

Identification of Novel Compounds Against *Plasmodium falciparum* Cytochrome *bc*₁ Complex Inhibiting the Trans-Membrane Electron Transfer Pathway: An *In Silico* Study

A thesis submitted in fulfilment of the requirements for the degree

of

DOCTOR OF PHILOSOPHY
IN BIOINFORMATICS

of

RHODES UNIVERSITY, SOUTH AFRICA

Research Unit in Bioinformatics (RUBi)

DEPARTMENT OF BIOCHEMISTRY and MICROBIOLOGY

Faculty of Science

by

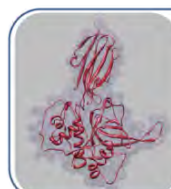
Lorna Jemosop Chebon

ORCID ID: <https://orcid.org/0000-0001-7236-8750>

August 2022



RHODES UNIVERSITY
Where leaders learn



RUBi
Research Unit in Bioinformatics

ABSTRACT

Malaria continues to be a burden globally with a myriad of challenges deterring eradication efforts. With most antimalarials facing drug resistance, such as atovaquone (ATQ), alternative compounds that can withstand resistance are warranted. The *Plasmodium falciparum* cytochrome *b* (PfCytb), a subunit of *P. falciparum* cytochrome *bc₁* complex, is a validated drug target. Structurally, cytochrome *b*, cytochrome *c₁*, and iron sulphur protein (ISP) subunits form the catalytic domain of the protein complex having heme bL, heme bH and iron-sulphur [2FE-2S] cluster cofactors. These cofactors have redox centres to aid in the electron transfer (ET) process. These subunits promote ET mainly through the enzyme's ubiquinol oxidation (Qo) and ubiquinone reduction (Qi) processes in the catalytic domain. ATQ drug has been used in the prevention and treatment of uncomplicated malaria by targeting PfCytb protein. Once the mitochondrial transmembrane ET pathway is inhibited, it causes a collapse in its membrane potential. Previously reported ATQ drug resistance has been associated with the point mutations Y268C, Y268N and Y268S. Thus, in finding alternatives to the ATQ drug, this research aimed to: i) employ *in silico* approaches incorporating protein into phospholipid bilayer for the first time to understand the parasites' resistance mechanism; ii) determine any sequence and structural differences that could be explored in drug design studies; and iii) screen for PfCytb-iron sulphur protein (Cytb-ISP) hit compounds from South African natural compound database (SANCDB) and Medicines for Malaria Venture (MMV) that can withstand the identified mutations.

Using computational tools, comparative sequence and structural analyses were performed on the cytochrome *b* protein, where the ultimate focus was on *P. falciparum* cytochrome *b* and its human homolog. Through multiple sequence alignment, motif discovery and phylogeny, differences between *P. falciparum* and *H. sapiens* cytochrome *b* were identified. Protein

modelling of both *P. falciparum* and *H. sapiens* cytochrome *b* - iron sulphur protein (PfCytb-ISP and HsCytb-ISP) was performed. Results showed that at the sequence level, there were few amino acid residue differences because the protein is highly conserved. Important to note is the four-residue deletion in *Plasmodium* spp. absent in the human homolog. Motif analysis discovered five unique motifs in *P. falciparum* cytochrome *b* protein which were mapped onto the predicted protein model. These motifs were not in regions of functional importance; hence their function is still unknown. At a structural level, the four-residue deletion was observed to alter the Qo substrate binding pocket as reported in previous studies and confirmed in this study. This deletion resulted in a 0.83 Å structural displacement. Also, there are currently no *in silico* studies that have performed experiments with *P. falciparum* cytochrome *b* protein incorporated into a phospholipid bilayer.

Using 350 ns molecular dynamics (MD) simulations of the holo and ATQ-bound systems, the study highlighted the resistance mechanism of the parasite protein where the loss of active site residue-residue interactions was identified, all linked to the three mutations. The identified compromised interactions are likely to destabilise the protein's function, specifically in the Qo substrate binding site. This showed the possible effect of mutations on ATQ drug activity, where all three mutations were reported to share a similar resistance mechanism.

Thereafter, this research work utilised *in silico* approaches where both Qo active site and interface pocket were targeted by screening the South African natural compounds database (SANCDB) and Medicines for Malaria Venture (MMV) compounds to identify novel selective hits. SANCDB compounds are known for their structural complexity that preserves the potency of the drug molecule. Both SANCDB and MMV compounds have not been explored as inhibitors against the PfCytb drug target. Molecular docking, molecular dynamics (MD) simulations, principal component, and dynamic residue network (DRN; global and local) analyses were utilised to identify and confirm the potential selective inhibitors. Docking results

identified compounds that bound selectively onto PfCytb-ISP with a binding energy ≤ -8.7 kcal/mol⁻¹.

Further, this work validated a total of eight potential selective compounds to inhibit PfCytb-ISP protein (Qo active site) not only in the wild-type but also in the presence of the point mutations Y268C, Y268N and Y268S. The selective binding of these hit compounds could be linked to the differences reported at sequence/residue level in chapter 3. DRN and residue contact map analyses of the eight compounds in holo and ligand-bound systems revealed reduced residue interactions and decreased protein communication. This suggests that the eight compounds show the possibility of inhibiting the parasite and disrupting important residue-residue interactions. Additionally, 13 selective compounds were identified to bind at the protein's heterodimer interface, where global and local analysis confirmed their effect on active site residues (distal location) as well as on the communication network. Based on the sequence differences between PfCytb and the human homolog, these findings suggest these selective compounds as potential allosteric modulators of the parasite enzyme, which may serve as possible replacements of the already resistant ATQ drug. Therefore, these findings pave the way for further *in vitro* studies to establish their anti-plasmodial inhibition levels.

DECLARATION

I, **LORNA JEMOSOP CