493, 542, 554, 568, 628, 687, 732, 767 = **13**	43, 67, 72, 204, 216, 499, 568, 595, 710, 740, 837, 928 = **12**	67, 90, 125, 126, 138, 158, 198, 214, 282, 337, 344, 347, 437, 694, 810 = **15**	158, 190, 198, 199, 201, 202, 214, 282, 344, 479, 480, 482, 495, 496, 499, 554, 568, 740, 766, 767, 775, 810, 815, 837 = **14**
D342N	**Retained hubs**	108, 110, 113, 116, 138, 155, 198, 199, 203, 282, 356, 482, 499, 808, 833, 928 = **16**	72, 110, 155, 116, 125, 194, 203, 225, 337, 482, 499, 554, 625, 732, 778, 779, 781, 790, 808, 833 = **20**	72, 108, 110, 113, 116, 155, 203, 335, 478, 542, 554, 628, 641, 653, 707, 710, 729, 732, 740, 776, 778, 779, 781, 790,	72, 108, 110, 155, 159, 203, 568, 625, 654, 707, 710, 732, 775, 776, 778, 779, 781, 790, 807, 808, 813, 833, 928 = **23**

				808, 813, 833, 837 = **28**	
	Gained hubs	122, 139, 335, 337, 344, 437, 469, 478, 479, 480, 486, 492, 495, 554, 662, 775, 778, 779, 781, 790, 813 = **21**	108, 218, 227, 238, 241, 344, 356, 595, 641, 653, 654, 667, 710, 734, 767, 775, 810, 837 = **18**	120, 125, 138, 159, 202, 282, 356, 372, 434, 469, 479, 482, 490, 493, 495, 499, 568, 765, 770, 774, 832, 835, 990 = **23**	138, 153, 198, 199, 201, 218, 282, 482, 499, 554, 606, 608, 641, 666, 740, 754, 765, 767, 777, 837 = **20**
D432N	**Retained hubs**	72, 108, 110, 122, 206, 282, 314, 356, 499, 710, 778, 779, 781, 790, 807, 808, 832, 833, 928 = **19**	108, 110, 155, 554, 625, 653, 707, 778, 779, 790, 807, 808, 833 = **13**	72, 108, 110, 113, 116, 155, 157, 159, 194, 199, 200, 203, 206, 207, 218, 282, 335, 478, 479, 480, 542, 554, 595, 625, 628, 654, 707, 710, 740, 778, 779, 781, 790, 807, 808, 813, 833 = **37**	72, 108, 110, 113, 138, 139, 153, 155, 159, 203, 207, 225, 241, 337, 356, 478, 568, 595, 603, 605, 625, 653, 654, 694, 707, 710, 729, 732, 776, 778, 779, 781, 790, 796, 807, 808, 813, 816, 832, 833, 928, 962 = **42**
	Gained hubs	67, 138, 155, 159, 198, 199, 201, 202, 203, 350, 479, 482, 493, 495, 554, 568, 595, 628, 732, 740, 768, 962 = **22**	113, 122, 126, 127, 153, 159, 197, 337, 493, 595, 740, 928 = **12**	72, 108, 110, 113, 138, 139, 155, 207, 337, 595, 625, 710, 732, 778, 779, 781, 790, 796, 807, 808, 833, 928 = **22**	31, 67, 90, 137, 198, 201, 206, 218, 282, 314, 469, 479, 480, 482, 495, 499, 548, 554, 606, 607, 734, 740, 775, 816, 829, 837 = **26**
S492L	**Retained hubs**	139, 155, 159, 356, 710, 740, 778, 779, 781, 790, 807, 808, 810, 832, 833 = **15**	72, 155, 202, 203, 282, 482, 495, 499, 554, 625, 653, 654, 729, 778, 779, 790, 808, 813, 833 = **19**	72, 108, 110, 113, 116, 155, 200, 203, 207, 314, 542, 554, 625, 628, 653, 654, 665, 710, 740, 778, 779, 781, 790, 807, 808, 813, 833, 837, 928 = **29**	72, 108, 110, 113, 138, 139, 155, 159, 337, 625, 653, 654, 710, 729, 732, 778, 779, 781, 790, 808, 813, 833, 928 = **23**
	Gained hubs	72, 108, 110, 113, 116, 138, 160, 344, 469, 554, 595, 605, 628, 707, 770, , 835, 928 = **17**	108, 110, 113, 127, 138, 139, 159, 197, 198, 199, 201, 205, 207, 218, 356, 595, 740, 765, 796, 807, 837 = **21**	93, 104, 120, 122, 125, 138, 154, 159, 160, 198, 199, 201, 202, 228, 282, 315, 356, 431, 490, 493, 496, 499, 548, 660, 662, 767, 832, 961, 990 = **29**	90, 116, 199, 201, 202, 218, 282, 335, 482, 554, 698, 740, 770, 837 = **14**
R886C	**Retained hubs**	72, 108, 110, 113, 155, 159, 200, 203, 207, 218, 314, 625, 740, 778, 779, 781, 790, 808, 835 = **19**	108, 110, 203, 220, 282, 625, 653, 707, 729, 732, 778, 779, 781, 790, 808, 813, 832 = **17**	72, 108, 113, 155, 199, 203, 206, 207, 218, 478, 480, 554, 595, 605, 625, 628, 641, 654, 666, 710, 729, 740, 778, 779, 781, 790, 807, 808, 813, 928 = **30**	72, 108, 110, 113, 138, 153, 155, 159, 207, 337, 356, 478, 595, 625, 653, 654, 710, 729, 778, 779, 781, 790, 807, 808, 813 = **25**
	Gained hubs	122, 125, 138, 195, 335, 431, 554, 568, 589, 595, 605, 628, 654, 698, 729, 732, 810, 833, 928 = **19**	72, 100, 138, 139, 140, 155, 159, 356, 482, 554, 568, 595, 605, 710, 740, 767, 807, 810, 833 = **19**	31, 67, 122, 125, 159, 198, 202, 282, 356, 403, 434, 479, 482, 490, 493, 495, 496, 499, 589, 606, 607, 691, 724, 837 = **24**	116, 158, 190, 198, 199, 201, 202, 214, 242, 282, 335, 344, 437, 479, 480, 482, 486, 495, 496, 499, 641, 667, 685, 691, 740, 835, 837 = **27**

L993R	**Retained hubs**	72, 90, 108, 110, 113, 116, 138, 155, 159, 206, 207, 218, 282, 337, 347, 482, 499, 554, 568, 569, 595, 625, 710, 835 = **24**	108, 110, 155, 200, 202, 203, 337, 482, 490, 493, 495, 496, 499, 554, 595, 625, 628, 653, 654, 729, 732, 778, 779, 781, 790, 808, 813, 832, 835 = **29**	72, 108, 110, 113, 116, 155, 157, 199, 200, 201, 203, 206, 207, 478, 480, 554, 595, 605, 625, 628, 666, 729, 732, 740, 778, 779, 781, 790, 807, 808, 813, 837 = **32**	72, 108, 110, 113, 138, 155, 159, 337, 356, 359, 568, 595, 625, 653, 654, 710, 729, 732, 778, 779, 781, 790, 808, 813, 837, 928 = **26**
	Gained hubs	122, 125, 194, 198, 199, 200, 201, 202, 203, 356, 437, 469, 479, 480, 486, 493, 495, 496, 542, 628, 653, 654, 691, 695, 724, 729, 732, 740, 778, 779, 781, 790, 807, 808, 832 = **35**	113, 122, 125, 153, 160, 198, 199, 201, 206, 207, 218, 282, 314, 347, 352, 356, 359, 437, 469, 478, 479, 480, 542, 568, 691, 695, 710, 724, 734, 740, 774, 792, 807, 832, 833 = **35**	31, 93, 120, 122, 125, 159, 198, 202, 282, 337, 344, 347, 348, 356, 403, 431, 437, 469, 479, 482, 490, 493, 495, 496, 499, 568, 653, 673, 691, 695, 708, 724, 810, 832, 835 = **35**	67, 90, 111, 116, 122, 125, 194, 195, 198, 199, 200, 201, 202, 203, 206, 207, 218, 282, 344, 347, 348, 437, 469, 479, 480, 482, 486, 493, 495, 496, 499, 542, 554, 572, 589, 606, 628, 657, 666, 667, 669, 691, 695, 724, 740, 832, 835 = **47**

Table S4.6.3. *Closeness centrality (CC)* hubs found in drug free and drug bound systems; hubs gained due to mutation, and hubs that were maintained despite the presence of mutation.

.

CC		**Drug free**		**Drug bound**	
Systems		**Monomer A**	**Monomer B**	**Monomer A**	**Monomer B**
WT	**Hubs**	71, 72, 75, 76, 79, 80, 101, 102, 104, 105, 137, 140, 141, 142, 145, 146, 147, 150, 151, 152, 155, 596, 597, 598, 599, 600, 601, 604, 605, 768 = **30**	67, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 101, 104, 105, 137, 140, 142, 145, 147, 149, 150, 151, 152, 155, 595, 596, 599, 602, 603, 605 = **32**	71, 72, 75, 76, 79, 80, 101, 102, 104, 105, 137, 140, 141, 142, 145, 146, 147, 150, 151, 152, 155, 596, 597, 598, 599, 600, 601, 604, 605, 768 = **30**	67, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 80, 101, 104, 105, 137, 140, 142, 145, 149, 150, 151, 152, 155, 595, 596, 599, 602, 603, 605 = **30**
S201R	**Retained hubs**	67, 71, 72, 73, 74, 75, 76, 77, 78, 79, 101, 104, 105, 150, 151, 152, 155, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 768 = **28**	73, 76, 77, 101, 104, 105, 146, 150, 151, 155, 596, 597, 598, 599, 600, 601, 605 = **17**	71, 72, 75, 76, 79, 80, 101, 102, 104, 105, 140, 141, 142, 145, 146, 147, 150, 151, 152, 155, 596, 597, 598, 599, 600, 601, 604, 605, 997 = **29**	67, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 80, 101, 104, 105, 137, 140, 142, 145, 149, 150, 151, 152, 155, 595, 596, 599, 602, 603 = **29**
	Gained hubs	68, 69, 70, 80, 81, 82, 102, 137, 140, 141, 142, 145, 146, 147, 148, 149, 153, 154, 156, 158, 594, 595, 769, 797, 997 = **25**	67, 70, 71, 72, 74, 75, 78, 79, 80, 81, 82, 97, 98, 102, 106, 137, 140, 141, 142, 145, 147, 148, 149, 152, 153, 154, 156, 158, 594, 595, 604, 769, 997 = **33**	67, 68, 70, 73, 74, 77, 78, 82, 149, 592, 595, 602, 603, 606, 768, 769 = **16**	66, 68, 79, 81, 146, 147, 153, 154, 156, 158, 159, 597, 598, 600, 601, 604, 605, 768, 769, 821, 861, 862, 997, 999 = **24**
K259E	**Retained hubs**	71, 72, 73, 74, 75, 76, 77, 78, 79, 150, 151, 152, 155, 596, 597, 598, 599, 601, 602, 603 = **20**	73, 76, 77, 101, 146, 150, 155, 598 = **8**	71, 72, 75, 76, 79, 80, 101, 104, 105, 140, 141, 142, 146, 147, 150, 151, 152, 155, 596, 597, 598, 599, 600, 601, 604, 605 = **26**	67, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 80, 101, 104, 105, 137, 140, 142, 149, 150, 151, 152, 155, 599, 602, 603, 605 = **27**
	Gained hubs	81, 97, 98, 101, 104, 105, 140, 142, 147, 605 = **10**	67, 68, 71, 72, 74, 75, 78, 79, 142 = **9**	67, 68, 70, 73, 74, 77, 78, 81, 82, 153, 158, 593, 595, 602, 603, 768, 997 = **17**	68, 79, 81, 97, 102, 146, 147, 153, 154, 156, 158, 159, 595, 596, 597, 598, 600, 601, 604, 821, 997 = **21**
D342N	**Retained hubs**	67, 71, 72, 73, 74, 75, 76, 77, 78, 79, 101, 104, 105, 150, 151, 152, 155, 596, 768 = **19**	73, 76, 77, 101, 104, 105, 146, 150, 151, 155, 596, 597, 598, 599, 600, 601, 603, 605 = **18**	71, 72, 75, 76, 79, 80, 101, 104, 105, 137, 140, 141, 142, 145, 146, 147, 150, 151, 152, 155, 597, 598, 599, 600, 601, 604, 605, 768 = **28**	67, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 101, 104, 150, 151, 152, 155, 595, 596, 599, 602, 603, 605 = **23**
	Gained hubs	68, 69, 70, 80, 81, 82, 85, 87, 98, 100, 102, 103, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 149, 153, 154,	66, 67, 68, 69, 70, 71, 72, 74, 75, 78, 79, 80, 81, 82, 98, 102, 103, 106, 140, 141, 142, 145, 147, 152, 153, 154, 156, 158,	67, 68, 69, 70, 73, 74, 77, 78, 81, 82, 87, 97, 138, 149, 153, 154, 156, 158, 603, 997, 999 = **21**	65, 68, 79, 146, 153, 154, 158, 597, 598, 600, 601 = **11**

		156, 158, 595, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 769, 821, 997 = **42**	595, 602, 604, 821, 859, 860, 862, 954, 956, 997, 998, 999 = **40**		
D432N	**Retained hubs**	67, 71, 72, 73, 74, 75, 76, 77, 78, 79, 101, 104, 105, 150, 151, 155, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 768 = **27**	76, 77, 101, 104, 105, 150, 151, 155, 596, 597, 598, 599, 600, 601, 602 = **15**	71, 72, 75, 76, 101, 104, 105, 146, 147, 150, 151, 152, 155, 596, 597, 598, 599 = **17**	67, 71, 72, 73, 74, 75, 76, 80, 101, 104, 105, 137, 140, 142, 149, 150, 151, 155, 596, 599 = **20**
	Gained hubs	68, 70, 80, 81, 140, 142, 595 = **7**	70, 72, 75, 79, 80, 81, 82, 140, 142, 145, 147, 149, 152, 604, 997 = **15**	73, 78, 595, 997 = **4**	68, 79, 81, 82, 141, 604 = **6**
S492L	**Retained hubs**	67, 71, 72, 73, 74, 75, 76, 77, 78, 79, 101, 104, 105, 155 = **14**	73, 74, 75, 76, 105, 146, 150, 151, 155, 597, 598, 599, 600, 601, 603, 605, 768 = **17**	71, 72, 75, 76, 79, 80, 101, 104, 105, 140, 142, 150, 151, 152, 155, 598, 599, 601 = **18**	73, 74, 75, 76, 77, 78, 80, 101, 105, 140, 142, 149, 150, 151, 152, 155, 599, 603, 605 = **19**
	Gained hubs	70, 80, 81, 82, 97, 102, 137, 140, 141, 142, 145, 146, 147, 149, 150, 151, 152, 153, 154, 156, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605 = **31**	67, 68, 69, 70, 71, 72, 77, 78, 79, 80, 81, 82, 87, 97, 98, 101, 102, 103, 104, 106, 137, 140, 141, 142, 145, 147, 149, 152, 153, 154, 595, 596, 602, 604, 797, 859, 954, 956, 995, 997, 998, 999 = **42**	65, 67, 68, 69, 70, 73, 74, 77, 78, 81, 97, 153, 154, 158, 595, 596, 597, 602 = **18**	79, 141, 146, 598, 600, 601, 604 = **7**
R886C	**Retained hubs**	67, 71, 72, 73, 74, 75, 76, 77, 78, 79, 101, 104, 105, 150, 151, 152, 155, 596, 597, 598, 599, 601, 603, 604, 605 = **25**	73, 76, 77, 101, 104, 146, 150, 151, 155, 596, 597, 598, 599, 600, 601, 602, 605 = **17**	71, 72, 75, 76, 79, 80, 101, 102, 104, 105, 140, 142, 145, 146, 147, 149, 150, 151, 152, 155, 596, 597, 598, 599, 600, 601, 604, 605, 768 = **28**	67, 70, 71, 72, 73, 74, 75, 76, 77, 78, 80, 101, 104, 105, 137, 140, 142, 145, 149, 150, 151, 152, 155, 595, 596, 599, 602, 603 = **28**
	Gained hubs	68, 70, 80, 81, 142, 145, 595, 997 = **8**	71, 72, 74, 75, 79, 80, 140, 142, 145, 147 = **10**	67, 68, 70, 73, 74, 77, 78, 143, 144, 154, 158, 595, 602, 603, 821 = **15**	79, 81, 82, 87, 97, 138, 141, 146, 147, 153, 154, 156, 158, 597, 598, 600, 601, 604, 605 = **19**
L993R	**Retained hubs**	65, 71, 72, 73, 74, 75, 76, 77, 78, 79, 101, 104, 105, 106, 150, 151, 152, 155, 596, 597, 598, 599, 600, 601, 602, 604, 605, 768 = **28**	67, 68, 69, 70, 71, 72, 76, 77, 101, 104, 105, 146, 150, 151, 155, 596, 597, 598, 599, 600, 601, 602, 603, 605, 768 = **25**	71, 72, 73, 74, 75, 76, 79, 80, 101, 104, 105, 140, 142, 146, 147, 150, 151, 155, 598, 599, 600, 601, 605, 606, 768 = **25**	67, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 80, 104, 105, 137, 140, 142, 145, 149, 150, 151, 152, 155, 595, 596, 599, 602, 603, 605 = **29**
	Gained hubs	67, 70, 80, 81, 97, 98, 100, 102, 140, 141, 142, 145, 146, 147, 148, 149, 153, 154, 156, 158, 594, 595, 769, 821, 999 = **25**	73, 74, 75, 78, 79, 80, 81, 87, 97, 98, 100, 102, 137, 140, 142, 145, 147, 148, 149, 152, 153, 154, 158, 595, 604, 821, 862, 995, 997 = **29**	67, 68, 69, 70, 77, 78, 81, 97, 158, 594, 595, 596, 597, 602, 603, 604, 769, 862, 997 = **19**	66, 68, 79, 81, 82, 97, 98, 101, 102, 107, 141, 144, 146, 147, 148, 154, 156, 158, 597, 598, 600, 601, 604, 768, 769, 821, 997 = **17**

Table S4.6.4. *Eigen centrality (EC)* hubs found in drug free and drug bound systems; hubs gained due to mutation, and hubs that were maintained despite the presence of mutation.

EC		**Drug free**		**Drug bound**	
Systems		**Monomer A**	**Monomer B**	**Monomer A**	**Monomer B**
WT	**Hubs**	121, 122, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 214, 216, 217, 225, 248, 253, 280, 281, 282, 283, 469, 477, 478, 479, 480, 482, 483, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 502, 503, 734, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 810 = **61**	None	100, 103, 104, 106, 108, 109, 110, 111, 112, 113, 114, 115, 116, 159, 665, 666, 667, 707, 708, 710, 717, 718, 721, 724, 729, 732, 733, 734, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 790, 791, 792, 796, 799, 802, 803, 804, 805, 806, 807, 808, 809, 810, 813, 829, 830, 832, 833, 834, 835, 836, 837, 852, 926, 928, 929 = **67**	595, 710, 732, 734, 740, 741, 744, 775, 776, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810 = **20**
S201R	**Retained hubs**	198, 479, 480, 495, 496, 499 = **6**	120, 121, 122, 123, 124, 125, 126, 127, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 214, 216, 217, 218, 224, 225, 227, 244, 245, 248, 280, 281, 282, 283, 284, 462, 464, 469, 471, 477, 478, 479, 480, 481, 482, 483, 486, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 503 = **64**	708, 710, 717, 721, 724, 732, 734, 774, 775, 776, 777, 778, 779, 780, 781, 782, 790, 792, 805, 808, 809, 810 = **22**	710, 732, 734, 772, 775, 776, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810 = **17**
	Gained hubs	31= **1**	None	740, 74 = **2**	640, 641, 665, 666, 667, 668, 669, 684, 685, 691, 692, 694, 695, 698, 706, 707, 708, 709, 711, 717, 718, 720, 721, 723, 724, 727, 729, 731, 733, 735, 771, 773, 774, 780, 791, 802, 804, 806, 807, 813, 926, 928 = **42**
K259E	**Retained hubs**	734, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 810 = **12**	None	734, 775, 777, 778, 779, 782, 790, 792, 805, 808, 809, 810 = **12**	732, 734, 775, 776, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810 = **15**

	Gained hubs	100, 103, 104, 106, 108, 109, 110, 804, 806, 807, 809, 813, 829, 830, 832, 833, 834, 835, 836, 837, 852, 926, 928, 929 = **24**	100, 103, 104, 106, 108, 109, 110, 111, 112, 113, 114, 115, 116, 155, 156, 158, 159, 160, 163, 710, 717, 718, 721, 732, 733, 735, 740, 767, 768, 769, 770, 771, 772, 773, 774, 776, 780, 783, 791, 793, 794, 795, 796, 799, 801, 802, 803, 804, 806, 807, 809, 813, 829, 830, 832, 833, 834, 835, 836, 837, 852, 926, 928, 929 = **64**	781 = **1**	196, 197, 198, 199, 200, 201, 202, 203, 205, 206, 280, 281, 282, 469, 477, 478, 479, 480, 482, 483, 486, 492, 493, 495, 496, 497, 498, 499, 500, 767, 774, 791, , 813 = **33**
D342N	**Retained hubs**	192, 193, 194, 197, 198, 199, 202, 203, 280, 281, 282, 283, 469, 477, 478, 479, 480, 482, 483, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 502, 503 = **31**	None	103, 108, 109, 110, 111, 112, 113, 116, 159, 707, 708, 710, 717, 721, 732, 733, 734, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 790, 791, 792, 796, 799, 802, 803, 804, 805, 806, 807, 808, 809, 810, 813, 829, 832, 833, 836, 926 = **49**	710, 732, 734, 740, 741, 775, 776, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810 = **18**
	Gained hubs	284, 464, 481, 484, 486 = **5**	31, 666, 667, 707, 708, 709, 710, 717, 721, 729, 732, 733, 734, 735, 774, 775, 776, 777, 778, 779, 780, 781, 782, 790, 791, 792, 793, 804, 805, 806, 807, 808, 809, 810, 813, 926 = **36**	479, 712, 735, 738, 740, 741, 765, 767, 770, 793, 794, 795, , 814, 928 = **14**	606, 608, 738, 739, 764, 765, 767, 770, 771, 772, 773, 774, 813 = **13**
D432N	**Retained hubs**	121, 122, 126, 127, 155, 159, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 214, 216, 217, 248, 253, 280, 281, 282, 283, 469, 477, 478, 479, 480, 482, 483, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 502, 503 = **50**	None	None	92, 93, 120, 121, 122, 126, 127, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 214, 216, 217, 218, 245, 248, 253, 280, 281, 282, 283, 284, 285, 462, 464, 469, 470, 471, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 502, 503, 504 = **77**
	Gained hubs	93, 104, 108, 109, 110, 111, 112, 113, 116, 120, 284, 462, 464, 470, 471, 481, 484, 486, 489, 490, 504, 833 = **22**	None	31 = **1**	None

S492L	**Retained hubs**	734, 777, 778, 779, 781, 782, 790, 792, 808, 810 = **10**	None	108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 159, , 734, 774, 775, 777, 778, 779, 782, 790, 792, 799, 804, 805, 807, 808, 809, 810, 813, 829, 832, 833, 836 = **32**	778 = **1**
	Gained hubs	595, 770, 773, 774, 775, 776 = **6**	120, 122, 126, 127, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 214, 216, 218, 225, 226, 280, 281, 282, 283, 469, 477, 478, 479, 480, 482, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 502, 503, 595, 605, 740, 765, 767, 769, 770, 771, 774, 775, 777, 778, 790, 794, 795, 805, 808, 810 = **77**	67, 72, 75, 76, 93, 100, 101, 103, 104, 120, 121, 122, 123, 124, 125, 126, 127, 130, 151, 152, 153, 154, 155, 156, 157, 158, 160, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 244, 282, 489, 490, 492, 493, 494, 495, 496, 497, 499, 500, 781 = **51**	None
R886C	**Retained hubs**	734, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 810 = **12**	None	None	None
	Gained hubs	595, 605, 707, 708, 710, 717, 718, 721, 732, 733, 735, 738, 740, 741, 765, 769, 770, 771, 772, 773, 774, 776, 780, 791, 793, 794, 795, 796, 802, 804, 806, 807, 809, 813, 926, 928 = **36**	108, 109, 110, 111, 112, 113, 159, 595, 605, 606, 707, 708, 710, 717, 718, 720, 721, 724, 729, 732, 733, 734, 735, 738, 740, 741, 765, 767, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 790, 791, 792, 793, 794, 796, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 813, 814, 832, 833, 926, 928, 929 = **65**	120, 121, 122, 126, 190, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 209, 211, 214, 216, 217, 218, 248, 253, 280, 281, 282, 283, 284, 462, 464, 469, 470, 471, 477, 478, 479, 480, 481, 482, 483, 486, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 502, 503, 504 = **90**	None
L993R	**Retained hubs**	734, 775, 777, 778, 779, 780, 781, 782, 790, 791, 792, 805, 808, 810 = **14**	None	None	None

	Gained hubs	100, 103, 104, 106, 108, 109, 110, 111, 112, 113, 114, 115, 116, 155, 156, 158, 159, 160, 163, 710, 717, 718, 721, 732, 733, 735, 740, 767, 768, 769, 770, 771, 772, 773, 774, 776, 783, 793, 794, 795, 796, 799, 801, 802, 803, 804, 806, 807, 809, 813, 829, 830, 832, 833, 834, 835, 836, 837, 852, 926, 928, 929 = **62**	595, 710, 717, 721, 732, 734, 738, 740, 741, 742, 743, 744, 750, 765, 769, 770, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 790, 792, 805, 808, 809, 810 = **32**	31 = **1**	92, 93, 116, 120, 121, 122, 123, 125, 126, 127, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 214, 216, 217, 218, 224, 225, 244, 248, 280, 281, 282, 283, 284, 462, 464, 469, 471, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 502, 503 = **79**

Table S4.6.5. *Katz centrality (KC)* hubs found in drug free and drug bound systems; hubs gained due to mutation, and hubs that were maintained despite the presence of mutation.

KC		**Drug free**		**Drug bound**	
Systems		**Monomer A**	**Monomer B**	**Monomer A**	**Monomer B**
WT	**Hubs**	103, 104, 108, 110, 111, 112, 113, 116, 155, 159, 198, 199, 201, 202, 203, 282, 479, 496, 499, 710, 732, 734, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 832, 833 = **36**	108, 110, 111, 155, 192, 193, 194, 195, 196, 198, 199, 200, 201, 202, 203, 206, 280, 282, 479, 480, 492, 493, 495, 496, 499, 500, 707, 710, 729, 732, 734, 774, 775, 776, 777, 778, 779, 781, 782, 790, 792, 805, 808, 810, 832, 833 = **46**	103, 104, 108, 109, 110, 111, 112, 113, 199, 203, 313, 350, 605, 640, 641, 665, 666, 667, 707, 710, 732, 734, 775, 776, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 813, 832, 833, 836, 837, 928 = **40**	108, 110, 111, 155, 159, 595, 707, 708, 710, 732, 734, 775, 778, 779, 781, 782, 790, 792, 805, 808, 810, 833 = **22**
S201R	**Retained hubs**	108, 155, 198, 199, 202, 203, 282, 479, 496, 499, 732, 734, 775, 778, 779, 781, 782, 790, 792, 805, 808, 810 = **22**	108, 110, 111, 155, 194, 195, 196, 199, 200, 201, 202, 203, 280, 282, 479, 480, 493, 495, 496, 499, 732, 734, 774, 775, 777, 778, 779, 781, 790, 808, 810, 832, 833 = **33**	108, 110, 111, 112, 113, 199, 203, 710, 732, 734, 775, 778, 779, 781, 782, 790, 792, 805, 808, 810, 832, 833 = **22**	108, 110, 155, 159, 708, 710, 732, 734, 775, 778, 779, 781, 782, 790, 792, 805, 808, 810, 833 = **19**
	Gained hubs	194, 280, 337, 347, 348, 469, 478, 480, 482, 492, 495, 691 = **12**	112, 113, 122, 123, 125, 127, 157, 159, 197, 198, 469, 478, 482, 595, 691, 724 = **15**	155, 159, 198, 202, 337, 347, 493, 496, 595, 691, 695, 708, 721, 724, 740, 777 = **16**	198, 199, 201, 202, 282, 479, 493, 495, 496, 641, 665, 666, 667, 669, 691, 695, 724, 729, 777 = **19**
K259E	**Retained hubs**	103, 104, 108, 110, 111, 112, 113, 116, 155, 159, 732, 734, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 829, 832, 833 = **26**	108, 155, 710, 732, 734, 774, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 810, 833 = **18**	108, 110, 111, 112, 710, 732, 734, 775, 776, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 832, 833 = **22**	155, 200 = **2**
	Gained hubs	109, 138, 710, 767, 770, 771, 772, 774, 776, 810, 813, 836, 928 = **13**	595, 721, 724, 740, 741, 770, 809, 813 = **8**	104, 155, 159, 710, 732, 734, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 810, 833 = **18**	108, 110, 112, 198, 199, 200, 201, 202, 203, 282, 478, 479, 480, 482, 495, 496, 499, 606, 766, 767, 774, 776, 809, 813, 816, 818 = **26**
D342N	**Retained hubs**	108, 110, 111, 112, 198, 199, 202, 203, 282, 496, 499, 734, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 810, 832, 833 = **25**	108, 110, 111, 196, 199, 707, 710, 732, 734, 774, 775, 776, 777, 778, 779, 781, 782, 790, 792, 805, 808, 810, 832, 833 = **24**	103, 104, 108, 109, 110, 111, 112, 113, 710, 732, 734, 775, 776, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 813, 832, 833, 836 = **28**	108, 155, 198, 710, 732, 734, 775, 778, 779, 781, 782, 790, 792, 805, 808, 810, 813 = **17**
	Gained hubs	469, 478, 479, 480, 482, 486, 492, 495, 498, 772, 776 = **11**	103, 112, 194, 195, 640, 641, 666, 667, 708, 709, 733, 809 = **12**	155, 200, 116, 120, 155, 159, 202, 282, 469, 478, 479, 480, 482, 492, 493, 495, 496, 499,	153, 199, 282, 479, 499, 608, 740, 765, 767, 774, 776, 777 = **12**

				740, 770, 772, 773, 774 = **23**	
D432N	**Retained hubs**	103, 104, 108, 110, 111, 112, 113, 155, 159, 198, 199, 201, 202, 203, 282, 479, 496, 499, 732, 734, 775, 776, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 832, 833 = **35**	108, 110, 111, 155, 493, 775, 776, 777, 778, 779, 781, 782, 790, 805, 808, 810, 832, 833 = **18**	108, 110, 111, 112, 710, 732, 734, 775, 776, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 832, 833 = **22**	108, 110, 155, 159, 710, 732, 734, 775, 778, 779, 781, 782, 790, 792, 805, 808, 810, 833 = **18**
	Gained hubs	67, 72, 109, 122, 154, 158, 192, 194, 197, 200, 206, 280, 469, 478, 480, 482, 492, 493, 495, 500, 503, 595, 772, 836 = **24**	113, 122, 127, 159, 197, 595, 740, 809, 928 = **9**	155, 200 = **2**	104, 112, 108, 110, 111, 112, 113, 199, 200, 201, 202, 203, 282, 478, 479, 480, 482, 495, 496, 499, 606, 766, 767, 774, 776, 777, 809, 813, 816, 818 = **30**
S492L	**Retained hubs**	108, 775, 775, 777, 778, 779, 781, 782, 790, 805, 808, 809, 810, 832, 833 = **15**	108, 110, 111, 155, 192, 194, 196, 198, 199, 200, 201, 202, 203, 206, 280, 282, 479, 480, 492, 493, 495, 496, 499, 500, 734, 774, 775, 776, 777, 778, 779, 781, 782, 790, 792, 805, 808, 810, 832, 833 = **40**	103, 104, 108, 109, 110, 111, 112, 113, 199, 203, 775, 776, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 813, 832 = **24**	108, 110, 111, 112, 155, 159, 710, 732, 734, 774, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 810 = **21**
	Gained hubs	109, 110, 111, 112, 113, 116, 155, 159, 595, 605, 606, 740, 770, 772, 773, 774, 792 = **17**	109, 112, 113, 126, 127, 159, 197, 205, 216, 482, 595, 605, 606, 740, 765, 767, 768, 769, 770, 771, 772, 773, 794, 795, 796, 809, 813, 816 = **28**	93, 116, 120, 122, 123, 124, 125, 126, 127, 153, 154, 155, 156, 159, 160, 197, 198, 201, 202, 200, 490, 493, 494, 496, 497, 732, 734, 774, 777, 833 = **30**	104, 113, 116, 813, 833 = **5**
R886C	**Retained hubs**	108, 110, 111, 112, 113, 155, 159, 200, 203, 710, 732, 734, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 832, 833 = **26**	108, 110, 111, 155, 710, 732, 734, 774, 775, 776, 777, 778, 779, 781, 782, 790, 792, 805, 808, 810, 832, 833 = **23**	103, 108, 109, 110, 111, 112, 113, 199, 203, 732, 734, 775, 778, 779, 781, 782, 790, 792, 805, 808, 810, 813, 833 = **23**	108, 110, 112, 155, 159, 708, 710, 732, 734, 775, 778, 779, 781, 782, 790, 792, 805, 808, 810 = **19**
	Gained hubs	595, 717, 740, 772, 774, 776, 836 = **7**	100, 103, 104, 109, 112, 113, 159, 721, 740, 809, 813 = **11**	122, 155, 159, 192, 194, 196, 197, 198, 200, 201, 202, 205, 206, 216, 280, 282, 469, 478, 479, 480, 482, 492, 493, 495, 496, 499, 500 = **27**	104, 198, 199, 201, 202, 282, 469, 478, 479, 480, 482, 486, 492, 495, 496, 499, 777 = **17**
L993R	**Retained hubs**	108, 110, 111, 112, 113, 116, 155, 159, 198, 199, 201, 202, 203, 282, 479, 496, 499, 732, 734, 778, 781, 790, 792, 808 = **24**	108, 110, 194, 195, 196, 199, 200, 201, 202, 203, 280, 282, 479, 480, 492, 493, 495, 496, 499, 500, 710, 732, 734, 774, 777, 778, 781, 790, 792, 805,	108, 110, 111, 112, 113, 199, 200, 201, 203, 732, 734, 778, 790, 808, 810 = **15**	108, 110, 111, 155, 159, 708, 710, 732, 734, 778, 781, 790, 808, 810, 833 = **15**

			808, 810 = **32**		
	Gained hubs	120, 122, 123, 125, 126, 127, 192, 194, 195, 196, 197, 200, 206, 216, 280, 469, 478, 480, 482, 483, 486, 490, 491, 492, 493, 494, 495, 497, 498, 500 = **30**	31, 122, 125, 197, 198, 469, 478, 482, 708 = **9**	31, 93, 116, 120, 122, 123, 125, 126, 127, 155, 156, 159, 197, 198, 202, 282, 469, 478, 479, 480, 482, 489, 490, 492, 493, 494, 495, 496, 497, 499, 500, 691, 781 = **33**	93, 100, 112, 113, 116, 120, 122, 123, 125, 126, 156, 160, 192, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 206, 216, 280, 282, 469, 478, 479, 480, 482, 483, 486, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 724, 832 = **48**

Table S5.1. DPD missense mutation associated with people of African population as identified by 1000 genomes and Exom sequencing project (ESP MAF), Variants were classified as rare or common based on their minor allele frequency (MAF) of 0.001. Prediction of the effect of 12 missense mutations on DPD structure and function using VAPOR web servers

Variant	SNP ID	DPD activity ± SD	P value	Exom sequencing project (ESP MAF)		Functional impact			Stability change	Reference
						HUMA-VAPOR				
				African American	African	Polyphen-2	Provean	PhD-SNP	MUpro	PHARMGKB
R592Q	rs138616379	n/a	n/a	0.00023	0.00012	Probably damaging	Disease	Disease	Decreases	Doi:10.1158/0008-5472.CAN-13-2482
A664S	rs138545885	0.92 ±0.05	0.26	0.00045	0.065	Probably damaging	Disease	Neutral	Decreases	Doi: 10.1158/0008-5472.CAN-13-2482
G674D	rs137999090	0.092±0.05	0.065	0.00023	nr	Probably damaging	Disease	Disease	Decreases	Doi:10.1158/0008-5472.CAN-13-2482
A721T	rs145548112	0.92 ±0.05	0.52	0.00045	0.065	Possibly damaging	Tolerated	Disease	Decreases	Doi: 10.1158/0008-5472.CAN-13-2482
V732G	rs60511679	0.00318	0	0.0020	0	Possibly damaging	Tolerated	Disease	Decreases	Doi: 10.1158/0008-5472.CAN-13-2482
T768K	rs56005131	na	na	0.00023	0	Possibly damaging	Tolerated	Disease	Decreases	Doi:10.1124/dmd.118.081737

Table S5.2. Estimated center of mass (COM) distances (Å) from the location of the missense mutant to either the nearest cofactors or the substrate. The calculations were based on the modeled native human DPD 3D structure.

Variant	Location (Domain)	COM distance to the nearest co-factor (Å)		COM distance to the substrate (Å)	
		Monomer A	Monomer B	Monomer A	Monomer B
R592Q	5-FU binding domain (II)	n/a	n/a	18.6	18.3
A664S	5-FU binding domain (II)	n/a	n/a	21.6	21.8
G674D	5-FU binding domain (II)	n/a	n/a	16.1	13.6
A721T	5-FU binding domain (II)	n/a	n/a	28.6	28.3
V732G	5-FU binding domain (II)	n/a	n/a	30.7	30.5
T768K	5-FU binding domain (II)	n/a	n/a	20.6	19.8

Table S5.3. Validation scores from the ERRAT, ProSA, and PROCHECK webservers for 13 drug free and drug bound modeled human DPD 3D structures. Both the drug free and drug bound models received the same score.

Protein	ERRAT	ProSA		PROCHECK Model 2		
	Overall quality factor (%)	Monomers z-Score		Ramachandran (residues location) (%)		
		A	B	Favored	Allowed	Disallowed
Template (PDB ID: 1H7X)	85.79	-13.56	-13.47	93.5	6.4	0.1
Wild type	88.86	-14.53	-14.27	89.1	10.7	0.2
R592Q	88.86	-14.53	-14.27	89.1	10.7	0.2
A664S	88.86	-14.53	-14.27	89.1	10.7	0.2
G674D	88.86	-14.53	-14.27	89.1	10.7	0.2
A721T	88.86	-14.53	-14.27	89.1	10.7	0.2
V732G	88.86	-14.53	-14.27	89.1	10.7	0.2
T768K	88.86	-14.53	-14.27	89.1	10.7	0.2

Table S5.4. The protonation states of the residues in the human DPD protein model that participate in the electron transfer channel. In both monomers, HIS was protonated as delta nitrogen (HID), epsilon nitrogen (HIE), or fully positively (+) charged (HIP). In both monomers of each system, protonated cysteine (CYM) glutamine (GLN), and aspartic acid (ASN) were discovered.

Residue	**AMBER protonated residue name**	**Residue number (DPD PDB)**
HIS	[1]HIP	426 and 323
	[2]HIE	23, 38, 62 and 976
	[3]HID	92, 233, 502, 671, 805 and 857
ASN	[4]ASH	420, 577 and 657
CYS	[5]CYM	27, 47, 50, *79, *82, *87, *91, 124, *130, *136, *140, 200, 255, 320, 322, 354, 383, 620, 642, 669, 682, 694, 874, 829, *953, *956, *959, *962, 969, *986, *989, *992, *996 and 999
GLN	[6]GLH	*156

*Represents residues coordinating to the $Fe^{2+}{}_4S^{2-}{}_4$ clusters; [1]HIP, [2]HIE, [3]HID represent the protonation state of histidine (HIS), [4]ASH represents the protonation state of aspartic acid (ASN), [5]CYM represents the protonation state of cysteine (CYS), and [6]GLH represents the protonation state of glutamine (Gln).

Table S5.5 MD simulations revealed the dynamics of 5-FU (1) retention, (2) release within the active site, and the 30 ns of equilibrated frame areas of Monomer A and Monomer B in activated (drug bound) systems.

System	Monomer A				Monomer B			
	Drug exit (Y/N)	Exit time (ns)	Exit frame	DRN calculation Monomer A frame range	Drug exit (Y/N)	Exit time (ns)	Exit frame	DRN calculation monomer B frame range
WT	N	N/A	N/A	55000 -60000	N	N/A	N/A	55000 -60000
R592Q	Y	N/A	N/A	15000-18000	N	220	220000	15000-18000
A664S	Y	180	180000	27000-30000	N	N/A	N/A	27000-30000
G674D	Y	160	160000	10000-13000	N	N/A	N/A	10000-13000
A721T	N	N/A	N/A	19000-22000	N	N/A	N/A	19000-22000
V732G	N	N/A	N/A	34000-37000	N	N/A	N/A	34000-37000
T768K	N	N/A	N/A	32000-35000	N	N/A	N/A	32000-35000

Table S5.6 A representation of the 30 ns of equilibrated frame areas of Monomer A and Monomer B in inactivated (drug free) systems.

System	Monomer A	Monomer B
	DRN calculation Monomer A frame range	DRN calculation monomer B frame range
WT	55000 -60000	55000 -60000
R592Q	150000-180000	150000-180000
A664S	27000-30000	27000-30000
G674D	10000-130000	10000-130000
A721T	19000-22000	19000-22000
V732G	34000-37000	34000-37000
T768K	32000-35000	32000-35000

Table S5.7.1. *Betweenness centrality (BC)* hubs found in drug free and drug bound systems; hubs gained due to mutation, and hubs that were maintained despite the presence of mutation.

BC		Drug free		Drug bound	
Systems		Monomer A	Monomer B	Monomer A	Monomer B
WT	Hubs	38, 61, 64, 66, 67, 71, 73, 74, 75, 76, 77, 80, 95, 97, 101, 102, 104, 108, 112, 116, 117, 120, 121, 132, 133, 136, 137, 142, 145, 146, 152, 153, 155, 235, 350, 370, 372, 373, 386, 493, 597, 598, 599, 603, 605, 768, 775, 799, 802, 803, 826, 862, 894 = 53	62, 63, 65, 66, 86, 90, 101, 102, 105, 108, 113, 114, 117, 121, 127, 130, 131, 132, 133, 137, 139, 142, 143, 144, 147, 149, 155, 160, 361, 367, 370, 372, 373, 376, 599, 601, 603, 604, 816, 818, 826, 955, 991 = 43	61, 65, 80, 85, 97, 101, 102, 108, 114, 116, 117, 121, 122, 123, 125, 126, 130, 132, 133, 137, 140, 142, 155, 200, 227, 231, 234, 310, 372, 373, 374, 375, 597, 598, 599, 604, 605, 768, 775, 776, 797, 799, 803, 804, 816, 825, 958, 962, 996, 997 = 50	22, 44, 45, 62, 69, 76, 77, 92, 95, 97, 101, 102, 113, 116, 120, 124, 125, 127, 131, 133, 135, 137, 138, 141, 142, 145, 149, 155, 228, 231, 234, 370, 372, 373, 374, 380, 383, 595, 602, 603, 605, 775, 799, 822, 853, 856 = 46
R592Q	Retained hubs	67, 80, 101, 102, 104, 108, 116, 117, 120, 132, 142, 153, 155, 372, 373, 386, 493, 597, 598, 599, 768, 799 = 22	63, 117, 120, 121, 137, 142, 146, 147, 149, 155, 370, 372, 373 = 13	85, 114, 117, 123, 125, 133, 142, 155, 372, 373, 597, 598, 599, 605, 958, 962, 997 = 17	95, 101, 102, 116, 120, 131, 137, 155, 373, 374, 775, 799, 822, 955 = 14
	Gained hubs	73, 70, 83, 85, 130, 131, 147, 157, 158, 159, 345, 374, 375, 376, 595, 596, 600, 604, 797, 798, 822, 996, 997 = 23	67, 70, 76, 77, 91, 116, 126, 129, 133, 150, 153, 156, 381, 493, 595, 596, 598, 599, 767, 799, 821 = 21	44, 46, 47, 60, 67, 70, 76, 82, 120, 124, 141, 144, 146, 147, 150, 153, 159, 160, 367, 595, 596, 832, 861 = 23	63, 65, 83, 85, 91, 93, 98, 117, 123, 130, 136, 143, 146, 493, 597, 598, 599, 772, 803, 804, 958, 962, 995 = 23
A664S	Retained hubs	64, 66, 67, 95, 108, 116, 117, 121, 132, 146, 155, 370, 493, 597, 598, 599, 775 = 17	62, 63, 65, 102, 108, 114, 117, 142, 155, 599 = 10	61, 80, 85, 97, 108, 116, 137, 138, 140, 142, 155, 231, 372, 373, 598, 599, 604 = 17	44, 45, 116, 125, 142, 372, 373, 595 = 8
	Gained hubs	44, 45, 62, 63, 65, 70, 128, 131, 151, 346, 347, 367, 594, 595, 596, 883, 962, 997 = 18	22, 73, 75, 76, 77, 80, 82, 97, 98, 116, 120, 125, 146, 150, 151, 349, 493, 768, 832, 859, 962, 995, 996, 998 = 24	38, 44, 46, 63, 74, 76, 81, 82, 130, 132, 141, 144, 146, 147, 493, 596, 601 = 17	38, 42, 43, 63, 64, 65, 68, 70, 71, 73, 83, 110, 117, 143, 144, 146, 147, 158, 235, 376, 598, 599, 797, 826, 829, 861, 958, 997 = 28
G674D	Retained hubs	38, 64, 73, 75, 95, 108, 116, 117, 121, 133, 142, 155, 373,	101, 102, 108, 113, 114, 117, 121, 132, 137, 142, 147, 149,	65, 85, 108, 114, 116, 117, 121, 130, 142, 155, 200, 372,	62, 65, 70, 76, 101, 102, 116, 120, 131, 142, 155, 370, 372, 373, 799 =

		591, 599 = 15	155, 160, 599 = 15	375, 598, 599, 799, 958, 997 = 18	15
	Gained hubs	68, 70, 113, 114, 131, 138, 147, 150, 151, 157, 158, 160, 234, 368, 370, 372, 604, 797, 821, 832, 962, 997 = 22	22, 23, 58, 70, 75, 80, 85, 95, 110, 153, 164, 243, 384, 597, 598, 799, 815, 832, 853, 883, 887, 891, 958, 962, 966, 995, 996, 997 = 28	63, 70, 82, 104, 113, 124, 138, 141, 146, 147, 150, 153, 157, 160, 235, 236, 346, 386, 493, 591, 802, 826, 829, 832, 995 = 25	80, 85, 108, 130, 132, 147, 151, 153, 363, 367, 493, 598, 958, 962, 996, 998 = 16
A721T	Retained hubs	67, 76, 77, 80, 108, 112, 116, 120, 121, 132, 142, 153, 155, 493, 595, 597, 599, 768, 797, 799, 803 = 21	63, 66, 101, 105, 113, 127, 130, 131, 132, 133, 142, 155, 372, 373, 599, 995 = 16	65, 80, 101, 102, 122, 126, 132, 133, 142, 155, 372, 373, 597, 598, 599, 604, 775, 803, 804, 958, 962 = 21	62, 76, 77, 101, 102, 113, 124, 125, 138, 141, 142, 155, 372, 373, 595, 799, 853 = 17
	Gained hubs	63, 78, 81, 91, 95, 123, 130, 147, 150, 151, 154, 156, 345, 375, 376, 381, 602 = 17	109, 112, 113, 116, 125, 572, 605, 606, 641, 666, 766, 767, 769, 770, 774, 794, 796, 813, 814, 816 = 20	44, 62, 63, 68, 70, 74, 77, 82, 96, 119, 120, 127, 131, 138, 143, 144, 146, 150, 153, 158, 346, 370, 528, 596, 603, 821, 862, 994, 995 = 29	61, 63, 81, 91, 95, 98, 108, 110, 114, 117, 121, 129, 140, 146, 147, 150, 152, 153, 156, 160, 361, 376, 600, 604, 765, 956, 998 = 27
V732G	Retained hubs	76, 77, 95, 101, 102, 110, 116, 120, 132, 133, 137, 142, 153, 155, 370, 373, 493, 600, 603, 803 = 21	65, 66, 113, 117, 121, 142, 144, 155, 372, 373, 376, 599 = 12	65, 101, 102, 114, 116, 117, 121, 130, 132, 133, 137, 142, 155, 373, 598, 605, 803 = 17	76, 101, 116, 125, 127, 131, 133, 137, 142, 155, 370, 372, 373 = 13
	Gained hubs	43, 44, 61, 63, 65, 113, 126, 127, 130, 131, 144, 146, 150, 157, 159, 160, 374, 376, 604, 837, 804, 958, 995 = 23	58, 67, 68, 73, 76, 77, 78, 82, 95, 110, 116, 120, 125, 129, 141, 146, 150, 157, 158, 159, 345, 348, 375, 381, 493, 528, 595, 596, 742, 803, 837, 995 = 32	23, 60, 63, 67, 73, 77, 104, 131, 143, 147, 149, 150, 159, 198, 367, 490, 493, 494, 597, 603, 604, 765, 766, 832, 861, 863, 958 = 27	63, 66, 81, 82, 108, 117, 123, 132, 144, 146, 150, 152, 154, 156, 157, 160, 345, 368, 597, 598, 599, 772 = 22
T7698K	Retained hubs	61, 66, 67, 71, 73, 75, 97, 108, 116, 120, 132, 137, 155, 370, 372, 373, 493 = 17	65, 101, 108, 127, 131, 132, 142, 144, 147, 155, 367, 370, 373, 599, 603, 604 = 16	97, 108, 116, 117, 123, 130, 142, 155, 372, 373, 598, 599, 604, 775, 797, 996 = 16	22, 62, 97, 116, 120, 125, 138, 140, 142, 144, 155, 370, 372, 373, 605, 799 = 16
	Gained hubs	62, 72, 98, 130, 131, 141, 144, 147, 156, 771, 797, 958, 996 = 13	63, 67, 76, 77, 91, 97, 116, 129, 138, 140, 141, 153, 156, 157, 231, 346, 375, 387, 595, 596,	22, 43, 58, 61, 67, 68, 70, 71, 72, 73, 81, 82, 91, 120, 131, 135, 138, 141, 143, 144, 146,	61, 63, 67, 91, 108, 123, 140, 147, 150, 151, 349, 367, 591, 598, 599, 600, 863, 958, 962 = 19

			598, 605, 799 = 23	147, 232, 370, 493, 589, 772, 862, 863, 995 = 30	

Table S5.7.2. *Degree centrality (DC)* hubs found in drug free and drug bound systems; hubs gained due to mutation, and hubs that were maintained despite the presence of mutation.

DC		Drug free		Drug bound	
Systems		Monomer A	Monomer B	Monomer A	Monomer B
WT	Hubs	72, 90, 108, 110, 113, 116, 138, 155, 159, 198, 199, 200, 201, 202, 203, 207, 218, 282, 314, 356, 435, 482, 499, 548, 576, 625, 710, 740, 778, 779, 781, 790, 808, 813, 833, 837, 928 = 37	108, 110, 155, 192, 193, 200, 202, 203, 225, 282, 335, 337, 482, 496, 499, 554, 625, 653, 654, 694, 707, 711, 729, 732, 778, 779, 781, 790, 808, 813, 833, 990 = 32	72, 108, 110, 113, 155, 157, 199, 200, 203, 206, 207, 218, 232, 313, 314, 335, 350, 351, 478, 480, 554, 605, 628, 641, 653, 654, 666, 677, 729, 732, 740, 778, 779, 781, 790, 807, 808, 813, 833, 837, 928 = 41	72, 108, 110, 113, 138, 139, 153, 155, 159, 203, 207, 225, 241, 337, 356, 478, 568, 595, 603, 605, 625, 653, 654, 694, 707, 729, 732, 778, 779, 781, 790, 807, 808, 832, 833, 928, 962 = 37
R592Q	Retained hubs	108, 110, 113, 116, 155, 159, 198, 199, 201, 202, 218, 282, 482, 499, 576, 625, 710, 740, 778, 779, 781, 790, 808, 813, 832, 833, 837, 928 = 28	108, 110, 155, 202, 203, 337, 482, 496, 554, 653, 654, 778, 781, 790, 808 = 15	72, 108, 110, 155, 199, 203, 554, 605, 707, 732, 740, 778, 779, 781, 790, 808, 833, 837, 928 = 29	108, 110, 113, 138, 155, 478, 605, 625, 707, 732, 778, 779, 781, 790, 807, 808, 833 = 17
	Gained hubs	67, 478, 480, 486, 554, 568, 595, 606, 628, 653, 707, 732, 768, 774 = 14	67, 111, 113, 116, 139, 159, 198, 199, 207, 218, 282, 344, 356, 479, 480, 493, 495, 589, 595, 672, 710, 734, 740, 767, 775, 835, 928, 961 = 28	125, 146, 193, 241, 337, 344, 403, 542, 595, 625, 665, 710 = 12	67, 154, 198, 199, 218, 282, 344, 435, 480, 482, 486, 495, 499, 554, 595, 653, 710, 740, 767, 775, 813, 816, , 837, 928 = 24
A664S	Retained hubs	108, 110, 113, 116, 155, 159, 198, 199, 218, 282, 356, 482, 499, 625, 740, 778, 779, 781, 790, 808, 833, 837, 928 = 23	108, 110, 155, 202, 203, 225, 554, 625, 778, 779, 781, 790, 808, 833 = 14	72, 108, 110, 113, 155, 199, 200, 203, 480, 628, 641, 653, 654, 732, 740, 778, 779, 781, 790, 808, 813, 833, 928 = 23	108, 110, 113, 139, 153, 155, 159, 203, 282, 568, 625, 653, 654, 778, 781, 790, 808, 833 = 28
	Gained hubs	203, 335, 431, 542, 554, 568, 589, 595, 628, 641, 732, 767, 775, 835 = 14	72, 113, 116, 151, 159, 198, 199, 201, 207, 218, 348, 493, 495, 595, 628, 721, 740, 770, 810, 813, 832, , 837 = 22	90, 122, 138, 139, 146, 159, 198, 202, 282, 335, 337, 356, 431, 479, 482, 486, 493, 495, 499, 554, 568, 691, 710, 770, 810, 835 = 26	31, 65, 72, 116, 154, 157, 198, 200, 314, 469, 480, 495, 496, 499, 542, 554, 710, 740, 767, 813, 837 = 22
G674D	Retained	108, 155, 199, 201, 203, 356,	108, 110, 155, 203, 337, 554,	108, 110, 113, 155, 200, 203,	72, 108, 110, 113, 139, 155, 159,

	hubs	625, 710, 740, 778, 779, 781, 790, 808, 813, 833, 928 = 17	625, 653, 694, 729, 732, 778, 779, 781, 790, 833 = 16	207, 218, 314, 554, 628, 653, 729, 732, 740, 778, 779, 781, 790, 807, 808, 813, 833, 928 = 24	203, 207, 356, 625, 778, 779, 781, 790, 808, 833 = 17
	Gained hubs	72, 122, 206, 403, 431, 589, 595, 628, 641, 653, 654, 657, 681, 694, 729, 732, 807, 832, 835 = 19	113, 207, 216, 314, 344, 347, 437, 480, 595, 605, 767, 810, 815, 816 = 14	122, 138, 159, 160, 194, 198, 199, 201, 202, 282, 344, 347, 437, 482, 486, 490, 495, 496, 499, 568, 589, 595, 625, 691, 710 = 25	116, 198, 199, 201, 202, 214, 218, 282, 314, 479, 480, 482, 486, 493, 495, 496, 499, 554, 721, 724, 767, 813 = 22
A721T	Retained hubs	108, 110, 113, 116, 159, 199, 203, 282, 314, 482, 499, 625, 710, 740, 778, 779, 781, 790, 808, 813, 833, 837, 928 = 22	108, 110, 155, 282, 337, 482, 499, 625, 653, 654, 778, 781, 790, 808 = 14	72, 108, 110, 155, 199, 203, 218, 554, 740, 775, 778, 779, 781, 790, 807, 808, 813, 833, 837, 928 = 20	72, 108, 110, 138, 155, 203, 356, 568, 625, 653, 778, 781, 790, 808, 928 = 15
	Gained hubs	122, 434, 478, 479, 480, 486, 554, 568, 595, 654, 662, 687, 691, 732, 792, 807 = 16	67, 113, 139, 198, 199, 201, 206, 356, 495, 595, 606, 618, 683, 691, 710, 734, 740, 767, 775, 816 = 20	90, 356, 499, 625, 718, 815 = 6	90, 201, 202, 469, 554, 606, 710, 734, 767, 775, 813, 816, 835 = 13
V732G	Retained hubs	72, 108, 113, 155, 159, 198, 199, 200, 202, 203, 218, 282, 356, 499, 576, 625, 740, 778, 779, 781, 790, 808, 833, 837, 928, 938 = 26	108, 110, 155, 200, 201, 202, 203, 282, 482, 496, 499, 554, 625, 654, 729, 778, 781, 790, 808, 833 = 20	108, 110, 155, 199, 203, 218, 554, 653, 707, 778, 779, 781, 790, 807, 808, 813, 832, 833, 928 = 19	72, 108, 110, 113, 155, 159, 203, 337, 595, 625, 653, 654, 729, 778, 779, 781, 790, 807, 808, 833 = 20
	Gained hubs	160, 194, 361, 431, 479, 493, 495, 496, 589, 595, 606, 628, 654, 767, 962 = 15	67, 113, 138, 159, 194, 195, 198, 199, 206, 218, 356, 437, 479, 480, 606, 685, 691, 710, 807, 810, 837 = 21	90, 122, 139, 198, 282, 314, 482, 548, 563, 589, 628, 665, 710, 792, 835 = 15	116, 125, 218, 238, 241, 344, 362, 464, 554, 710, 740, 770, 813, 816, 835, 837, 928 = 17
T768K	Retained hubs	90, 108, 110, 113, 138, 155, 159, 198, 199, 201, 202, 203, 218, 282, 314, 482, 499, 625, 740, 778, 779, 781, 790, 808, 813, 837, 928 = 27	108, 110, 155, 200, 202, 203, 282, 337, 482, 496, 499, 554, 625, 653, 707, 729, 732, 778, 779, 781, 790, 808, 813 = 14	108, 110, 113, 155, 199, 200, 203, 206, 207, 218, 478, 480, 499, 554, 605, 625, 628, 653, 654, 729, 732, 740, 778, 779, 781, 790, 808, 813, 837, 928 = 30	108, 110, 138, 155, 159, 203, 337, 356, 595, 605, 625, 653, 710, 729, 732, 778, 779, 781, 790, 807, 808, 832, 833, 928 =24

	Gained hubs	31, 122, 125, 156, 157, 206, 337, 344, 347, 437, 469, 478, 479, 480, 493, 495, 496, 542, 554, 568, 595, 605, 628, 653, 654, 667, 673, 691, 695, 707, 729, 732, 807, 832, 835 = 35	67, 72, 90, 93, 125, 138, 153, 194, 198, 199, 201, 206, 344, 347, 356, 437, 469, 478, 479, 480, 486, 490, 493, 495, 595, 605, 628, 667, 669, 691, 695, 710, 724, 740, 807, 835, 837 = 37	31, 116, 122, 138, 159, 198, 201, 202, 282, 337, 347, 348, 356, 437, 469, 479, 482, 486, 493, 495, 496, 568, 589, 595, 691, 695, 724, 810, 832, 835 = 30	31, 72, 93, 116, 120, 122, 125, 198, 199, 201, 202, 206, 218, 282, 314, 344, 347, 437, 469, 479, 480, 482, 490, 493, 495, 496, 499, 589, 628, 691, 695, 740, 813, 835 = 34

Table S5.7.3. *Closeness centrality (CC)* hubs found in drug free and drug bound systems; hubs gained due to mutation, and hubs that were maintained despite the presence of mutation.

.

CC		**Drug free**		**Drug bound**	
Systems		Monomer A	Monomer B	Monomer A	Monomer B
WT	Hubs	67, 71, 72, 73, 74, 75, 76, 77, 78, 79, 101, 104, 105, 150, 151, 155, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 768 = **27**	73, 76, 101, 104, 105, 146, 150, 151, 155, 596, 597, 598, 599, 600, 601, 603, 605, 768 = **18**	71, 72, 75, 76, 79, 80, 101, 102, 104, 105, 137, 140, 141, 142, 145, 146, 147, 150, 151, 152, 155, 596, 597, 598, 599, 600, 601, 604, 605, 768 = **30**	67, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 80, 101, 104, 105, 137, 140, 142, 145, 147, 149, 150, 151, 152, 155, 595, 596, 599, 602, 603, 605 = **31**
R592Q	Retained hubs	67, 71, 72, 73, 74, 75, 76, 77, 78, 79, 101, 104, 105, 150, 151, 155, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 768 = **27**	73, 76, 101, 104, 105, 146, 150, 151, 155, 596, 597, 598, 599, 600, 601, 602, 605 = **17**	71, 72, 75, 76, 79, 80, 101, 104, 105, 140, 141, 142, 145, 146, 147, 150, 151, 152, 155, 596, 597, 598, 599, 600, 601, 605 = **26**	67, 71, 72, 73, 74, 75, 76, 77, 78, 101, 104, 105, 137, 142, 145, 147, 150, 151, 152, 155, 596, 599, 605 = **23**
	Gained hubs	66, 68, 70, 80, 81, 82, 87, 98, 102, 103, 107, 137, 140, 141, 142, 145, 146, 147, 148, 149, 152, 153, 154, 156, 158, 159, 593, 595, 606, 797, 997 = **31**	67, 68, 69, 70, 71, 72, 74, 75, 77, 78, 79, 80, 81, 82, 97, 102, 140, 141, 142, 145, 147, 148, 149, 152, 153, 154, 156, 158, 595, 603, 821, 997 = **32**	67, 68, 69, 70, 73, 74, 77, 78, 81, 82, 143, 144, 148, 149, 154, 158, 595, 602, 997 = **19**	79, 102, 146, 153, 154, 597, 598, 600 = **8**
A664S	Retained hubs	67, 71, 72, 73, 74, 75, 76, 77, 78, 105, 150, 151, 155, 597, 598, 599, 603, 604, 605, 768 = **20**	73, 76, 101, 104, 105, 146, 150, 151, 155, 596, 597, 598, 599, 600, 603, 605, 768 = **16**	71, 72, 75, 76, 79, 80, 101, 102, 104, 105, 140, 141, 142, 145, 146, 147, 150, 151, 152, 155, 596, 597, 598, 599, 600, 601, 604, 605, 768 = **19**	67, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 80, 101, 104, 105, 137, 140, 142, 145, 147, 149, 150, 151, 152, 155, 595, 596, 599, 602, 603, 605 = **32**
	Gained hubs	65, 68, 69, 70, 146, 147, 154, 594, 595, 596, 769, 997 = **12**	67, 70, 71, 72, 74, 75, 77, 78, 79, 80, 81, 82, 87, 97, 98, 102, 140, 141, 142, 149, 152, 153, 595, 602, 769, 995 = **26**	65, 67, 68, 69, 70, 73, 74, 77, 78, 81, 82, 85, 86, 87, 88, 97, 98, 99, 100, 103, 106, 112, 136, 137, 138, 139, 143, 144, 148, 149, 153, 154, 156, 158, 159, 595, 602, 603, 606, 769, 797, 821, 822 = **45**	43, 64, 65, 66, 68, 79, 81, 87, 97, 98, 102, 103, 106, 141, 143, 144, 146, 148, 153, 154, 156, 157, 158, 159, 597, 598, 600, 601, 604, 606, 768, 769, 797, 798, 861, 862, 954, 956, 958, 996, 997, 998 = **42**

G674D	Retained hubs	71, 72, 73, 74, 75, 76, 77, 78, 79, 101, 104, 105, 150, 151, 155, 599, 600, 601 = 18	76, 150, 598, 599 = **4**	71, 72, 75, 76, 79, 80, 101, 102, 104, 105, 140, 141, 142, 145, 146, 147, 150, 151, 152, 155, 596, 598, 599, 600, 601, 604, 605 = **27**	67, 70, 71, 72, 73, 74, 75, 76, 77, 78, 80, 101, 104, 105, 140, 142, 147, 150, 151, 152, 155, 596, 599, 602, 603, 605 = **26**
	Gained hubs	70, 80, 81, 140, 141, 142, 147, 152, 997 = **9**	70, 71, 72, 73, 74, 75, 997 = **7**	67, 68, 69, 70, 73, 74, 77, 78, 81, 82, 97, 98, 103, 106, 149, 153, 154, 158, 602, 603, 997 = **21**	65, 68, 79, 81, 82, 97, 98, 102, 141, 153, 154, 158, 598, 600, 601, 821, 997, 998 = **18**
A721T	Retained hubs	67, 71, 72, 73, 74, 75, 76, 77, 78, 79, 101, 104, 105, 150, 151, 155, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 768 = **27**	73, 76, 101, 104, 105, 146, 150, 151, 155, 596, 597, 598, 599, 600, 601, 603, 605 = **17**	71, 72, 75, 76, 79, 80, 101, 105, 155, 596, 597, 598, 599, 601, 604 = **15**	71, 72, 74, 75, 76, 77, 78, 101, 105, 140, 142, 147, 150, 151, 152, 155, 599 = **17**
	Gained hubs	68, 69, 70, 80, 81, 82, 97, 98, 100, 102, 106, 132, 137, 140, 141, 142, 143, 145, 146, 147, 148, 149, 156, 158, 159, 594, 595, 606, 797, 821, 997 = **31**	66, 67, 68, 69, 70, 71, 72, 74, 75, 77, 78, 79, 80, 81, 82, 87, 97, 98, 102, 140, 141, 142, 143, 144, 145, 152, 153, 154, 156, 158, 159, 595, 604, 861, 958, 995, 997, 998, 999 = **39**	70, 73, 74, 77, 78, 150 = **6**	79, 598, 600, 601= **4**
V732G	Retained hubs	67, 72, 73, 75, 76, 77, 78, 101, 150, 151, 155, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605 = **21**	73, 76, 101, 104, 105, 146, 150, 151, 155, 596, 597, 598, 599, 600, 601 = **15**	71, 72, 75, 76, 79, 80, 104, 140, 142, 147, 150, 151, 155, 596, 597, 598, 599, 600, 601, 604, 605 = **21**	70, 71, 72, 73, 74, 75, 76, 77, 78, 80, 101, 104, 105, 140, 142, 145, 147, 149, 150, 151, 152, 155, 599, 603 = **22**
	Gained hubs	142, 146, 147, 152, 153, 158 = **6**	67, 68, 69, 70, 71, 72, 74, 75, 77, 78, 79, 140, 142, 147, 152, 154, 158, 595, 604 = **17**	67, 73, 74, 77, 78, 149, 603 = **7**	67, 68, 79, 81, 146, 154, 597, 598, 601, 997 = **10**
T768K	Retained hubs	67, 71, 72, 73, 74, 75, 76, 77, 78, 79, 101, 104, 105, 150, 151, 155, 596, 597, 598, 599, 600, 601, 602, 604, 605, 768 = **26**	73, 76, 101, 104, 105, 146, 150, 151, 155, 596, 597, 598, 599, 600, 601, 603, 605, 768 = **18**	71, 72, 75, 76, 79, 80, 101, 104, 105, 140, 141, 142, 146, 147, 150, 151, 152, 155, 595, 596, 597, 598, 599, 600, 601, 604, 605, 768 = **28**	67, 70, 71, 72, 73, 74, 75, 76, 77, 78, 80, 101, 104, 140, 142, 145, 147, 149, 150, 151, 152, 155, 595, 596, 599, 600, 602, 603, 605 = **29**

	Gained hubs	68, 69, 70, 80, 81, 82, 97, 98, 100, 102, 103, 106, 137, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 152, 153, 154, 156, 158, 159, 595, 769, 797, 821, 822, 997 = **35**	67, 68, 69, 70, 71, 72, 74, 75, 77, 78, 79, 80, 81, 82, 85, 87, 97, 98, 102, 106, 137, 138, 140, 141, 142, 143, 144, 145, 147, 148, 149, 152, 153, 154, 156, 158, 159, 593, 594, 595, 602, 604, 606, 769, 997 = **45**	67, 68, 69, 70, 73, 74, 77, 78, 81, 82, 97, 98, 158, 602, 603, 995, 997, 998, 999 = **19**	68, 79, 81, 82, 97, 141, 144, 146, 158, 597, 598, 601, 604, 768, 769, 997 = **16**

Table S5.7.4. *Eigen centrality (EC)* hubs found in drug free and drug bound systems; hubs gained due to mutation, and hubs that were maintained despite the presence of mutation.

EC		Drug free		Drug bound	
Systems		Monomer A	Monomer B	Monomer A	Monomer B
WT	Hubs	778 = 1	121, 122, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 214, 216, 217, 225, 248, 253, 280, 281, 282, 283, 469, 477, 478, 479, 480, 482, 483, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 502, 503, 734, 775, 778, 779, 781, 790, 808, 810 = 57	103, 104, 108, 109, 110, 111, 112, 113, 114, 115, 159, 665, 666, 667, 707, 708, 710, 717, 721, 724, 729, 732, 733, 734, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 790, 791, 792, 796, 799, 802, 803, 804, 805, 806, 807, 808, 809, 810, 813, 829, 832, 833, 834, 835, 836, 837, 852, 926, 928 = 61	929, 595, 710, 732, 734, 740, 741, 744, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810 = 20
R592Q	Retained hubs	778 = 1	194, 197, 198, 199, 200, 201, 202, 203, 280, 281, 282, 469, 477, 478, 479, 480, 482, 483, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 734, 775, 778, 790 = 32	775, 776, 778, 779, 781, 782, 790, 805, 808, 809, 810 = 11	710, 732, 733, 734, 774, 775, 777, 778, 779, 781, 782, 790, 792, 808, 809, 810 = 16
	Gained hubs	734, 770, 771, 773, 774, 775, 776, 777, 779, 781, 782, 790, 792, 805, 808, 810 = 16	22, 481, 486, 740, 774, 777, 792 = 7	595, 605, 606, 740, 741, 769 = 6	549, 595, 605, 606, 707, 708, 717, 718, 721, 724, 735, 765, 766, 767, 768, 769, 770, 771, 772, 773, 776, 780, 791, 793, 794, 795, 796, 802, 803, 804, 805, 806, 807, 811, 813, 814, 815, 816, 926, 928, 929 = 41
A664S	Retained hubs	778 = 1	734, 775, 778, 779, 781, 790, 808, 810 = 8	72, 108, 110, 113, 155, 199, 200, 203, 480, 628, 641, 653, 654, 732, 740, 778, 779, 781, 790, 808, 813, 833, 928 = 23	108, 110, 113, 139, 153, 155, 159, 203, 282, 568, 625, 653, 654, 778, 781, 790, 808, 833 = 18

	Gained hubs	100, 103, 108, 109, 110, 111, 112, 113, 114, 115, 116, 159, 732, 734, 772, 774, 775, 776, 777, 779, 781, 782, 790, 791, 792, 799, 802, 803, 804, 805, 806, 807, 808, 809, 810, 813, 829, 832, 833, 834, 835, 836, 837, 852, 926, 928 = 46	100, 103, 104, 108, 109, 110, 111, 112, 113, 114, 115, 116, 155, 159, 740, 741, 744, 774, 776, 777, 782, 792, 799, 804, 805, 809, 829, 832, 833, 836 = 30	92, 93, 120, 121, 122, 125, 126, 127, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 214, 216, 218, 225, 244, 248, 253, 280, 281, 282, 283, 284, 461, 462, 464, 469, 470, 471, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 502, 503, 504 = 69	31 =1
G674D	Retained hubs	778 = 1	734, 775, 778, 779, 780, 781, 808, 810 = 8	108, 110, 113, 155, 200, 203, 207, 218, 314, 554, 628, 653, 729, 732, 740, 778, 779, 781, 790, 807, 808, 813, 833, 928 = 12	72, 108, 110, 113, 139, 155, 159, 203, 207, 356, 625, 778, 779, 781, 790, 808, 833
	Gained hubs	595, 605, 606, 666, 707, 708, 709, 710, 711, 712, 717, 718, 720, 721, 724, 729, 731, 732, 733, 734, 735, 738, 740, 741, 765, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 779, 780, 781, 782, 783, 784, 790, 791, 792, 793, 794, 795, 796, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 813, 926, 928, 929, 932 = 64	595, 605, 606, 710, 717, 721, 732, 733, 740, 741, 742, 743, 744, 750, 765, 767, 768, 769, 770, 771, 772, 773, 774, 776, 777, 782, 792, 794, 805, 806, 807, 790, 791, 809, 813, 814, 815, 816, 926, 928 = 40	31 = 1	92, 93, 120, 122, 126, 127, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 214, 216, 280, 281, 282, 283, 284, 462, 464, 469, 471, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 502, 503 - 79
A721T	Retained hubs	778 = 1	734, 775, 778, 779, 781, 790, 808, 810 = 8	710, 717, 721, 732, 733, 734, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 790, 791, 792, 796, 802, 803, 804, 805, 806, 807, 808, 809, 810, 813, 926, 928 = 34	605, 606, 712, 721, 735, 738, 739, 740, 743, 750, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 776, 791, 793, 794, 795, 796, 802, 804, 806, 807, 813, 816, 817, 926, 928 = 36
	Gained hubs	595, 707, 708, 710, 711, 712, 717, 718, 720, 721, 724, 732, 733, 734, 735, 738, 740, 765,	593, 595, 605, 606, 607, 608, 672, 683, 708, 710, 711, 712, 713, 717, 721, 732, 733, 735,	595, 606, 712, 718, 735, 738, 740, 741, 764, 765, 767, 768, 769, 770, 783, 784, 793, 794, 795, 929 = 10	595, 710, 732, 734, 741, 744, 775, 777, 778, 779, 780, 781, 782, 790, 792, 805, 808, 809, 810 = 19

		767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 779, 780, 781, 782, 783, 790, 791, 792, 793, 794, 795, 796, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 813, 926, 928, 929 = 55	736, 737, 738, 739, 740, 741, 742, 743, 744, 762, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 776, 777, 780, 782, 791, 792, 793, 794, 796, 804, 805, 806, 807, 809, 813, 816, 926, 928 = 58		
V732G	Retained hubs	72, 108, 113, 155, 159, 198, 199, 200, 202, 203, 218, 282, 356, 499, 576, 625, 740, 778, 779, 781, 790, 808, 833, 837, 928, 938 = 26	121, 122, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 214, 216, 217, 225, 248, 253, 280, 281, 282, 283, 469, 477, 478, 479, 480, 482, 483, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 502, 503 = 67	734, 772, 773, 774, 775, 776, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810 = 17	790, 792, 805, 808, 809, 810 = 6
	Gained hubs	198, 199, 201, 202, 282, 479, 492, 493, 495, 496, 499 = 11	218, 284, 462, 464, 481, 484, 48 = 7	595, 605, 740, 741, 769, 929 = 6	103, 108, 109, 110, 111, 112, 113, 159, 547, 548, 549, 572, 595, 605, 606, 666, 667, 707, 708, 709, 710, 712, 717, 718, 721, 724, 729, 731, 732, 733, 734, 735, 738, 740, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 783, 791, 793, 794, 795, 796, 799, 801, 802, 803, 804, 806, 807, 811, 812, 813, 814, 815, 816, 829, 832, 833, 836, 837, 926, 928, 929 = 79
T768K	Retained hubs	90, 108, 110, 113, 138, 155, 159, 198, 199, 201, 202, 203, 218, 282, 314, 482, 499, 625, 740, 778, 779, 781, 790, 808, 813, 837, 928 = 27	121, 122, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 216, 225, 248, 280, 281, 282, 283, 469, 477, 478, 479, 480, 482, 483, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 502,	108, 110, 113, 155, 199, 200, 203, 206, 207, 218, 478, 480, 499, 554, 605, 625, 628, 653, 654, 729, 732, 740, 778, 779, 781, 790, 808, 813, 837, 928 = 30	108, 110, 138, 155, 159, 203, 337, 356, 595, 605, 625, 653, 710, 729, 732, 778, 779, 781, 790, 807, 808, 832, 833, 928 = 12

			503 = 43		
	Gained hubs	31, 198, 201, 202, 282, 479, 480, 492, 493, 495, 496, 499 = 12	92, 93, 120, 123, 125, 126, 227, 284, 464, 481, 484, 485, 486, 488, 489, 490 = 36	31, 122, 196, 197, 198, 199, 200, 201, 202, 203, 280, 282, 469, 478, 479, 480, 482, 483, 486, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500 = 30	93, 120, 121, 122, 192, 194, 196, 197, 198, 199, 200, 201, 202, 203, 205, 206, 280, 281, 282, 469, 477, 478, 479, 480, 482, 483, 486, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500 = 39

Table S5.7.5. *Katz centrality (KC)* hubs found in drug free and drug bound systems; hubs gained due to mutation, and hubs that were maintained despite the presence of mutation.

KC		Drug free		Drug bound	
Systems		Monomer A	Monomer B	Monomer A	Monomer B
WT	Hubs	103, 104, 108, 109, 110, 111, 112, 113, 199, 203, 313, 350, 351, 605, 640, 641, 665, 666, 667, 707, 710, 732, 734, 775, 776, 777, 778, 779, 781, 782, 790, 792, 805, 807, 808, 809, 810, 813, 832, 833, 836, 837, 928 = **42**	108, 110, 111, 155, 159, 595, 707, 708, 710, 732, 734, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 832, 833 = **25**	103, 104, 108, 110, 111, 112, 113, 116, 155, 159, 198, 199, 201, 202, 203, 282, 479, 496, 499, 710, 732, 734, 775, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 832, 833 = **35**	108, 110, 111, 155, 192, 193, 194, 195, 196, 198, 199, 200, 201, 202, 203, 206, 280, 282, 479, 480, 492, 493, 495, 496, 499, 500, 707, 710, 729, 732, 734, 774, 775, 776, 777, 778, 779, 781, 782, 790, 792, 805, 808, 810, 832, 833 = **23**
R592Q	Retained hubs	108, 110, 111, 203, 605, 710, 732, 734, 775, 776, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 832, 833, 836 = **2**	108, 110, 111, 155, 595, 710, 732, 734, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 813, 832, 836 = **23**	103, 108, 110, 111, 112, 113, 159, 198, 199, 201, 202, 282, 479, 496, 499, 732, 734, 775, 776, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 832, 833 = **33**	108, 110, 111, 155, 198, 199, 201, 202, 203, 282, 479, 480, 492, 493, 495, 496, 499, 710, 732, 734, 774, 775, 776, 777, 778, 779, 781, 790, 792, 805, 808, 810, 832, 833 = **34**
	Gained hubs	200, 595 = **2**	67, 198, 199, 201, 202, 282, 478, 479, 480, 482, 486, 492, 493, 495, 496, 499, 605, 606, 721, 767, 769, 770, 774, 796, 816 = **24**	280, 478, 480, 482, 486, 492, 495, 595, 606, 768, 770, 771, 772, 773, 774, 777 = **36**	103, 104, 109, 112, 113, 116, 159, 469, 478, 482, 483, 595, 735, 740, 741, 765, 770, 773 = **38**
A664S	Retained hubs	108, 110, 111, 112, 113, 199, 203, 710, 732, 734, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 810, 832, 833 = **23**	108, 110, 155, 159, 734, 775, 777, 778, 781, 805, 808, 810 = **12**	103, 108, 110, 111, 112, 113, 116, 159, 198, 282, 499, 732, 734, 775, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 832, 833 = **26**	108, 110, 111, 155, 198, 199, 200, 201, 202, 203, 493, 495, 496, 499, 732, 734, 775, 776, 777, 778, 779, 781, 782, 790, 792, 805, 808, 810, 832, 833 = **30**
	Gained hubs	104, 122, 155, 159, 196, 197, 198, 200, 201, 202, 206, 280, 282, 469, 478, 479, 480, 482, 483, 486,	67, 72, 104, 113, 153, 154, 157, 158, 198, 202, 282, 479, 480, 495, 496, 499, 790, 813 = **18**	740, 776, 777, 813, 829, 836, 837 = **7**	100, 103, 104, 109, 112, 113, 116, 152, 159, 595, 740, 770, 809, 813 = **7**

		492, 493, 494, 495, 496, 497, 498, 499, 500, 735, 774 = **31**			
G674D	Retained hubs	108, 110, 111, 112, 113, 199, 203, 666, 710, 732, 734, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 833 = **12**	108, 110, 111, 155, 159, 775, 777, 778, 779, 781, 790, 805, 808, 810, 832, 833 = **16**	108, 110, 111, 155, 201, 710, 732, 734, 775, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 832, 833 = **21**	108, 110, 203, 732, 734, 774, 775, 776, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 833 = **10**
	Gained hubs	122, 155, 159, 194, 196, 197, 198, 200, 201, 202, 282, 479, 480, 482, 492, 493, 495, 496, 499, 500, 554, 572, 774 = **23**	112, 113, 122, 196, 197, 198, 199, 201, 202, 203, 280, 282, 469, 478, 479, 480, 482, 483, 486, 491, 492, 493, 494, 495, 496, 499, 500, 734, 774 = **27**	641, 707, 721, 724, 729, 770, 772, 773, 774, 776, 777, 813, 928 = **13**	549, 554, 572, 595, 605, 767, 813, 815 = **8**
A721T	Retained hubs	108, 110, 111, 199, 734, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 832, 833 = **39**	108, 110, 111, 155, 159, 595, 707, 710, 732, 734, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 832, 833 = **24**	108, 110, 111, 112, 113, 116, 159, 199, 282, 479, 499, 710, 732, 734, 775, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 832, 833 = **27**	108, 110, 155, 198, 199, 282, 479, 480, 495, 499, 710, 732, 734, 774, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 810, 833 = **26**
	Gained hubs	155, 197, 198, 282, 482 = **5**	109, 112, 113, 116, 125, 572, 605, 606, 641, 666, 766, 767, 769, 770, 774, 794, 796, 813, 814, 816 = **20**	478, 480, 482, 495, 771, 772, 773, 774, 776, 777, 813 = **11**	113, 482, 595, 606, 738, 739, 740, 741, 764, 765, 767, 770, 809 = **13**
V732G	Retained hubs	108, 110, 111, 155, 734, 775, 777, 778, 779, 781, 782, 790, 792, 805, 808, 809, 810, 832, 833 = **19**	794, 796 = **2**	108, 155, 159, 198, 199, 201, 202, 203, 282, 479, 496, 499, 734, 775, 778, 779, 781, 790, 792, 808, 810, 833 = **22**	108, 110, 155, 192, 194, 195, 196, 198, 199, 200, 201, 202, 203, 206, 280, 282, 479, 480, 492, 493, 495, 496, 499, 500, 710, 732, 734, 775, 778, 779, 781, 782, 790, 808, 810, 833 = **38**
	Gained hubs	197, 198, 199, 282, 482 = **5**	100, 103, 104, 108, 109, 110, 111, 112, 113, 116, 125, 155, 159, 572, 595, 605, 606, 641, 666, 707, 710, 732, 734, 766, 767, 769, 770, 774, 775, 777,	194, 197, 200, 280, 480, 492, 493, 495, 500, 777 = **10**	159, 216, 478, 482, 666, 708 = **6**

			778, 779, 781, 782, 813, 814, 816, 832, 833 = **39**		
T768K	Retained hubs	103, 108, 109, 110, 111, 112, 113, 199, 203, 732, 734, 775, 778, 779, 781, 790, 808, 810, 832, 833 = **22**	108, 110, 111, 155, 159, 708, 710, 732, 734, 778, 781, 790, 808, 810, 833 = **15**	90, 108, 110, 113, 138, 155, 159, 198, 199, 201, 202, 203, 218, 282, 314, 482, 499, 625, 740, 778, 779, 781, 790, 808, 813, 837, 928 = **27**	108, 110, 155, 200, 202, 203, 282, 337, 482, 496, 499, 554, 625, 653, 707, 729, 732, 778, 779, 781, 790, 808, 813 = **23**
	Gained hubs	93, 100, 116, 120, 122, 123, 126, 155, 156, 159, 160, 197, 198, 200, 201, 202, 206, 280, 282, 469, 478, 479, 480, 482, 483, 486, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 691 = **38**	93, 112, 113, 116, 120, 122, 194, 196, 197, 198, 199, 200, 201, 202, 203, 206, 280, 282, 469, 478, 479, 480, 482, 483, 486, 490, 492, 493, 494, 495, 496, 497, 498, 499, 500, 691 = **36**	31, 122, 125, 156, 157, 206, 337, 344, 347, 437, 469, 478, 479, 480, 493, 495, 496, 542, 554, 568, 595, 605, 628, 653, 654, 667, 673, 691, 695, 707, 729, 732, 807, 832, 835 = **35**	67, 72, 90, 93, 125, 138, 153, 194, 198, 199, 201, 206, 344, 347, 356, 437, 469, 478, 479, 480, 486, 490, 493, 495, 595, 605, 628, 667, 669, 691, 695, 710, 724, 740, 807, 835, 837 = **37**

REFERENCES

1. Dorling D (2021) World population prospects at the UN: our numbers are not our problem? *The Struggle for Social Sustainability*, (Policy Press), pp 129-154.
2. Manufacturers IFoP & Associations (2017) The Pharmaceutical Industry and Global Health. Facts and Figures 2017.
3. Kinch MS, Haynesworth A, Kinch SL, & Hoyer D (2014) An overview of FDA-approved new molecular entities: 1827–2013. *Drug discovery today* 19(8):1033-1039.
4. Schmid EF, James K, & Smith DA (2001) The impact of technological advances on drug discovery today. *Drug information journal: DIJ/Drug Information Association* 35(1):41-45.
5. VC Guido R, Oliva G, & D Andricopulo A (2011) Modern drug discovery technologies: opportunities and challenges in lead discovery. *Combinatorial chemistry & high throughput screening* 14(10):830-839.
6. T Issa N, Wathieu H, Ojo A, W Byers S, & Dakshanamurthy S (2017) Drug metabolism in preclinical drug development: a survey of the discovery process, toxicology, and computational tools. *Current drug metabolism* 18(6):556-565.
7. Zhang Z & Tang W (2018) Drug metabolism in drug discovery and development. *Acta Pharmaceutica Sinica B* 8(5):721-732.
8. Patwardhan B & Chaguturu R (2016) Innovative approaches in drug discovery: Ethnopharmacology, systems biology and holistic targeting.
9. Ahmed S, Zhou Z, Zhou J, & Chen S-Q (2016) Pharmacogenomics of drug metabolizing enzymes and transporters: relevance to precision medicine. *Genomics, proteomics & bioinformatics* 14(5):298-313.
10. Lennard L (2001) Therapeutic drug monitoring of cytotoxic drugs. *British Journal of Clinical Pharmacology* 52(S1):75-87.
11. Bertholee D, Maring JG, & van Kuilenburg AB (2017) Genotypes affecting the pharmacokinetics of anticancer drugs. *Clinical pharmacokinetics* 56(4):317-337.
12. Biemar F & Foti M (2013) Global progress against cancer—challenges and opportunities. *Cancer biology & medicine* 10(4):183.
13. Prager GW*, et al.* (2018) Global cancer control: responding to the growing burden, rising costs and inequalities in access. *ESMO open* 3(2):e000285.
14. Fitzmaurice C*, et al.* (2018) Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2016: a systematic analysis for the global burden of disease study. *JAMA oncology* 4(11):1553-1568.
15. Bray F*, et al.* (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* 68(6):394-424.
16. Adedeji OA (2017) *Cancer in Sub-Saharan Africa: current practice and future* (Springer).
17. Ferlay J*, et al.* (2019) Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *International journal of cancer* 144(8):1941-1953.
18. Gouda HN*, et al.* (2019) Burden of non-communicable diseases in sub-Saharan Africa, 1990–2017: results from the Global Burden of Disease Study 2017. *The Lancet Global Health* 7(10):e1375-e1387.

19. Compare G (2016) Institute of health metrics and evaluation. *Washington: University of Washington*.
20. Ngwa W*, et al.* (2022) Cancer in sub-Saharan Africa: a Lancet Oncology Commission. *The Lancet Oncology*.
21. Ferlay J*, et al.* (2021) Cancer statistics for the year 2020: An overview. *International journal of cancer* 149(4):778-789.
22. Siegel RL, Miller KD, Fuchs HE, & Jemal A (2022) Cancer statistics, 2022. *CA: a cancer journal for clinicians* 72(1):7-33.
23. Pramesh C*, et al.* (2022) Priorities for cancer research in low-and middle-income countries: a global perspective. *Nature Medicine* 28(4):649-657.
24. Trieu PDY, Mello-Thoms CR, Barron ML, & Lewis SJ (2023) Look how far we have come: BREAST cancer detection education on the international stage. *Frontiers in oncology* 12:1023714.
25. Pace LE & Shulman LN (2016) Breast cancer in sub-Saharan Africa: challenges and opportunities to reduce mortality. *The oncologist* 21(6):739-744.
26. Martei YM, Dauda B, & Vanderpuye V (2022) Breast cancer screening in sub-Saharan Africa: a systematic review and ethical appraisal. *BMC cancer* 22(1):203.
27. de Martel C, Georges D, Bray F, Ferlay J, & Clifford GM (2020) Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. *The Lancet Global Health* 8(2):e180-e190.
28. Rotimi SO, Rotimi OA, & Salhia B (2021) A review of cancer genetics and genomics studies in Africa. *Frontiers in Oncology* 10:606400.
29. Tishkoff SA*, et al.* (2009) The genetic structure and history of Africans and African Americans. *science* 324(5930):1035-1044.
30. Lachance J*, et al.* (2012) Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers. *Cell* 150(3):457-469.
31. Hamdi Y*, et al.* (2021) Human OMICs and computational biology research in Africa: current challenges and prospects. *Omics: a journal of integrative biology* 25(4):213-233.
32. Tucci S & Akey JM (2019) The long walk to African genomics. *Genome biology* 20(1):1-3.
33. Campbell MC, Hirbo JB, Townsend JP, & Tishkoff SA (2014) The peopling of the African continent and the diaspora into the new world. *Current opinion in genetics & development* 29:120-132.
34. Gurdasani D*, et al.* (2015) The African genome variation project shapes medical genetics in Africa. *Nature* 517(7534):327-332.
35. Skoglund P*, et al.* (2017) Reconstructing prehistoric African population structure. *Cell* 171(1):59-71. e21.
36. Schlebusch CM*, et al.* (2017) Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. *Science* 358(6363):652-655.
37. Gurdasani D, Barroso I, Zeggini E, & Sandhu MS (2019) Genomics of disease risk in globally diverse populations. *Nature Reviews Genetics* 20(9):520-535.
38. Huang T, Shu Y, & Cai Y-D (2015) Genetic differences among ethnic groups. *BMC genomics* 16(1):1-10.
39. Caswell-Jin JL*, et al.* (2018) Racial/ethnic differences in multiple-gene sequencing results for hereditary cancer risk. *Genetics in Medicine* 20(2):234-239.
40. Jing L, Su L, & Ring BZ (2014) Ethnic background and genetic variation in the evaluation of cancer risk: a systematic review. *PloS one* 9(6):e97522.
41. Lilyquist J, Ruddy KJ, Vachon CM, & Couch FJ (2018) Common Genetic Variation and Breast Cancer Risk—Past, Present, and FutureCommon Genetic Variation and Breast Cancer Risk. *Cancer Epidemiology, Biomarkers & Prevention* 27(4):380-394.

42. Ahmed Z, Zeeshan S, Mendhe D, & Dong X (2020) Human gene and disease associations for clinical-genomics and precision medicine research. *Clinical and translational medicine* 10(1):297-318.
43. MacArthur J*, et al.* (2017) The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic acids research* 45(D1):D896-D901.
44. Hindorff LA*, et al.* (2018) Prioritizing diversity in human genomics research. *Nature Reviews Genetics* 19(3):175-185.
45. Pertea M (2012) The human transcriptome: an unfinished story. *Genes* 3(3):344-360.
46. Ohno S (1972) So much'junk'DNA in our genome. *Evolution of Genetic Systems, Brookhaven Symp. Biol.*, pp 366-370.
47. Mattick JS (2011) The central role of RNA in human development and cognition. *FEBS letters* 585(11):1600-1616.
48. Wilkins MR*, et al.* (1996) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnology and genetic engineering reviews* 13(1):19-50.
49. Holman JD, Dasari S, & Tabb DL (2013) Informatics of protein and posttranslational modification detection via shotgun proteomics. *Proteomics for Biomarker Discovery*, (Springer), pp 167-179.
50. Hanash S (2003) Disease proteomics. *Nature* 422(6928):226-232.
51. Alberts B (1998) The cell as a collection of protein machines: preparing the next generation of molecular biologists. *cell* 92(3):291-294.
52. Ho Y*, et al.* (2002) Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature* 415(6868):180-183.
53. Banks RE*, et al.* (2000) Proteomics: new perspectives, new biomedical opportunities. *The Lancet* 356(9243):1749-1756.
54. Dunn MJ (2000) Studying heart disease using the proteomic approach. *Drug discovery today* 5(2):76-84.
55. Kustatscher G*, et al.* (2022) Understudied proteins: opportunities and challenges for functional proteomics. *Nature Methods*:1-6.
56. Jungbauer A & Hahn R (2009) Ion-exchange chromatography. *Methods in enzymology* 463:349-371.
57. Tastan Bishop Ö & Kroon M (2011) Study of protein complexes via homology modeling, applied to cysteine proteases and their protein inhibitors. *Journal of molecular modeling* 17(12):3163-3172.
58. Lieberman RL (2019) How does a protein's structure spell the difference between health and disease? Our journey to understand glaucoma-associated myocilin. *PLoS Biology* 17(4):e3000237.
59. Bishop ÖT, Musyoka TM, & Barozi V (2022) Allostery and missense mutations as intermittently linked promising aspects of modern computational drug discovery. *Journal of Molecular Biology*:167610.
60. Musyoka T & Bishop ÖT (2019) South African Abietane diterpenoids and their analogs as potential antimalarials: novel insights from hybrid computational approaches. *Molecules* 24(22):4036.
61. Kumar B, Prakash A, Ruhela RK, & Medhi B (2014) Potential of metabolomics in preclinical and clinical drug development. *Pharmacological Reports* 66(6):956-963.
62. Tian X, Li S, Liu Y, & Liu X (2016) Transcriptomic profiling reveals metabolic and regulatory pathways in the desiccation tolerance of Mungbean (Vigna radiata [L.] R. Wilczek). *Frontiers in plant science* 7:1921.

63. Dubey D*, et al.* (2018) NMR-based serum metabolomics revealed distinctive metabolic patterns in reactive arthritis compared with rheumatoid arthritis. *Journal of proteome research* 18(1):130-146.
64. Zhou J*, et al.* (2016) Development and evaluation of a parallel reaction monitoring strategy for large-scale targeted metabolomics quantification. *Analytical chemistry* 88(8):4478-4486.
65. Schilling B*, et al.* (2015) Multiplexed, scheduled, high-resolution parallel reaction monitoring on a full scan QqTOF instrument with integrated data-dependent and targeted mass spectrometric workflows. *Analytical chemistry* 87(20):10222-10229.
66. Garcia-Perez I*, et al.* (2020) Identifying unknown metabolites using NMR-based metabolic profiling techniques. *Nature Protocols* 15(8):2538-2567.
67. Ma W*, et al.* (2019) Comparative transcriptome and metabolic profiling analysis of buckwheat (Fagopyrum tataricum (L.) Gaertn.) under salinity stress. *Metabolites* 9(10):225.
68. Ashrafian H*, et al.* (2021) Metabolomics: The stethoscope for the twenty-first century. *Medical Principles and Practice* 30(4):301-310.
69. Beger RD*, et al.* (2016) Metabolomics enables precision medicine:"a white paper, community perspective". *Metabolomics* 12(9):1-15.
70. Shah SH & Newgard CB (2015) Integrated metabolomics and genomics: systems approaches to biomarkers and mechanisms of cardiovascular disease. *Circulation: cardiovascular genetics* 8(2):410-419.
71. Sun L, Zou L-X, & Chen M-J (2017) Make precision medicine work for chronic kidney disease. *Medical Principles and Practice* 26(2):101-107.
72. McGrath S (2018) THE INFLUENCE OF 'OMICS'IN SHAPING PRECISION MEDICINE. *INNOVATIONS*.
73. Ota M & Fujio K (2021) Multi-omics approach to precision medicine for immune-mediated diseases. *Inflammation and Regeneration* 41(1):1-6.
74. Hu C & Jia W (2021) Multi-omics profiling: the way toward precision medicine in metabolic diseases. *Journal of Molecular Cell Biology* 13(8):576-593.
75. Collins FS & Varmus H (2015) A new initiative on precision medicine. *New England journal of medicine* 372(9):793-795.
76. Hasanzad M*, et al.* (2021) Precision medicine journey through omics approach. *Journal of Diabetes & Metabolic Disorders*:1-8.
77. Ahmed Z*, et al.* (2021) Integrative clinical, genomics and metabolomics data analysis for mainstream precision medicine to investigate COVID-19. *BMJ innovations* 7(1).
78. Ahmed Z (2020) Practicing precision medicine with intelligently integrative clinical and multi-omics data analysis. *Human genomics* 14(1):1-5.
79. Weinshilboum RM & Wang L (2017) Pharmacogenomics: precision medicine and drug response. *Mayo Clinic Proceedings*, (Elsevier), pp 1711-1722.
80. Oscarson M (2003) Pharmacogenetics of drug metabolising enzymes: importance for personalised medicine.
81. Lamb DC, Waterman MR, Kelly SL, & Guengerich FP (2007) Cytochromes P450 and drug discovery. *Current opinion in biotechnology* 18(6):504-512.
82. Jancova PI (drug metabolizing enzymes, Biomed. *Pap. Med. Fac. Univ. Palacký Olomouc Czechoslov* (154):103.
83. Kimura S, Umeno M, Skoda R, Meyer U, & Gonzalez F (1989) The human debrisoquine 4-hydroxylase (CYP2D) locus: sequence and identification of the polymorphic CYP2D6 gene, a related gene, and a pseudogene. *American journal of human genetics* 45(6):889.
84. Gough AC*, et al.* (1993) Localization of the CYP2D gene locus to human chromosome 22q13. 1 by polymerase chain reaction, in situ hybridization, and linkage analysis. *Genomics* 15(2):430-432.

85. Evans WE & Relling MV (1999) Pharmacogenomics: translating functional genomics into rational therapeutics. *science* 286(5439):487-491.
86. Zanger UM & Schwab M (2013) Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacology & therapeutics* 138(1):103-141.
87. Carvalho Henriques B*, et al.* (2020) How can drug metabolism and transporter genetics inform psychotropic prescribing? *Frontiers in Genetics* 11:491895.
88. van der Weide J & Steijns LS (1999) Cytochrome P450 enzyme system: genetic polymorphisms and impact on clinical pharmacology. *Annals of clinical biochemistry* 36(6):722-729.
89. Ingelman-Sundberg M, Sim SC, Gomez A, & Rodriguez-Antona C (2007) Influence of cytochrome P450 polymorphisms on drug therapies: pharmacogenetic, pharmacoepigenetic and clinical aspects. *Pharmacology & therapeutics* 116(3):496-526.
90. Johansson I & Ingelman-Sundberg M (2011) Genetic polymorphism and toxicology—with emphasis on cytochrome p450. *Toxicological sciences* 120(1):1-13.
91. Sim E, Abuhammad A, & Ryan A (2014) Arylamine N-acetyltransferases: from drug metabolism and pharmacogenetics to drug discovery. *British Journal of Pharmacology* 171(11):2705-2725.
92. Szumlanski C*, et al.* (1996) Thiopurine methyltransferase pharmacogenetics: human gene cloning and characterization of a common polymorphism. *DNA and cell biology* 15(1):17-30.
93. Peregud-Pogorzelski J*, et al.* (2011) Thiopurine S-methyltransferase (TPMT) polymorphisms in children with acute lymphoblastic leukemia, and the need for reduction or cessation of 6-mercaptopurine doses during maintenance therapy: The Polish multicenter analysis. *Pediatric blood & cancer* 57(4):578-582.
94. Elion GB, Burgi E, & Hitchings GH (1952) Studies on condensed pyrimidine systems. IX. The synthesis of some 6-substituted purines. *Journal of the American Chemical Society* 74(2):411-414.
95. Burchenal JH*, et al.* (1953) Clinical evaluation of a new antimetabolite, 6-mercaptopurine, in the treatment of leukemia and allied diseases. *Blood* 8(11):965-999.
96. Silverman LB & Sallan SE (2003) Newly diagnosed childhood acute lymphoblastic leukemia: update on prognostic factors and treatment. *Current opinion in hematology* 10(4):290-296.
97. Pui C-H, Campana D, & Evans WE (2001) Childhood acute lymphoblastic leukaemia–current status and future perspectives. *The lancet oncology* 2(10):597-607.
98. McLeod H, Krynetski E, Relling M, & Evans W (2000) Genetic polymorphism of thiopurine methyltransferase and its clinical relevance for childhood acute lymphoblastic leukemia. *Leukemia* 14(4):567-572.
99. Relling MV, Hancock ML, Boyett JM, Pui C-H, & Evans WE (1999) Prognostic importance of 6-mercaptopurine dose intensity in acute lymphoblastic leukemia. *Blood, The Journal of the American Society of Hematology* 93(9):2817-2823.
100. Krynetski EY*, et al.* (1996) Genetic polymorphism of thiopurine S-methyltransferase: clinical importance and molecular mechanisms. *Pharmacogenetics* 6(4):279-290.
101. Remy CN (1963) Metabolism of thiopyrimidines and thiopurines: S-methylation with S-adenosylmethionine transmethylase and catabolism in mammalian tissues. *Journal of Biological Chemistry* 238(3):1078-1084.
102. Schaeffeler E*, et al.* (2004) Comprehensive analysis of thiopurine S-methyltransferase phenotype–genotype correlation in a large population of German-Caucasians and identification of novel TPMT variants. *Pharmacogenetics and Genomics* 14(7):407-417.
103. Indjova D*, et al.* (2003) Phenotypic and genotypic analysis of thiopurine S-methyltransferase polymorphism in the Bulgarian population. *Therapeutic drug monitoring* 25(5):631-636.

104. Fakhoury M*, et al.* (2007) Should TPMT genotype and activity be used to monitor 6-mercaptopurine treatment in children with acute lymphoblastic leukaemia? *Journal of clinical pharmacy and therapeutics* 32(6):633-639.
105. Evans WE (2004) Pharmacogenetics of thiopurine S-methyltransferase and thiopurine therapy. *Therapeutic drug monitoring* 26(2):186-191.
106. Relling MV*, et al.* (1999) Mercaptopurine therapy intolerance and heterozygosity at the thiopurine S-methyltransferase gene locus. *Journal of the National Cancer Institute* 91(23):2001-2008.
107. Group JCO (1999) National Cancer Institute-common toxicity criteria. *Jpn J Cancer Chemother* 26:1084-1144.
108. Stanulla M*, et al.* (2005) Thiopurine methyltransferase (TPMT) genotype and early treatment response to mercaptopurine in childhood acute lymphoblastic leukemia. *Jama* 293(12):1485-1489.
109. Lennard L, Cartwright CS, Wade R, & Vora A (2015) Thiopurine methyltransferase and treatment outcome in the UK acute lymphoblastic leukaemia trial ALL 2003. *British journal of haematology* 170(4):550-558.
110. Innocenti F, Iyer L, & Ratain M (2000) Pharmacogenetics-A tool for individualising antineoplastic therapy. *Clinical Pharmacokinetics* 39(5):315-325.
111. Tai H-L*, et al.* (1999) Enhanced proteasomal degradation of mutant human thiopurine S-methyltransferase (TPMT) in mammalian cells: mechanism for TPMT protein deficiency inherited by TPMT* 2, TPMT* 3A, TPMT* 3B or TPMT* 3C. *Pharmacogenetics* 9(5):641-650.
112. Tai H-L*, et al.* (1996) Thiopurine S-methyltransferase deficiency: two nucleotide transitions define the most prevalent mutant allele associated with loss of catalytic activity in Caucasians. *American journal of human genetics* 58(4):694.
113. Collie-Duguid E*, et al.* (1999) The frequency and distribution of thiopurine methyltransferase alleles in Caucasian and Asian populations. *Pharmacogenetics* 9(1):37-42.
114. McLeod HL & Siva C (2002) The thiopurine S-methyltransferase gene locus–implications for clinical pharmacogenomics. *Pharmacogenomics* 3(1):89-98.
115. Lennard L, Gibson B, Nicole T, & Lilleyman J (1993) Congenital thiopurine methyltransferase deficiency and 6-mercaptopurine toxicity during treatment for acute lymphoblastic leukaemia. *Archives of disease in childhood* 69(5):577-579.
116. Lee W, Lockhart AC, Kim RB, & Rothenberg ML (2005) Cancer pharmacogenomics: powerful tools in cancer chemotherapy and drug development. *The Oncologist* 10(2):104-111.
117. Guillemette C (2003) Pharmacogenomics of human UDP-glucuronosyltransferase enzymes. *The pharmacogenomics journal* 3(3):136-158.
118. Chabot GG (1997) Clinical pharmacokinetics of irinotecan. *Clinical pharmacokinetics* 33(4):245-259.
119. Rothenberg ML*, et al.* (1993) Phase I and pharmacokinetic trial of weekly CPT-11. *Journal of Clinical Oncology* 11(11):2194-2204.
120. Pommier Y, Pourquier P, Fan Y, & Strumberg D (1998) Mechanism of action of eukaryotic DNA topoisomerase I and drugs targeted to the enzyme. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression* 1400(1-3):83-106.
121. Di Paolo A*, et al.* (2011) Pharmacokinetic and pharmacogenetic predictive markers of irinotecan activity and toxicity. *Current drug metabolism* 12(10):932-943.
122. Slatter JG*, et al.* (2000) Pharmacokinetics, metabolism, and excretion of irinotecan (CPT-11) following iv infusion of [14C] CPT-11 in cancer patients. *Drug Metabolism and Disposition* 28(4):423-433.

123. Mick R, Gupta E, Vokes EE, & Ratain MJ (1996) Limited-sampling models for irinotecan pharmacokinetics-pharmacodynamics: prediction of biliary index and intestinal toxicity. *Journal of clinical oncology* 14(7):2012-2019.
124. Bosma PJ*, et al.* (1995) The genetic basis of the reduced expression of bilirubin UDP-glucuronosyltransferase 1 in Gilbert's syndrome. *New England Journal of Medicine* 333(18):1171-1175.
125. Innocenti F*, et al.* (2004) Genetic variants in the UDP-glucuronosyltransferase 1A1 gene predict the risk of severe neutropenia of irinotecan. *Journal of Clinical Oncology* 22(8):1382-1388.
126. Beutler E, Gelbart T, & Demina A (1998) Racial variability in the UDP-glucuronosyltransferase 1 (UGT1A1) promoter: a balanced polymorphism for regulation of bilirubin metabolism? *Proceedings of the National Academy of Sciences* 95(14):8170-8174.
127. Innocenti F*, et al.* (2002) Haplotype structure of the UDP-glucuronosyltransferase 1A1 promoter in different ethnic groups. *Pharmacogenetics and Genomics* 12(9):725-733.
128. McLeod HL & Watters JW (2004) Irinotecan pharmacogenetics: is it time to intervene? (American Society of Clinical Oncology), pp 1356-1359.
129. Dobritzsch D, Schneider G, Schnackerz KD, & Lindqvist Y (2001) Crystal structure of dihydropyrimidine dehydrogenase, a major determinant of the pharmacokinetics of the anti-cancer drug 5-fluorouracil. *The EMBO journal* 20(4):650-660.
130. Dobritzsch D, Ricagno S, Schneider G, Schnackerz KD, & Lindqvist Y (2002) Crystal Structure of the Productive Ternary Complex of Dihydropyrimidine Dehydrogenase with NADPH and 5-Iodouracil: IMPLICATIONS FOR MECHANISM OF INHIBITION AND ELECTRON TRANSFER∗. *Journal of Biological Chemistry* 277(15):13155-13166.
131. Lohkamp B, Voevodskaya N, Lindqvist Y, & Dobritzsch D (2010) Insights into the mechanism of dihydropyrimidine dehydrogenase from site-directed mutagenesis targeting the active site loop and redox cofactor coordination. *Biochimica Et Biophysica Acta (BBA)-Proteins and Proteomics* 1804(12):2198-2206.
132. Miteva-Marcheva NN, Ivanov HY, Dimitrov DK, & Stoyanova VK (2020) Application of pharmacogenetics in oncology. *Biomarker Research* 8(1):1-10.
133. van Kuilenburg AB*, et al.* (2017) Severe fluoropyrimidine toxicity due to novel and rare DPYD missense mutations, deletion and genomic amplification affecting DPD activity and mRNA splicing. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* 1863(3):721-730.
134. Meyerhardt JA & Mayer RJ (2005) Systemic therapy for colorectal cancer. *New England journal of medicine* 352(5):476-487.
135. Twelves C*, et al.* (2005) Capecitabine as adjuvant treatment for stage III colon cancer. *New England Journal of Medicine* 352(26):2696-2704.
136. Heidelberger C*, et al.* (1957) Fluorinated pyrimidines, a new class of tumour-inhibitory compounds. *Nature* 179(4561):663-666.
137. Christensen S*, et al.* (2019) 5-Fluorouracil treatment induces characteristic T> G mutations in human cancer. *Nature communications* 10(1):1-11.
138. Sethy C & Kundu CN (2021) 5-Fluorouracil (5-FU) resistance and the new strategy to enhance the sensitivity against cancer: Implication of DNA repair inhibition. *Biomedicine & Pharmacotherapy* 137:111285.
139. Mittal B, Tulsyan S, Kumar S, Mittal RD, & Agarwal G (2015) Cytochrome P450 in cancer susceptibility and treatment. *Advances in clinical chemistry* 71:77-139.
140. Dean L & Kane M (2020) Capecitabine therapy and DPYD genotype.
141. Maring JG*, et al.* (2002) Reduced 5-FU clearance in a patient with low DPD activity due to heterozygosity for a mutant allele of the DPYD gene. *British journal of cancer* 86(7):1028-1033.

142. Longley DB, Harkin DP, & Johnston PG (2003) 5-fluorouracil: mechanisms of action and clinical strategies. *Nature reviews cancer* 3(5):330-338.
143. Lu Z, Zhang R, & Diasio RB (1993) Dihydropyrimidine dehydrogenase activity in human peripheral blood mononuclear cells and liver: population characteristics, newly identified deficient patients, and clinical implication in 5-fluorouracil chemotherapy. *Cancer research* 53(22):5433-5438.
144. Vodenkova S*, et al.* (2020) 5-fluorouracil and other fluoropyrimidines in colorectal cancer: Past, present and future. *Pharmacology & therapeutics* 206:107447.
145. Malet-Martino M & Martino R (2002) Clinical studies of three oral prodrugs of 5-fluorouracil (capecitabine, UFT, S-1): a review. *The oncologist* 7(4):288-323.
146. Van Kuilenburg AB, De Abreu RA, & Van Gennip AH (2003) Pharmacogenetic and clinical aspects of dihydropyrimidine dehydrogenase deficiency. *Annals of clinical biochemistry* 40(1):41-45.
147. Tendwa MB, Chebon-Bore L, Lobb K, Musyoka TM, & Tastan Bishop Ö (2021) Force Field Parameters for Fe2+ 4S2− 4 Clusters of Dihydropyrimidine Dehydrogenase, the 5-Fluorouracil Cancer Drug Deactivation Protein: A Step towards In Silico Pharmacogenomics Studies. *Molecules* 26(10):2929.
148. Wigle TJ, Tsvetkova EV, Welch SA, & Kim RB (2019) DPYD and fluorouracil-based chemotherapy: mini review and case report. *Pharmaceutics* 11(5):199.
149. García-González X*, et al.* (2020) New DPYD variants causing DPD deficiency in patients treated with fluoropyrimidine. *Cancer Chemotherapy and Pharmacology* 86(1):45-54.
150. Lu Z-H, Zhang R, & Diasio R (1992) Purification and characterization of dihydropyrimidine dehydrogenase from human liver. *Journal of biological chemistry* 267(24):17102-17109.
151. Berman HM*, et al.* (2000) The protein data bank. *Nucleic acids research* 28(1):235-242.
152. Mander L & Liu H-w (2010) *Comprehensive natural products II: chemistry and biology* (Elsevier).
153. Heine T, Van Berkel WJ, Gassner G, Van Pée K-H, & Tischler D (2018) Two-component FAD-dependent monooxygenases: current knowledge and biotechnological opportunities. *Biology* 7(3):42.
154. Pitsawong W, Haynes CA, Koder Jr RL, Rodgers DW, & Miller A-F (2017) Mechanism-informed refinement reveals altered substrate-binding mode for catalytically competent nitroreductase. *Structure* 25(7):978-987. e974.
155. Schnackerz KD, Dobritzsch D, Lindqvist Y, & Cook PF (2004) Dihydropyrimidine dehydrogenase: a flavoprotein with four iron–sulfur clusters. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* 1701(1-2):61-74.
156. Podschun B, Cook P, & Schnackerz K (1990) Kinetic mechanism of dihydropyrimidine dehydrogenase from pig liver. *Journal of Biological Chemistry* 265(22):12966-12972.
157. Dudev T & Lim C (2008) Metal binding affinity and selectivity in metalloproteins: insights from computational studies. *Annu. Rev. Biophys.* 37:97-116.
158. Palmirotta R*, et al.* (2019) Rare Dihydropyrimidine Dehydrogenase Variants and Toxicity by Floropyrimidines: A Case Report. *Frontiers in Oncology* 9:139.
159. Palmirotta R*, et al.* (2017) Characterization of a rare nonpathogenic sequence variant (c. 1905C> T) of the dihydropyrimidine dehydrogenase gene (DPYD). *The International Journal of Biological Markers* 32(3):357-360.
160. Palmirotta R*, et al.* (2018) SNPs in predicting clinical efficacy and toxicity of chemotherapy: walking through the quicksand. *Oncotarget* 9(38):25355.
161. Ruzzo A*, et al.* (2017) Dihydropyrimidine dehydrogenase pharmacogenetics for predicting fluoropyrimidine-related toxicity in the randomised, phase III adjuvant TOSCA trial in high-risk colon cancer patients. *British journal of cancer* 117(9):1269-1277.

162. Henricks LM*, et al.* (2018) DPYD genotype-guided dose individualisation of fluoropyrimidine therapy in patients with cancer: a prospective safety analysis. *The Lancet Oncology* 19(11):1459-1467.
163. Henricks LM*, et al.* (2015) Translating DPYD genotype into DPD phenotype: using the DPYD gene activity score. *Pharmacogenomics* 16(11):1275-1284.
164. Amstutz U*, et al.* (2018) Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for dihydropyrimidine dehydrogenase genotype and fluoropyrimidine dosing: 2017 update. *Clinical Pharmacology & Therapeutics* 103(2):210-216.
165. Lunenburg CA*, et al.* (2020) Dutch Pharmacogenetics Working Group (DPWG) guideline for the gene–drug interaction of DPYD and fluoropyrimidines. *European Journal of Human Genetics* 28(4):508-517.
166. Meulendijks D*, et al.* (2015) Clinical relevance of DPYD variants c. 1679T> G, c. 1236G> A/HapB3, and c. 1601G> A as predictors of severe fluoropyrimidine-associated toxicity: a systematic review and meta-analysis of individual patient data. *The Lancet Oncology* 16(16):1639-1650.
167. Deenen MJ*, et al.* (2011) Relationship between Single Nucleotide Polymorphisms and Haplotypes in DPYD and Toxicity and Efficacy of Capecitabine in Advanced Colorectal CancerSNPs and Haplotypes in DPYD and Outcome of Capecitabine. *Clinical Cancer Research* 17(10):3455-3468.
168. Offer SM*, et al.* (2014) Comparative Functional Analysis of DPYD Variants of Potential Clinical Relevance to Dihydropyrimidine Dehydrogenase ActivityScreening of DPYD Variants for Altered 5-FU Catabolism. *Cancer research* 74(9):2545-2554.
169. Saif M, Ezzeldin H, Vance K, Sellers S, & Diasio RB (2007) DPYD* 2A mutation: the most common mutation associated with DPD deficiency. *Cancer chemotherapy and pharmacology* 60:503-507.
170. Salgueiro N*, et al.* (2004) Mutations in exon 14 of dihydropyrimidine dehydrogenase and 5-Fluorouracil toxicity in Portuguese colorectal cancer patients. *Genetics in Medicine* 6(2):102-107.
171. van KUILENBURG AB*, et al.* (2002) Novel disease-causing mutations in the dihydropyrimidine dehydrogenase gene interpreted by analysis of the three-dimensional protein structure. *Biochemical Journal* 364(1):157-163.
172. Schwab M*, et al.* (2008) Role of genetic and nongenetic factors for fluorouracil treatment-related severe toxicity: a prospective clinical trial by the German 5-FU Toxicity Study Group. *Journal of clinical oncology* 26(13):2131-2138.
173. Morel A*, et al.* (2006) Clinical relevance of different dihydropyrimidine dehydrogenase gene single nucleotide polymorphisms on 5-fluorouracil tolerance. *Molecular cancer therapeutics* 5(11):2895-2904.
174. Offer SM*, et al.* (2014) Comparative functional analysis of DPYD variants of potential clinical relevance to dihydropyrimidine dehydrogenase activity. *Cancer research* 74(9):2545-2554.
175. Milano G (2016) Highlight on DPYD gene polymorphisms and treatment by capecitabine. *Scandinavian Journal of Clinical and Laboratory Investigation* 76(sup245):S30-S33.
176. Offer SM*, et al.* (2013) A DPYD variant (Y186C) in individuals of african ancestry is associated with reduced DPD enzyme activity. *Clinical Pharmacology & Therapeutics* 94(1):158-166.
177. Saif MW*, et al.* (2014) A DPYD variant (Y186C) specific to individuals of African descent in a patient with life-threatening 5-FU toxic effects: potential for an individualized medicine approach. *Mayo Clinic Proceedings*, (Elsevier), pp 131-136.
178. Etienne-Grimaldi M-C*, et al.* (2017) New advances in DPYD genotype and risk of severe toxicity under capecitabine. *PLoS One* 12(5):e0175998.
179. van Kuilenburg AB*, et al.* (2016) Phenotypic and clinical implications of variants in the dihydropyrimidine dehydrogenase gene. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* 1862(4):754-762.

180. Gross E*, et al.* (2003) Detailed analysis of five mutations in dihydropyrimidine dehydrogenase detected in cancer patients with 5-fluorouracil-related side effects. *Human mutation* 22(6):498-498.
181. Fredj RB*, et al.* (2007) Mutational spectrum of dihydropyrimidine dehydrogenase gene (DPYD) in the Tunisian population. *Comptes rendus biologies* 330(10):764-769.
182. Da Rocha JE, Lombard Z, & Ramsay M (2021) Potential Impact of DPYD Variation on Fluoropyrimidine Drug Response in sub-Saharan African Populations. *Frontiers in genetics* 12:626954.
183. Morsman JM*, et al.* (2000) Evaluation of dihydropyrimidine dehydrogenase activity in South-west Asian, Kenyan and Ghanaian populations. *British journal of clinical pharmacology* 50(3):269-272.
184. Etienne M*, et al.* (1994) Population study of dihydropyrimidine dehydrogenase in cancer patients. *Journal of Clinical Oncology* 12(11):2248-2253.
185. Sohn D-R, Cho MS, & Chung P-J (1999) Dihydropyrimidine dehydrogenase activity in a Korean population. *Therapeutic drug monitoring* 21(2):152-154.
186. White C*, et al.* (2021) Ethnic Diversity of DPD Activity and the DPYD Gene: Review of the Literature. *Pharmacogenomics and Personalized Medicine* 14:1603.
187. Gudmundsson S*, et al.* (2022) Variant interpretation using population databases: Lessons from gnomAD. *Human mutation* 43(8):1012-1030.
188. Karczewski KJ*, et al.* (2017) The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic acids research* 45(D1):D840-D845.
189. Meier A & Söding J (2015) Automatic prediction of protein 3D structures by probabilistic multi-template homology modeling. *PLoS computational biology* 11(10):e1004343.
190. Tastan Bishop AO, De Beer TA, & Joubert F (2008) Protein homology modelling and its use in South Africa: research in action. *South African Journal of Science* 104(1):2-6.
191. Söding J, Biegert A, & Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic acids research* 33(suppl_2):W244-W248.
192. Hatherley R, Brown DK, Glenister M, & Tastan Bishop Ö (2016) PRIMO: an interactive homology modeling pipeline. *PLoS One* 11(11):e0166698.
193. Bourne PE*, et al.* (2004) The distribution and query systems of the RCSB Protein Data Bank. *Nucleic acids research* 32(suppl_1):D223-D225.
194. Steinegger M*, et al.* (2019) HH-suite3 for fast remote homology detection and deep protein annotation. *BMC bioinformatics* 20(1):1-15.
195. Rost B & Sander C (2000) Third generation prediction of secondary structures. *Protein structure prediction*, (Springer), pp 71-95.
196. Muhammed MT & Aki-Yalcin E (2019) Homology modeling in drug discovery: Overview, current applications, and future perspectives. *Chemical biology & drug design* 93(1):12-20.
197. Levasseur A, Pontarotti P, Poch O, & Thompson JD (2008) Strategies for reliable exploitation of evolutionary concepts in high throughput biology. *Evolutionary Bioinformatics* 4:EBO. S597.
198. Thompson JD, Linard B, Lecompte O, & Poch O (2011) A comprehensive benchmark study of multiple sequence alignment methods: current challenges and future perspectives. *PloS one* 6(3):e18093.
199. Thompson JD, Higgins DG, & Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research* 22(22):4673-4680.
200. Notredame C, Holm L, & Higgins DG (1998) COFFEE: an objective function for multiple sequence alignments. *Bioinformatics (Oxford, England)* 14(5):407-422.

201. Notredame C, Higgins DG, & Heringa J (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of molecular biology* 302(1):205-217.
202. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* 32(5):1792-1797.
203. Katoh K, Misawa K, Kuma Ki, & Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research* 30(14):3059-3066.
204. Pei J, Kim B-H, & Grishin NV (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic acids research* 36(7):2295-2300.
205. Waterhouse A*, et al.* (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research* 46(W1):W296-W303.
206. Webb B & Sali A (2016) Comparative protein structure modeling using MODELLER. *Current protocols in bioinformatics* 54(1):5.6. 1-5.6. 37.
207. Zheng W*, et al.* (2021) Folding non-homologous proteins by coupling deep-learning contact maps with I-TASSER assembly simulations. *Cell reports methods* 1(3):100014.
208. Eswar N*, et al.* (2006) Comparative protein structure modeling using Modeller. *Current protocols in bioinformatics* 15(1):5.6. 1-5.6. 30.
209. Laskowski RA, MacArthur MW, Moss DS, & Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography* 26(2):283-291.
210. Wiederstein M & Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research* 35(suppl_2):W407-W410.
211. Colovos C & Yeates T (1993) ERRAT: an empirical atom-based method for validating protein structures. *Protein Sci* 2(9):1511-1519.
212. Gopalakrishnan K, Sowmiya G, Sheik S, & Sekar K (2007) Ramachandran plot on the web (2.0). *Protein and peptide letters* 14(7):669-671.
213. Messaoudi A, Belguith H, & Ben Hamida J (2013) Homology modeling and virtual screening approaches to identify potent inhibitors of VEB-1 β-lactamase. *Theoretical Biology and Medical Modelling* 10(1):1-10.
214. Jumper J*, et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596(7873):583-589.
215. Edgar RC & Batzoglou S (2006) Multiple sequence alignment. *Current opinion in structural biology* 16(3):368-373.
216. Cheng J, Randall A, & Baldi P (2006) Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins: Structure, Function, and Bioinformatics* 62(4):1125-1132.
217. Capriotti E, Fariselli P, & Casadio R (2005) I-Mutant2. 0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic acids research* 33(suppl_2):W306-W310.
218. Capriotti E, Calabrese R, & Casadio R (2006) Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics* 22(22):2729-2734.
219. Brown DK & Tastan Bishop Ö (2018) HUMA: A platform for the analysis of genetic variation in humans. *Human mutation* 39(1):40-51.
220. Ponzoni L, Peñaherrera DA, Oltvai ZN, & Bahar I (2020) Rhapsody: predicting the pathogenicity of human missense variants. *Bioinformatics* 36(10):3084-3092.
221. Bahar I, Lezon TR, Yang L-W, & Eyal E (2010) Global dynamics of proteins: bridging between structure and function. *Annual review of biophysics* 39:23-42.

222. Putignano V, Rosato A, Banci L, & Andreini C (2018) MetalPDB in 2018: a database of metal sites in biological macromolecular structures. *Nucleic acids research* 46(D1):D459-D464.
223. Khairallah A, Tastan Bishop Ö, & Moses V (2021) AMBER force field parameters for the Zn (II) ions of the tunneling-fold enzymes GTP cyclohydrolase I and 6-pyruvoyl tetrahydropterin synthase. *Journal of Biomolecular Structure and Dynamics* 39(16):5843-5860.
224. Musyoka T, Bishop ÖT, Lobb K, & Moses V (2018) The determination of CHARMM force field parameters for the Mg2+ containing HIV-1 integrase. *Chemical Physics Letters* 711:1-7.
225. Moses V (2017) The investigation of type-specific features of the copper coordinating AA9 proteins and their effect on the interaction with crystalline cellulose using molecular dynamics studies. (RHODES UNIVERSITY, SOUTH AFRICA).
226. Lopes PE, Guvench O, & MacKerell AD (2015) Current status of protein force fields for molecular dynamics simulations. *Molecular modeling of proteins*, (Springer), pp 47-71.
227. Robustelli P, Piana S, & Shaw DE (2018) Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences* 115(21):E4758-E4766.
228. Apostolopoulos V*, et al.* (2021) A global review on short peptides: Frontiers and perspectives. *Molecules* 26(2):430.
229. Shi W & Chance MR (2011) Metalloproteomics: forward and reverse approaches in metalloprotein structural and functional characterization. *Current opinion in chemical biology* 15(1):144-148.
230. Bray S & Furriols M (2001) Notch pathway: making sense of suppressor of hairless. *Current Biology* 11(6):R217-R221.
231. Carloni P, Rothlisberger U, & Parrinello M (2002) The role and perspective of ab initio molecular dynamics in the study of biological systems. *Accounts of Chemical Research* 35(6):455-464.
232. Wu Q*, et al.* (2017) Poly [n] catenanes: Synthesis of molecular interlocked chains. *Science* 358(6369):1434-1439.
233. Oostenbrink C, Villa A, Mark AE, & Van Gunsteren WF (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *Journal of computational chemistry* 25(13):1656-1676.
234. Cornell WD*, et al.* (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society* 117(19):5179-5197.
235. Best RB*, et al.* (2012) 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. *Journal of chemical theory and computation* 8(9):3257-3273.
236. MacKerell Jr AD*, et al.* (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The journal of physical chemistry B* 102(18):3586-3616.
237. Guvench O & MacKerell AD (2008) Comparison of protein force fields for molecular dynamics simulations. *Molecular modeling of proteins*:63-88.
238. Kaminski GA, Friesner RA, Tirado-Rives J, & Jorgensen WL (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *The Journal of Physical Chemistry B* 105(28):6474-6487.
239. Price ML, Ostrovsky D, & Jorgensen WL (2001) Gas-phase and liquid-state properties of esters, nitriles, and nitro compounds with the OPLS-AA force field. *Journal of Computational Chemistry* 22(13):1340-1352.
240. Mohamadi F*, et al.* (1990) Macromodel—an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *Journal of Computational Chemistry* 11(4):440-467.

241. Vanommeslaeghe K*, et al.* (2010) CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of computational chemistry* 31(4):671-690.
242. Yu W, He X, Vanommeslaeghe K, & MacKerell Jr AD (2012) Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. *Journal of computational chemistry* 33(31):2451-2468.
243. Vanommeslaeghe K, Raman EP, & MacKerell Jr AD (2012) Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges. *Journal of chemical information and modeling* 52(12):3155-3168.
244. Vanommeslaeghe K & MacKerell Jr AD (2012) Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing. *Journal of chemical information and modeling* 52(12):3144-3154.
245. Hancock RD (1990) Molecular mechanics calculations and metal ion recognition. *Accounts of Chemical Research* 23(8):253-257.
246. Stote RH & Karplus M (1995) Zinc binding in proteins and solution: a simple but accurate nonbonded representation. *Proteins: Structure, Function, and Bioinformatics* 23(1):12-31.
247. Wang J, Wolf RM, Caldwell JW, Kollman PA, & Case DA (2004) Development and testing of a general amber force field. *Journal of computational chemistry* 25(9):1157-1174.
248. Case DA*, et al.* (2008) Amber 10. (University of California).
249. Li P & Merz Jr KM (2016) MCPB. py: a python based metal center parameter builder. (ACS Publications).
250. Zheng S*, et al.* (2016) VFFDT: a new software for preparing AMBER force field parameters for metal-containing molecular systems. *Journal of chemical information and modeling* 56(4):811-818.
251. Pang Y-P (1999) Novel zinc protein molecular dynamics simulations: Steps toward antiangiogenesis for cancer treatment. *Molecular modeling annual* 5(10):196-202.
252. Oda A, Yamaotsu N, & Hirono S (2005) New AMBER force field parameters of heme iron for cytochrome P450s determined by quantum chemical calculations of simplified models. *Journal of computational chemistry* 26(8):818-826.
253. Donini OA & Kollman PA (2000) Calculation and prediction of binding free energies for the matrix metalloproteinases. *Journal of medicinal chemistry* 43(22):4180-4188.
254. Hu L & Ryde U (2011) Comparison of methods to obtain force-field parameters for metal sites. *Journal of Chemical Theory and Computation* 7(8):2452-2463.
255. Pang Y-P, Xu K, El Yazal J, & Prendergast FG (2000) Successful molecular dynamics simulation of the zinc-bound farnesyltransferase using the cationic dummy atom approach. *Protein Science* 9(10):1857-1865.
256. Liao Q, Pabis A, Strodel B, & Kamerlin SCL (2017) Extending the nonbonded cationic dummy model to account for ion-induced dipole interactions. *The journal of physical chemistry letters* 8(21):5408-5414.
257. Hoops SC, Anderson KW, & Merz Jr KM (1991) Force field design for metalloproteins. *Journal of the American Chemical Society* 113(22):8262-8270.
258. Li P & Merz Jr KM (2017) Metal ion modeling using classical mechanics. *Chemical reviews* 117(3):1564-1686.
259. Carvalho AT, Teixeira AF, & Ramos MJ (2013) Parameters for molecular dynamics simulations of iron-sulfur proteins. *Journal of Computational Chemistry* 34(18):1540-1548.
260. Seminario JM (1996) Calculation of intramolecular force fields from second-derivative tensors. *International journal of quantum chemistry* 60(7):1271-1277.

261. Li P, Song LF, & Merz Jr KM (2015) Systematic parameterization of monovalent ions employing the nonbonded model. *Journal of chemical theory and computation* 11(4):1645-1657.
262. Kazmi S*, et al.* (2022) Molecular dynamics simulation of entecavir-silver nanoparticles at different biological pH. *Pakistan Journal of Pharmaceutical Sciences* 35(4).
263. Helgaker T, Jorgensen P, & Olsen J (2014) *Molecular electronic-structure theory* (John Wiley & Sons).
264. Parr RG & Yang W (1995) Density-functional theory of the electronic structure of molecules. *Annual review of physical chemistry* 46(1):701-728.
265. Vuillon L & Lesieur C (2015) From local to global changes in proteins: a network view. *Current opinion in structural biology* 31:1-8.
266. Sheik Amamuddy O, Verkhivker GM, & Tastan Bishop Oz (2020) Impact of early pandemic stage mutations on molecular dynamics of SARS-CoV-2 Mpro. *Journal of chemical information and modeling* 60(10):5080-5102.
267. Monzon AM*, et al.* (2017) Conformational diversity analysis reveals three functional mechanisms in proteins. *PLoS computational biology* 13(2):e1005398.
268. Nussinov R & Tsai C-J (2015) Allostery without a conformational change? Revisiting the paradigm. *Current opinion in structural biology* 30:17-24.
269. Amusengeri A, Astl L, Lobb K, Verkhivker GM, & Tastan Bishop Ö (2019) Establishing computational approaches towards identifying malarial allosteric modulators: A case study of plasmodium falciparum hsp70s. *International journal of molecular sciences* 20(22):5574.
270. Sanyanga TA, Nizami B, & Tastan Bishop Ö (2019) Mechanism of action of non-synonymous single nucleotide variations associated with α-carbonic anhydrase II deficiency. *Molecules* 24(21):3987.
271. Sankar S*, et al.* (2020) Deciphering the role of filamin B calponin-homology domain in causing the Larsen syndrome, boomerang dysplasia, and atelosteogenesis type I spectrum disorders via a computational approach. *Molecules* 25(23):5543.
272. Humphrey W, Dalke A, & Schulten K (1996) J. l\/lolec. *J. Mol. Graphics* 14:33-38.
273. Ishak SNH*, et al.* (2017) Molecular dynamic simulation of space and earth-grown crystal structures of thermostable T1 lipase Geobacillus zalihae revealed a better structure. *Molecules* 22(10):1574.
274. Dong Y-w, Liao M-l, Meng X-l, & Somero GN (2018) Structural flexibility and protein adaptation to temperature: Molecular dynamics analysis of malate dehydrogenases of marine molluscs. *Proceedings of the National Academy of Sciences* 115(6):1274-1279.
275. Jiang Z, You L, Dou W, Sun T, & Xu P (2019) Effects of an electric field on the conformational transition of the protein: A molecular dynamics simulation study. *Polymers* 11(2):282.
276. Ivankov DN, Bogatyreva NS, Lobanov MY, & Galzitskaya OV (2009) Coupling between properties of the protein shape and the rate of protein folding. *PloS one* 4(8):e6476.
277. Kasahara K, Fukuda I, & Nakamura H (2014) A novel approach of dynamic cross correlation analysis on molecular dynamics simulations and its application to Ets1 dimer–DNA complex. *PloS one* 9(11):e112419.
278. Tekpinar M, Neron B, & Delarue M (2021) Extracting Dynamical Correlations and Identifying Key Residues for Allosteric Communication in Proteins by correlationplus. *Journal of Chemical Information and Modeling* 61(10):4832-4838.
279. Brown DK & Bishop ÖT (2017) Role of structural bioinformatics in drug discovery by computational SNP analysis: analyzing variation at the protein level. *Global heart* 12(2):151-161.
280. Jolliffe IT & Cadima J (2016) Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374(2065):20150202.

281. Hayward S & Groot BLd (2008) Normal modes and essential dynamics. *Molecular Modeling of Proteins*, (Springer), pp 89-106.
282. David CC, Avery CS, & Jacobs DJ (2021) JEDi: java essential dynamics inspector—a molecular trajectory analysis toolkit. *BMC bioinformatics* 22(1):1-23.
283. Amamuddy OS, Boateng RA, Barozi V, Nyamai DW, & Bishop ÖT (2021) Novel dynamic residue network analysis approaches to study allosteric modulation: SARS-CoV-2 Mpro and its evolutionary mutations as a case study. *Computational and Structural Biotechnology Journal* 19:6431-6455.
284. Amamuddy OS, Glenister M, Tshabalala T, & Bishop ÖT (2021) MDM-TASK-web: MD-TASK and MODE-TASK web server for analyzing protein dynamics. *Computational and Structural Biotechnology Journal* 19:5059-5071.
285. Barozi V, Musyoka TM, Sheik Amamuddy O, & Tastan Bishop Oz (2022) Deciphering Isoniazid Drug Resistance Mechanisms on Dimeric Mycobacterium tuberculosis KatG via Post-molecular Dynamics Analyses Including Combined Dynamic Residue Network Metrics. *ACS omega*.
286. Barozi V, Edkins AL, & Bishop ÖT (2022) Evolutionary progression of collective mutations in Omicron sub-lineages towards efficient RBD-hACE2: allosteric communications between and within viral and human proteins. *Computational and Structural Biotechnology Journal* 20:4562-4578.
287. Okeke CJ, Musyoka TM, Amamuddy OS, Barozi V, & Bishop ÖT (2021) 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.
288. Golbeck J (2013) Network structure and measures. *Analyzing the social web*:25-44.
289. Coccia M, Roshani S, & Mosleh M (2022) Evolution of quantum computing: Theoretical and innovation management implications for emerging quantum industry. *IEEE Transactions on Engineering Management*.
290. Duarte JM, Sathyapriya R, Stehr H, Filippis I, & Lappe M (2010) Optimal contact definition for reconstruction of contact maps. *BMC bioinformatics* 11(1):1-10.
291. Jones DT, Buchan DW, Cozzetto D, & Pontil M (2012) PSICOV: precise structural contact prediction using sparse inverse covariance estimation on large multiple sequence alignments. *Bioinformatics* 28(2):184-190.
292. Sung H*, et al.* (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* 71(3):209-249.
293. Hamilton JG & Waters EA (2018) How are multifactorial beliefs about the role of genetics and behavior in cancer causation associated with cancer risk cognitions and emotions in the US population? *Psycho-Oncology* 27(2):640-647.
294. Shrestha S*, et al.* (2018) Gene-specific variant classifier (DPYD-varifier) to identify deleterious alleles of dihydropyrimidine dehydrogenase. *Clinical Pharmacology & Therapeutics* 104(4):709-718.
295. Mazzuca F*, et al.* (2016) Pre-treatment evaluation of 5-fluorouracil degradation rate: association of poor and ultra-rapid metabolism with severe toxicity in a colorectal cancer patients cohort. *Oncotarget* 7(15):20612.
296. Saif MW (2013) Dihydropyrimidine dehydrogenase gene (DPYD) polymorphism among Caucasian and non-Caucasian patients with 5-FU-and capecitabine-related toxicity using full sequencing of DPYD. *Cancer genomics & proteomics* 10(2):89-92.
297. Tozer T, Heale K, Manto Chagas C, de Barros ALB, & Alisaraie L (2019) Interdomain twists of human thymidine phosphorylase and its active–inactive conformations: Binding of 5-FU and its

analogues to human thymidine phosphorylase versus dihydropyrimidine dehydrogenase. *Chemical Biology & Drug Design* 94(5):1956-1972.
298. Dal Peraro M*, et al.* (2007) Modeling the charge distribution at metal sites in proteins for molecular dynamics simulations. *Journal of structural biology* 157(3):444-453.
299. Vanommeslaeghe K & Guvench O (2014) Molecular mechanics. *Current pharmaceutical design* 20(20):3281-3292.
300. Teixeira MH*, et al.* (2019) Modeling the Hydrolysis of Iron–Sulfur Clusters. *Journal of Chemical Information and Modeling* 60(2):653-660.
301. Amamuddy OS, Musyoka TM, Boateng RA, Zabo S, & Bishop ÖT (2020) Determining the unbinding events and conserved motions associated with the pyrazinamide release due to resistance mutations of Mycobacterium tuberculosis pyrazinamidase. *Computational and structural biotechnology journal* 18:1103-1120.
302. Consortium U (2007) The universal protein resource (UniProt). *Nucleic acids research* 36(suppl_1):D190-D195.
303. Kelm S, Shi J, & Deane CM (2010) MEDELLER: homology-based coordinate generation for membrane proteins. *Bioinformatics* 26(22):2833-2840.
304. BIOVIA DS (2017) Discovery Studio Modeling Environment. (Dassault Systèmes San Diego, CA, USA).
305. Lüthy R, Bowie JU, & Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. *Nature* 356(6364):83-85.
306. Laskowski RA, MacArthur MW, & Thornton JM (1998) Validation of protein models derived from experiment. *Current opinion in structural biology* 8(5):631-639.
307. Benkert P, Künzli M, & Schwede T (2009) QMEAN server for protein model quality estimation. *Nucleic acids research* 37(suppl_2):W510-W514.
308. Anandakrishnan R, Aguilar B, & Onufriev AV (2012) H++ 3.0: automating p K prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic acids research* 40(W1):W537-W541.
309. Bowers KJ*, et al.* (2006) Scalable algorithms for molecular dynamics simulations on commodity clusters. *SC'06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*, (IEEE), pp 43-43.
310. Ryde U (1995) Molecular dynamics simulations of alcohol dehydrogenase with a four-or five-coordinate catalytic zinc ion. *Proteins: Structure, Function, and Bioinformatics* 21(1):40-56.
311. Ryde U (2003) Combined quantum and molecular mechanics calculations on metalloproteins. *Current opinion in chemical biology* 7(1):136-142.
312. Lee C, Yang W, & Parr RG (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical review B* 37(2):785.
313. Frisch M*, et al.* (2009) Gaussian Inc. *Wallingford Ct* 2.
314. Case DA*, et al.* (2006) AMBER 9. *University of California, San Francisco* 45.
315. Bayly CI, Cieplak P, Cornell W, & Kollman PA (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *The Journal of Physical Chemistry* 97(40):10269-10280.
316. Yang Y, Weaver MN, & Merz Jr KM (2009) Assessment of the "6-31+ G**+ LANL2DZ" mixed basis set coupled with density functional theory methods and the effective core potential: prediction of heats of formation and ionization potentials for first-row-transition-metal complexes. *The Journal of Physical Chemistry A* 113(36):9843-9851.
317. Becke AD (1988) Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical review A* 38(6):3098.

318. Bannwarth C, Ehlert S, & Grimme S (2019) GFN2-xTB—An accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *Journal of chemical theory and computation* 15(3):1652-1671.
319. Grimme S, Bannwarth C, & Shushkov P (2017) A robust and accurate tight-binding quantum chemical method for structures, vibrational frequencies, and noncovalent interactions of large molecular systems parametrized for all spd-block elements (Z= 1–86). *Journal of chemical theory and computation* 13(5):1989-2009.
320. Van Der Spoel D*, et al.* (2005) GROMACS: fast, flexible, and free. *Journal of computational chemistry* 26(16):1701-1718.
321. Sousa da Silva AW & Vranken WF (2012) ACPYPE-Antechamber python parser interface. *BMC research notes* 5(1):1-8.
322. Glättli A, Daura X, & van Gunsteren WF (2002) Derivation of an improved simple point charge model for liquid water: SPC/A and SPC/L. *The Journal of chemical physics* 116(22):9811-9828.
323. Hockney RW, Goel S, & Eastwood J (1974) Quiet high-resolution computer models of a plasma. *Journal of Computational Physics* 14(2):148-158.
324. Petersen HG (1995) Accuracy and efficiency of the particle mesh Ewald method. *The Journal of chemical physics* 103(9):3668-3679.
325. Essmann U*, et al.* (1995) A smooth particle mesh Ewald method. *The Journal of chemical physics* 103(19):8577-8593.
326. Ryckaert J-P, Ciccotti G, & Berendsen HJ (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of computational physics* 23(3):327-341.
327. Team R (2015) RStudio: integrated development for R. RStudio. *Inc., Boston, MA* 700:879.
328. Humphrey W, Dalke A, & Schulten K (1996) VMD: visual molecular dynamics. *Journal of molecular graphics* 14(1):33-38.
329. Brown DK*, et al.* (2017) MD-TASK: a software suite for analyzing molecular dynamics trajectories. *Bioinformatics* 33(17):2768-2771.
330. Schrödinger L (2015) The PyMOL molecular graphics system, version 1.7. 6.6. *Schrödinger LLC*.
331. Kluyver T*, et al.* (2016) *Jupyter Notebooks-a publishing format for reproducible computational workflows*.
332. Hunter JD (2007) Matplotlib: A 2D graphics environment. *Computing in science & engineering* 9(03):90-95.
333. McKinney W (2010) Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*, (Austin, TX), pp 51-56.
334. Van Der Walt S, Colbert SC, & Varoquaux G (2011) The NumPy array: a structure for efficient numerical computation. *Computing in science & engineering* 13(2):22-30.
335. Nguyen H, Case DA, & Rose AS (2018) NGLview–interactive molecular graphics for Jupyter notebooks. *Bioinformatics* 34(7):1241-1242.
336. Chen H & Kihara D (2008) Estimating quality of template-based protein models by alignment stability. *Proteins: Structure, Function, and Bioinformatics* 71(3):1255-1274.
337. Eisenberg D, Lüthy R, & Bowie JU (1997) [20] VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods in enzymology*, (Elsevier), Vol 277, pp 396-404.
338. Grimme S (2006) Semiempirical hybrid density functional with perturbative second-order correlation. *The Journal of chemical physics* 124(3):034108.
339. Li X, Hayik SA, & Merz Jr KM (2010) QM/MM X-ray refinement of zinc metalloenzymes. *Journal of inorganic biochemistry* 104(5):512-522.

340. Harding MM (1999) The geometry of metal–ligand interactions relevant to proteins. *Acta Crystallographica Section D: Biological Crystallography* 55(8):1432-1443.
341. Harding MM (2006) Small revisions to predicted distances around metal sites in proteins. *Acta Crystallographica Section D: Biological Crystallography* 62(6):678-682.
342. Wei C, Lazim R, & Zhang D (2014) Importance of polarization effect in the study of metalloproteins: Application of polarized protein specific charge scheme in predicting the reduction potential of azurin. *Proteins: Structure, Function, and Bioinformatics* 82(9):2209-2219.
343. Tuccinardi T*, et al.* (2006) Amber force field implementation, molecular modelling study, synthesis and MMP-1/MMP-2 inhibition profile of (R)-and (S)-N-hydroxy-2-(N-isopropoxybiphenyl-4-ylsulfonamido)-3-methylbutanamides. *Bioorganic & medicinal chemistry* 14(12):4260-4276.
344. Smith DM, Xiong Y, Straatsma T, Rosso KM, & Squier TC (2012) Force-field development and molecular dynamics of [NiFe] hydrogenase. *Journal of Chemical Theory and Computation* 8(6):2103-2114.
345. Chebon-Bore L, Sanyanga TA, Manyumwa CV, Khairallah A, & Tastan Bishop Ö (2021) Decoding the molecular effects of atovaquone linked resistant mutations on Plasmodium falciparum Cytb-ISP complex in the phospholipid bilayer membrane. *International journal of molecular sciences* 22(4):2138.
346. Cieplak P, Cornell WD, Bayly C, & Kollman PA (1995) Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *Journal of Computational Chemistry* 16(11):1357-1377.
347. Haspel N, Moll M, Baker ML, Chiu W, & Kavraki LE (2010) Tracing conformational changes in proteins. *BMC structural biology* 10(1):1-11.
348. Orellana L (2019) Large-scale conformational changes and protein function: breaking the in silico barrier. *Frontiers in molecular biosciences* 6:117.
349. Sara JD*, et al.* (2018) 5-fluorouracil and cardiotoxicity: a review. *Therapeutic advances in medical oncology* 10:1758835918780140.
350. Lonsdale R, Harvey JN, & Mulholland AJ (2012) Effects of dispersion in density functional based quantum mechanical/molecular mechanical calculations on cytochrome P450 catalyzed reactions. *Journal of Chemical Theory and Computation* 8(11):4637-4645.
351. Smith DG, Burns LA, Patkowski K, & Sherrill CD (2016) Revised damping parameters for the D3 dispersion correction to density functional theory. *The journal of physical chemistry letters* 7(12):2197-2203.
352. Grimme S, Antony J, Ehrlich S, & Krieg H (2010) A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of chemical physics* 132(15):154104.
353. Pinto N & Eileen Dolan M (2011) Clinically relevant genetic variations in drug metabolizing enzymes. *Current drug metabolism* 12(5):487-497.
354. Nzabonimpa GS, Rasmussen HB, Brunak S, Taboureau O, & Consortium I (2016) Investigating the impact of missense mutations in hCES1 by in silico structure-based approaches. *Drug Metabolism and Personalized Therapy* 31(2):97-106.
355. Joerger AC, Ang HC, & Fersht AR (2006) Structural basis for understanding oncogenic p53 mutations and designing rescue drugs. *Proceedings of the National Academy of Sciences* 103(41):15056-15061.
356. Bohl CE, Wu Z, Miller DD, Bell CE, & Dalton JT (2007) Crystal structure of the T877A human androgen receptor ligand-binding domain complexed to cyproterone acetate provides insight for ligand-induced conformational changes and structure-based drug design. *Journal of Biological Chemistry* 282(18):13648-13655.

357. Tomalik-Scharte D, Lazar A, Fuhr U, & Kirchheiner J (2008) The clinical role of genetic polymorphisms in drug-metabolizing enzymes. *The pharmacogenomics journal* 8(1):4-15.
358. Gonzalez FJ & Fernandez-Salguaro P (1995) Diagnostic analysis, clinical importance and molecular basis of dihydropyrimidine dehydrogenase deficiency. *Trends in pharmacological sciences* 16(10):325-327.
359. Wei X, McLeod HL, McMurrough J, Gonzalez FJ, & Fernandez-Salguero P (1996) Molecular basis of the human dihydropyrimidine dehydrogenase deficiency and 5-fluorouracil toxicity. *The Journal of clinical investigation* 98(3):610-615.
360. Miller A, Hoogstraten B, Staquet M, & Winkler A (1981) Reporting results of cancer treatment. *cancer* 47(1):207-214.
361. Meulendijks D, Cats A, Beijnen JH, & Schellens JH (2016) Improving safety of fluoropyrimidine chemotherapy by individualizing treatment based on dihydropyrimidine dehydrogenase activity–ready for clinical practice? *Cancer treatment reviews* 50:23-34.
362. Tuchman M*, et al.* (1985) Familial pyrimidinemia and pyrimidinuria associated with severe fluorouracil toxicity. *New England Journal of Medicine* 313(4):245-249.
363. Diasio RB, Beavers T, & Carpenter J (1988) Familial deficiency of dihydropyrimidine dehydrogenase. Biochemical basis for familial pyrimidinemia and severe 5-fluorouracil-induced toxicity. *The Journal of clinical investigation* 81(1):47-51.
364. Van Kuilenburg A*, et al.* (1999) Genotype and phenotype in patients with dihydropyrimidine dehydrogenase deficiency. *Human genetics* 104(1):1-9.
365. van Kuilenburg AB*, et al.* (2012) Evaluation of 5-fluorouracil pharmacokinetics in cancer patients with a c. 1905+ 1G> A mutation in DPYD by means of a Bayesian limited sampling strategy. *Clinical pharmacokinetics* 51(3):163-174.
366. Deenen MJ*, et al.* (2016) Upfront genotyping of DPYD* 2A to individualize fluoropyrimidine therapy: a safety and cost analysis. *Journal of Clinical Oncology* 34(3):227-234.
367. Vreken P*, et al.* (1996) A point mutation in an invariant splice donor site leads to exon skipping in two unrelated Dutch patients with dihydropyrimidine dehydrogenase deficiency. *Journal of inherited metabolic disease* 19(5):645-654.
368. Terrazzino S*, et al.* (2013) DPYD IVS14+ 1G> A and 2846A> T genotyping for the prediction of severe fluoropyrimidine-related toxicity: a meta-analysis. *Pharmacogenomics* 14(11):1255-1272.
369. Balboa-Beltrán E, Duran G, Lamas MJ, Carracedo A, & Barros F (2015) Long Survival and Severe Toxicity Under 5-Fluorouracil–Based Therapy in a Patient With Colorectal Cancer Who Harbors a Germline Codon-Stop Mutation in TYMS. *Mayo Clinic Proceedings*, (Elsevier), pp 1298-1303.
370. Zalcberg J, Kerr D, Seymour L, Palmer M, & Group TIS (1998) Haematological and non-haematological toxicity after 5-fluorouracil and leucovorin in patients with advanced colorectal cancer is significantly associated with gender, increasing age and cycle number. *European journal of cancer* 34(12):1871-1875.
371. Stein BN*, et al.* (1995) Age and sex are independent predictors of 5-fluorouracil toxicity. Analysis of a large scale phase III trial. *Cancer* 75(1):11-17.
372. van Kuilenburg AB*, et al.* (2010) Intragenic deletions and a deep intronic mutation affecting pre-mRNA splicing in the dihydropyrimidine dehydrogenase gene as novel mechanisms causing 5-fluorouracil toxicity. *Human genetics* 128(5):529-538.
373. van Kuilenburg AB*, et al.* (2001) Lethal outcome of a patient with a complete dihydropyrimidine dehydrogenase (DPD) deficiency after administration of 5-fluorouracil: frequency of the common IVS14+ 1G> A mutation causing DPD deficiency. *Clinical cancer research* 7(5):1149-1153.
374. Barozi V, Musyoka TM, Sheik Amamuddy O, & Tastan Bishop Oz (2022) Deciphering Isoniazid Drug Resistance Mechanisms on Dimeric Mycobacterium tuberculosis KatG via Post-molecular

Dynamics Analyses Including Combined Dynamic Residue Network Metrics. *ACS omega* 7(15):13313-13332.
375. Sotomayor-Vivas C, Hernández-Lemus E, & Dorantes-Gilardi R (2022) Linking protein structural and functional change to mutation using amino acid networks. *Plos one* 17(1):e0261829.
376. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, & Hamosh A (2015) OMIM. org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic acids research* 43(D1):D789-D798.
377. Landrum MJ*, et al.* (2018) ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic acids research* 46(D1):D1062-D1067.
378. Whirl-Carrillo M*, et al.* (2012) Pharmacogenomics knowledge for personalized medicine. *Clinical Pharmacology & Therapeutics* 92(4):414-417.
379. Wishart DS*, et al.* (2018) DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic acids research* 46(D1):D1074-D1082.
380. Server EV & Project NGES (2016) Internet. *NHLBI GO Exome Sequencing Project (ESP).[cited 28 March 2017]. http://evs. gs. washington. edu/EVS*.
381. Yates AD*, et al.* (2020) Ensembl 2020. *Nucleic acids research* 48(D1):D682-D688.
382. Consortium U (2014) Activities at the universal protein resource (UniProt). *Nucleic acids research* 42(D1):D191-D198.
383. Burley SK*, et al.* (2019) RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic acids research* 47(D1):D464-D474.
384. Feng HBJWZ & Weissig GGTBH (IN Shindyalov PE Bourne (2000). *Nucleic Acids Res* 28:235-242.
385. Webb B & Sali A (2017) Protein structure modeling with MODELLER. *Functional genomics*, (Springer), pp 39-54.
386. Colovos C & Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. *Protein science* 2(9):1511-1519.
387. Grimsley GR, Scholtz JM, & Pace CN (2009) A summary of the measured pK values of the ionizable groups in folded proteins. *Protein Science* 18(1):247-251.
388. Pace CN, Grimsley GR, & Scholtz JM (2009) Protein ionizable groups: pK values and their contribution to protein stability and solubility. *Journal of Biological Chemistry* 284(20):13285-13289.
389. Myers J, Grothaus G, Narayanan S, & Onufriev A (2006) A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules. *Proteins: Structure, Function, and Bioinformatics* 63(4):928-938.
390. Gordon JC*, et al.* (2005) H++: a server for estimating p K as and adding missing hydrogens to macromolecules. *Nucleic acids research* 33(suppl_2):W368-W371.
391. Salam NK, Adzhigirey M, Sherman W, & Pearlman DA (2014) Structure-based approach to the prediction of disulfide bonds in proteins. *Protein Engineering, Design & Selection* 27(10):365-374.
392. Farhoodi R*, et al.* (2017) Predicting the effect of point mutations on protein structural stability. *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, pp 247-252.
393. Abraham MJ*, et al.* (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1:19-25.
394. Maier JA*, et al.* (2015) ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of chemical theory and computation* 11(8):3696-3713.
395. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, & Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* 79(2):926-935.

396. Parrinello M & Rahman A (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics* 52(12):7182-7190.
397. Hess B, Bekker H, Berendsen HJ, & Fraaije JG (1997) LINCS: a linear constraint solver for molecular simulations. *Journal of computational chemistry* 18(12):1463-1472.
398. Laskowski RA & Swindells MB (2011) LigPlot+: multiple ligand–protein interaction diagrams for drug discovery. (ACS Publications).
399. Waskom M (2021) seaborn: statistical data visualization, J. Open Source Softw., 6, 3021.
400. Hubbard RE & Haider MK (2010) Hydrogen bonds in proteins: role and strength. *eLS*.
401. Pace CN*, et al.* (2014) Contribution of hydrogen bonds to protein stability. *Protein Science* 23(5):652-661.
402. Roe DR & Cheatham III TE (2013) PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *Journal of chemical theory and computation* 9(7):3084-3095.
403. Krissinel E & Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *Journal of molecular biology* 372(3):774-797.
404. Sukhwal A & Sowdhamini R (2015) PPCheck: A webserver for the quantitative analysis of protein-protein interfaces and prediction of residue hotspots. *Bioinformatics and biology insights* 9:BBI. S25928.
405. Schrodinger L (2017) The PyMol molecular graphics system, version 2.0. Schrödinger, LLC, New York, NY.
406. Kortemme T, Kim DE, & Baker D (2004) Computational alanine scanning of protein-protein interfaces. *Science's STKE* 2004(219):pl2-pl2.
407. Sherry ST*, et al.* (2001) dbSNP: the NCBI database of genetic variation. *Nucleic acids research* 29(1):308-311.
408. Consortium GP (2015) A global reference for human genetic variation. *Nature* 526(7571):68.
409. Prabantu VM, Naveenkumar N, & Srinivasan N (2021) Influence of disease-causing mutations on protein structural networks. *Frontiers in molecular biosciences* 7:620554.
410. Thorn CF, Klein TE, & Altman RB (2013) PharmGKB: the pharmacogenomics knowledge base. *Pharmacogenomics*, (Springer), pp 311-320.
411. Wang Y*, et al.* (2021) A crowdsourcing open platform for literature curation in UniProt. *PLoS Biology* 19(12):e3001464.
412. Zardecki C, Goodsell DS, Dutta S, Voigt M, & Burley SK (2022) PDB-101: Molecular Explorations through Biology and Medicine. *FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology* 36.
413. Gerlt JA (1987) Relationships between enzymatic catalysis and active site structure revealed by applications of site-directed mutagenesis. *Chemical reviews* 87(5):1079-1105.
414. Kanaya S*, et al.* (1990) Identification of the amino acid residues involved in an active site of Escherichia coli ribonuclease H by site-directed mutagenesis. *Journal of Biological Chemistry* 265(8):4615-4621.
415. Jack BR, Meyer AG, Echave J, & Wilke CO (2016) Functional sites induce long-range evolutionary constraints in enzymes. *PLoS Biology* 14(5):e1002452.
416. Nisthal A, Wang CY, Ary ML, & Mayo SL (2019) Protein stability engineering insights revealed by domain-wide comprehensive mutagenesis. *Proceedings of the National Academy of Sciences* 116(33):16367-16377.
417. Islam MM, Kobayashi K, Kidokoro SI, & Kuroda Y (2019) Hydrophobic surface residues can stabilize a protein through improved water–protein interactions. *The FEBS Journal* 286(20):4122-4134.

418. Malleshappa Gowder S, Chatterjee J, Chaudhuri T, & Paul K (2014) Prediction and analysis of surface hydrophobic residues in tertiary structure of proteins. *The Scientific World Journal* 2014.
419. Sakharov DV & Lim C (2005) Zn protein simulations including charge transfer and local polarization effects. *Journal of the American Chemical Society* 127(13):4921-4929.
420. Penkler DL & Tastan Bishop Ö (2019) Modulation of human Hsp90α conformational dynamics by allosteric ligand interaction at the C-terminal domain. *Scientific reports* 9(1):1-17.
421. Osuna S, Jimenez-Oses G, Noey EL, & Houk K (2015) Molecular dynamics explorations of active site structure in designed and evolved enzymes. *Accounts of chemical research* 48(4):1080-1089.
422. Penkler DL, Atilgan C, & Tastan Bishop Oz (2018) Allosteric modulation of human Hsp90α conformational dynamics. *Journal of Chemical Information and Modeling* 58(2):383-404.
423. Kufareva I & Abagyan R (2011) Methods of protein structure comparison. *Homology modeling*, (Springer), pp 231-257.
424. Komatsu TS*, et al.* (2020) Drug binding dynamics of the dimeric SARS-CoV-2 main protease, determined by molecular dynamics simulation. *Scientific reports* 10(1):1-11.
425. Gross E*, et al.* (2008) Strong association of a common dihydropyrimidine dehydrogenase gene polymorphism with fluoropyrimidine-related toxicity in cancer patients. *PloS one* 3(12):e4003.
426. Seck K*, et al.* (2005) Analysis of the DPYD gene implicated in 5-fluorouracil catabolism in a cohort of Caucasian individuals. *Clinical cancer research* 11(16):5886-5892.
427. Loganayagam A*, et al.* (2013) Pharmacogenetic variants in the DPYD, TYMS, CDA and MTHFR genes are clinically significant predictors of fluoropyrimidine toxicity. *British journal of cancer* 108(12):2505-2515.
428. Patil VM*, et al.* (2016) Dihydropyrimidine dehydrogenase mutation in neoadjuvant chemotherapy in head and neck cancers: Myth or reality? *South Asian Journal of Cancer* 5(04):182-185.
429. Maharjan AS*, et al.* (2019) The prevalence of DPYD* 9A (c. 85T> C) genotype and the genotype-phenotype correlation in patients with gastrointestinal malignancies treated with fluoropyrimidines: updated analysis. *Clinical colorectal cancer* 18(3):e280-e286.
430. Varma A, Jayanthi M, Dubashi B, Shewade DG, & Sundaram R (2020) Genetic influence of DPYD* 9A polymorphism on plasma levels of 5-fluorouracil and subsequent toxicity after oral administration of capecitabine in colorectal cancer patients of South Indian origin. *Drug Metabolism and Personalized Therapy* 35(4).
431. Sheik Amamuddy O, Musyoka T, Boateng R, Zabo S, & Tastan Bishop Ö (2020) Determining the unbinding events and conserved motions associated with the pyrazinamide release due to resistance mutations of Mycobacterium tuberculosis pyrazinamidase. *Comput. Struct. Biotechnol. J* 18:1103-1120.
432. Yang G, Hong N, Baier F, Jackson CJ, & Tokuriki N (2016) Conformational tinkering drives evolution of a promiscuous activity through indirect mutational effects. *Biochemistry* 55(32):4583-4593.
433. Naganathan AN (2019) Modulation of allosteric coupling by mutations: from protein dynamics and packing to altered native ensembles and function. *Current opinion in structural biology* 54:1-9.
434. Cunningham F*, et al.* (2022) Ensembl 2022. *Nucleic acids research* 50(D1):D988-D995.
435. Collie-Duguid E, Etienne M, Milano G, & McLeod H (2000) Known variant DPYD alleles do not explain DPD deficiency in cancer patients. *Pharmacogenetics and Genomics* 10(3):217-223.
436. Mattison LK, Johnson MR, & Diasio RB (2002) A comparative analysis of translated dihydropyrimidine dehydrogenase cDNA; conservation of functional domains and relevance to genetic polymorphisms. *Pharmacogenetics and Genomics* 12(2):133-144.

437. Vendruscolo M, Dokholyan NV, Paci E, & Karplus M (2002) Small-world view of the amino acids that play a key role in protein folding. *Physical Review E* 65(6):061910.
438. Brown DK, Amamuddy OS, & Bishop ÖT (2017) Structure-based analysis of single nucleotide variants in the renin-angiotensinogen complex. *Global heart* 12(2):121-132.
439. Amusengeri A, Tata RB, & Tastan Bishop Ö (2020) Understanding the pyrimethamine drug resistance mechanism via combined molecular dynamics and dynamic residue network analysis. *Molecules* 25(4):904.
440. Jayashree S, Murugavel P, Sowdhamini R, & Srinivasan N (2019) Interface residues of transient protein-protein complexes have extensive intra-protein interactions apart from inter-protein interactions. *Biology direct* 14(1):1-14.
441. FUJII H, KURAHASHI T, CONG Z, OHTSUKI A, & TANIZAWA MS (Structure-Function Relationship of Metalloproteins.
442. Naushad SM, Hussain T, Alrokayan SA, & Kutala VK (2021) Pharmacogenetic profiling of dihydropyrimidine dehydrogenase (DPYD) variants in the Indian population. *The Journal of Gene Medicine* 23(1):e3289.
443. Hariprakash JM*, et al.* (2018) Pharmacogenetic landscape of DPYD variants in south Asian populations by integration of genome-scale data. *Pharmacogenomics* 19(3):227-241.
444. Medwid S, Wigle TJ, & Kim RB (2022) Fluoropyrimidine-associated toxicity and DPYD variants c. 85T> C, c. 496A> G, and c. 1236G> A: impact of haplotype. *Cancer Chemotherapy and Pharmacology*:1-6.
445. Offer SM, Wegner NJ, Fossum C, Wang K, & Diasio RB (2013) Phenotypic profiling of DPYD variations relevant to 5-fluorouracil sensitivity using real-time cellular analysis and in vitro measurement of enzyme activity. *Cancer research* 73(6):1958-1968.
446. Sanyanga TA & Tastan Bishop Ö (2020) Structural characterization of carbonic anhydrase VIII and effects of missense single nucleotide variations to protein structure and function. *International journal of molecular sciences* 21(8):2764.
447. Negre CF*, et al.* (2018) Eigenvector centrality for characterization of protein allosteric pathways. *Proceedings of the National Academy of Sciences* 115(52):E12201-E12208.
448. Nyamai DW & Tastan Bishop Ö (2020) Identification of selective novel hits against plasmodium falciparum prolyl tRNA synthetase active site and a predicted allosteric site using in silico approaches. *International journal of molecular sciences* 21(11):3803.
449. Goodsell DS & Olson AJ (2000) Structural symmetry and protein function. *Annual review of biophysics and biomolecular structure* 29:105.
450. Hishinuma E*, et al.* (2022) Importance of Rare DPYD Genetic Polymorphisms for 5-Fluorouracil Therapy in the Japanese Population. *Frontiers in Pharmacology* 13.
451. Hishinuma E, Gutiérrez Rico E, & Hiratsuka M (2020) In Vitro Assessment of Fluoropyrimidine-Metabolizing Enzymes: Dihydropyrimidine Dehydrogenase, Dihydropyrimidinase, and β-Ureidopropionase. *Journal of Clinical Medicine* 9(8):2342.
452. Wachnowsky C, Wesley NA, Fidai I, & Cowan J (2017) Understanding the molecular basis of multiple mitochondrial dysfunctions syndrome 1 (MMDS1)—impact of a disease-causing Gly208Cys substitution on structure and activity of NFU1 in the Fe/S cluster biosynthetic pathway. *Journal of molecular biology* 429(6):790-807.
453. Li M, Petukh M, Alexov E, & Panchenko AR (2014) Predicting the impact of missense mutations on protein–protein binding affinity. *Journal of chemical theory and computation* 10(4):1770-1780.
454. Zhou W & Freed CR (2004) Tyrosine-to-cysteine modification of human α-synuclein enhances protein aggregation and cellular toxicity. *Journal of Biological Chemistry* 279(11):10128-10135.

455. Diasio RB & Offer SM (2022) Testing for Dihydropyrimidine Dehydrogenase Deficiency to Individualize 5-Fluorouracil Therapy. *Cancers* 14(13):3207.
456. Saif M*, et al.* (2013) A DPYD variant (Y186C) specific to individuals of African descent in a patient with life-threatening 5-FU toxicity potential for an individualized medicine approach. *Mayo Clin Proc*.
457. Betts MJ & Russell RB (2003) Amino acid properties and consequences of substitutions. *Bioinformatics for geneticists* 317(289):10.1002.
458. Zhang X-p*, et al.* (2012) Polymorphisms of dihydropyrimidine dehydrogenase gene and clinical outcomes of gastric cancer patients treated with fluorouracil-based adjuvant chemotherapy in Chinese population. *Chinese medical journal* 125(05):741-746.
459. Zhang H, Li Y-M, Zhang H, & Jin X (2007) DPYD* 5 gene mutation contributes to the reduced DPYD enzyme activity and chemotherapeutic toxicity of 5-FU: results from genotyping study on 75 gastric carcinoma and colon carcinoma patients. *Medical Oncology* 24:251-258.
460. Aledo JC (2019) Methionine in proteins: The Cinderella of the proteinogenic amino acids. *Protein Science* 28(10):1785-1796.
461. Volk J*, et al.* (2001) Safe administration of irinotecan, oxaliplatin and raltitrexed in a DPD-deficient patient with metastatic colon cancer. *Annals of oncology* 12(4):569-571.
462. Van Kuilenburg A*, et al.* (1997) Heterozygosity for a point mutation in an invariant splice donor site of dihydropyrimidine dehydrogenase and severe 5-fluorouracil related toxicity. *European Journal of Cancer* 33(13):2258-2264.
463. Lee AM*, et al.* (2014) DPYD variants as predictors of 5-fluorouracil toxicity in adjuvant colon cancer treatment (NCCTG N0147). *JNCI: Journal of the National Cancer Institute* 106(12).
464. Gentile G*, et al.* (2016) Genotype–phenotype correlations in 5-fluorouracil metabolism: a candidate DPYD haplotype to improve toxicity prediction. *The pharmacogenomics journal* 16(4):320-325.
465. Jeffrey GA & Saenger W (2012) *Hydrogen bonding in biological structures* (Springer Science & Business Media).
466. Glusker JP (1998) Directional aspects of intermolecular interactions. *Design of organic solids*:1-56.
467. Negureanu L & Salsbury Jr FR (2014) Non-specificity and synergy at the binding site of the carboplatin-induced DNA adduct via molecular dynamics simulations of the MutSα–DNA recognition complex. *Journal of Biomolecular Structure and Dynamics* 32(6):969-992.
468. Krishnamoorthy N & Fakhro K (2021) Identification of mutation resistance coldspots for targeting the SARS-CoV2 main protease. *IUBMB life* 73(4):670-675.
469. Shirian J, Sharabi O, & Shifman JM (2016) Cold spots in protein binding. *Trends in biochemical sciences* 41(9):739-745.
470. Hientz K, Mohr A, Bhakta-Guha D, & Efferth T (2017) The role of p53 in cancer drug resistance and targeted chemotherapy. *Oncotarget* 8(5):8921.
471. Tang C*, et al.* (2021) Hypomorph mutation-directed small-molecule protein-protein interaction inducers to restore mutant SMAD4-suppressed TGF-β signaling. *Cell chemical biology* 28(5):636-647. e635.
472. Patel AY*, et al.* (2021) Effects of ionic liquids on metalloproteins. *Molecules* 26(2):514.
473. Yang W*, et al.* (2016) Genomics of drug sensitivity in cancer (GDSC): a resource for biomarker discovery in cancer cells. *European Journal of Cancer* 1(69):S82.
474. Good BM, Ainscough BJ, McMichael JF, Su AI, & Griffith OL (2014) Organizing knowledge to enable personalization of medicine in cancer. *Genome biology* 15(8):1-9.
475. Costello JC*, et al.* (2014) A community effort to assess and improve drug sensitivity prediction algorithms. *Nature biotechnology* 32(12):1202-1212.

476. Basu A*, et al.* (2013) An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell* 154(5):1151-1161.
477. Mo Q*, et al.* (2013) Pattern discovery and cancer gene identification in integrated cancer genomic data. *Proceedings of the National Academy of Sciences* 110(11):4245-4250.
478. Pardo EP & Godzik A (2015) Analysis of individual protein regions provides novel insights on cancer pharmacogenomics. *PLoS computational biology* 11(1):e1004024.
479. Kobayashi S*, et al.* (2005) EGFR mutation and resistance of non–small-cell lung cancer to gefitinib. *New England Journal of Medicine* 352(8):786-792.
480. Wong A, Tian X, Gehring C, & Marondedze C (2018) Discovery of novel functional centers with rationally designed amino acid motifs. *Computational and Structural Biotechnology Journal* 16:70-76.
481. Popovych N, Sun S, Ebright RH, & Kalodimos CG (2006) Dynamically driven protein allostery. *Nature structural & molecular biology* 13(9):831-838.
482. Rajasekaran N, Sekhar A, & Naganathan AN (2017) A universal pattern in the percolation and dissipation of protein structural perturbations. *The journal of physical chemistry letters* 8(19):4779-4784.
483. Ronda L, Bruno S, & Bettati S (2013) Tertiary and quaternary effects in the allosteric regulation of animal hemoglobins. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* 1834(9):1860-1872.
484. Hauser AS*, et al.* (2018) Pharmacogenomics of GPCR drug targets. *Cell* 172(1-2):41-54. e19.
485. Caudle K*, et al.* (2013) Clinical Pharmacogenetics Implementation Consortium guidelines for dihydropyrimidine dehydrogenase genotype and fluoropyrimidine dosing. *Clinical Pharmacology & Therapeutics* 94(6):640-645.
486. Tennessen JA*, et al.* (2012) Broad GO; Seattle GO; NHLBI Exome Sequencing Project. *Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science* 337(6090):64-69.
487. Mallick S*, et al.* (2016) The Simons genome diversity project: 300 genomes from 142 diverse populations. *Nature* 538(7624):201-206.
488. Choudhury A*, et al.* (2020) High-depth African genomes inform human migration and health. *Nature* 586(7831):741-748.
489. National Center for Biotechnology Information NLoM (2015) Database of single nucleotide polymorphisms (dbSNP). (National Center for Biotechnology Information, National Library of Medicine …).
490. Anonymous (2021) UniProt: the universal protein knowledgebase in 2021. *Nucleic acids research* 49(D1):D480-D489.
491. Sangkuhl K, Berlin DS, Altman RB, & Klein TE (2008) PharmGKB: understanding the effects of individual genetic variants. *Drug metabolism reviews* 40(4):539-551.
492. Landrum MJ*, et al.* (2016) ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic acids research* 44(D1):D862-D868.
493. Choi Y, Sims GE, Murphy S, Miller JR, & Chan AP (2012) Predicting the functional effect of amino acid substitutions and indels.
494. Adzhubei I, Jordan DM, & Sunyaev SR (2013) Predicting functional effect of human missense mutations using PolyPhen-2. *Current protocols in human genetics* 76(1):7.20. 21-27.20. 41.
495. Capriotti E & Fariselli P (2017) PhD-SNPg: a webserver and lightweight tool for scoring single nucleotide variants. *Nucleic acids research* 45(W1):W247-W252.
496. Caimano MJ*, et al.* (2014) A model system for studying the transcriptomic and physiological changes associated with mammalian host-adaptation by Leptospira interrogans serovar Copenhageni. *PLoS pathogens* 10(3):e1004004.

497. De Luca O*, et al.* (2022) Predicting dihydropyrimidine dehydrogenase deficiency and related 5-fluorouracil toxicity: opportunities and challenges of DPYD exon sequencing and the role of phenotyping assays. *International Journal of Molecular Sciences* 23(22):13923.
498. Son K (2015) Prediction of 5-Fluorouracil Toxicity Associated With Dihydropyrimidine Dehydrogenase Gene (DPYD) Polymorphism Using the Secondary Structure Prediction Programs. *Journal of Pharmacogenomics & Pharmacoproteomics* 6(4):1.
499. Amamuddy OS, Glenister M, & Bishop ÖT (2021) MDM-TASK-web: A web platform for protein dynamic residue networks and modal analysis. *bioRxiv*:2021.2001. 2029.428734.
500. Chen D*, et al.* (2016) Regulation of protein-ligand binding affinity by hydrogen bond pairing. *Science advances* 2(3):e1501240.
501. Zhang F*, et al.* (2019) Molecular mechanism regarding allosteric modulation of ligand binding and the impact of mutations on dimerization for CCR5 homodimer. *Journal of Chemical Information and Modeling* 59(5):1965-1976.
502. Laine E, Auclair C, & Tchertanov L (2012) Allosteric communication across the native and mutated KIT receptor tyrosine kinase.
503. Ogura K*, et al.* (2005) Dihydropyrimidine dehydrogenase activity in 150 healthy Japanese volunteers and identification of novel mutations. *Clinical cancer research* 11(14):5104-5111.
504. Froehlich TK, Amstutz U, Aebi S, Joerger M, & Largiadér CR (2015) Clinical importance of risk variants in the dihydropyrimidine dehydrogenase gene for the prediction of early-onset fluoropyrimidine toxicity. *International journal of cancer* 136(3):730-739.
505. Stavraka C*, et al.* (2019) Clinical implementation of pre-treatment DPYD genotyping in capecitabine-treated metastatic breast cancer patients. *Breast Cancer Research and Treatment* 175(2):511-517.
506. Foutch D, Pham B, & Shen T (2021) Protein conformational switch discerned via network centrality properties. *Computational and Structural Biotechnology Journal* 19:3599-3608.
507. Malone ER, Oliva M, Sabatini PJ, Stockley TL, & Siu LL (2020) Molecular profiling for precision cancer therapies. *Genome medicine* 12(1):1-19.
508. Jean-Quartier C, Jeanquartier F, Jurisica I, & Holzinger A (2018) In silico cancer research towards 3R. *BMC cancer* 18(1):1-12.
509. Nigam SK, Bush KT, Bhatnagar V, Poloyac SM, & Momper JD (2020) The systems biology of drug metabolizing enzymes and transporters: relevance to quantitative systems pharmacology. *Clinical Pharmacology & Therapeutics* 108(1):40-53.
510. MohammadiPeyhani H*, et al.* (2021) NICEdrug. ch, a workflow for rational drug design and systems-level analysis of drug metabolism. *Elife* 10:e65543.
511. Sharma V, Gupta SK, & Verma M (2019) Dihydropyrimidine dehydrogenase in the metabolism of the anticancer drugs. *Cancer Chemotherapy and Pharmacology* 84(6):1157-1166.
512. Tawfik DS (2014) Accuracy-rate tradeoffs: how do enzymes meet demands of selectivity and catalytic efficiency? *Current opinion in chemical biology* 21:73-80.
513. Murphy PJ (2008) The evolution of drug metabolism research. *Handbook of Drug Metabolism*, (CRC Press), pp 23-38.
514. Izidoro SC, de Melo-Minardi RC, & Pappa GL (2015) GASS: identifying enzyme active sites with genetic algorithms. *Bioinformatics* 31(6):864-870.
515. Munoz-Munoz J, Cartmell A, Terrapon N, Henrissat B, & Gilbert HJ (2017) Unusual active site location and catalytic apparatus in a glycoside hydrolase family. *Proceedings of the National Academy of Sciences* 114(19):4936-4941.
516. Riziotis IG, Ribeiro AJ, Borkakoti N, & Thornton JM (2022) Conformational variation in enzyme catalysis: A structural study on catalytic residues. *Journal of Molecular Biology* 434(7):167517.

517. Warshel A (2014) Multiscale modeling of biological functions: from enzymes to molecular machines (Nobel Lecture). *Angewandte Chemie International Edition* 53(38):10020-10031.
518. Weng Y-Z, Chang DT-H, Huang Y-F, & Lin C-W (2011) A study on the flexibility of enzyme active sites. *BMC bioinformatics* 12(1):1-11.
519. Medlock AE*, et al.* (2021) Insight into the function of active site residues in the catalytic mechanism of human ferrochelatase. *Biochemical Journal* 478(17):3239-3252.
520. Perche-Letuvée P, Molle T, Forouhar F, Mulliez E, & Atta M (2014) Wybutosine biosynthesis: structural and mechanistic overview. *RNA biology* 11(12):1508-1518.
521. Verma R & Mitchell-Koch K (2017) In Silico studies of small molecule interactions with enzymes reveal aspects of catalytic function. *Catalysts* 7(7):212.
522. Desjardins M*, et al.* (2017) Systematic Functional Analysis of Active-Site Residues in l-Threonine Dehydrogenase from Thermoplasma volcanium. *ACS omega* 2(7):3308-3314.
523. Friedman R (2022) Computational studies of protein–drug binding affinity changes upon mutations in the drug target. *Wiley Interdisciplinary Reviews: Computational Molecular Science* 12(1):e1563.
524. Zaidi ZH & Smith DL (1996) *Protein Structure--function Relationship* (Springer).
525. Bolduc JM*, et al.* (1995) Mutagenesis and Laue structures of enzyme intermediates: isocitrate dehydrogenase. *Science* 268(5215):1312-1318.
526. Schwans JP*, et al.* (2013) Use of anion–aromatic interactions to position the general base in the ketosteroid isomerase active site. *Proceedings of the National Academy of Sciences* 110(28):11308-11313.
527. Sunden F, Peck A, Salzman J, Ressl S, & Herschlag D (2015) Extensive site-directed mutagenesis reveals interconnected functional units in the alkaline phosphatase active site. *Elife* 4:e06181.
528. Zhang X & Houk K (2005) Why enzymes are proficient catalysts: beyond the Pauling paradigm. *Accounts of chemical research* 38(5):379-385.
529. Warshel A*, et al.* (2006) Electrostatic basis for enzyme catalysis. *Chemical reviews* 106(8):3210-3235.
530. Carlin DA*, et al.* (2016) Kinetic characterization of 100 glycoside hydrolase mutants enables the discovery of structural features correlated with kinetic constants. *PloS one* 11(1):e0147596.
531. Lucas JE & Siegel JB (2015) Quantitative functional characterization of conserved molecular interactions in the active site of mannitol 2-dehydrogenase. *Protein Science* 24(6):936-945.
532. Elraiyah T*, et al.* (2017) Novel deleterious dihydropyrimidine dehydrogenase variants may contribute to 5-fluorouracil sensitivity in an East African population. *Clinical Pharmacology & Therapeutics* 101(3):382-390.
533. Vehlow C*, et al.* (2011) CMView: interactive contact map visualization and analysis. *Bioinformatics* 27(11):1573-1574.
534. Kozma D, Simon I, & Tusnady GE (2012) CMWeb: an interactive on-line tool for analysing residue–residue contacts and contact prediction methods. *Nucleic acids research* 40(W1):W329-W333.
535. Ho HK, Kuiper MJ, & Kotagiri R (2008) PConPy—a Python module for generating 2D protein maps. *Bioinformatics* 24(24):2934-2935.
536. Mandelker D*, et al.* (2009) A frequent kinase domain mutation that changes the interaction between PI3Kα and the membrane. *Proceedings of the National Academy of Sciences* 106(40):16996-17001.
537. Karp JM, Sparks S, & Cowburn D (2017) Effects of FGFR2 kinase activation loop dynamics on catalytic activity. *PLoS computational biology* 13(2):e1005360.
538. Rufer AC (2021) Drug discovery for enzymes. *Drug Discovery Today* 26(4):875-886.

539. Zhao W & Meng H (2022) Effects of genetic polymorphism of drug-metabolizing enzymes on the plasma concentrations of antiepileptic drugs in Chinese population. *Bioengineered* 13(3):7709-7745.
540. Sileshi T, Mekonen G, Makonnen E, & Aklillu E (2022) Effect of Genetic Variations in Drug-Metabolizing Enzymes and Drug Transporters on the Pharmacokinetics of Rifamycins: A Systematic Review. *Pharmacogenomics and Personalized Medicine* 15:561.
541. Fogel DB (2018) Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: a review. *Contemporary clinical trials communications* 11:156-164.
542. Kiriiri GK, Njogu PM, & Mwangi AN (2020) Exploring different approaches to improve the success of drug discovery and development projects: a review. *Future Journal of Pharmaceutical Sciences* 6(1):1-12.
543. Lunenburg CA*, et al.* (2016) Prospective DPYD genotyping to reduce the risk of fluoropyrimidine-induced severe toxicity: ready for prime time. *European Journal of Cancer* 54:40-48.
544. Amorosi CJ*, et al.* (2021) Massively parallel characterization of CYP2C9 variant enzyme activity and abundance. *The American Journal of Human Genetics* 108(9):1735-1751.
545. Fowler DM & Fields S (2014) Deep mutational scanning: a new style of protein science. *Nature methods* 11(8):801-807.
546. Matreyek KA*, et al.* (2018) Multiplex assessment of protein variant abundance by massively parallel sequencing. *Nature genetics* 50(6):874-882.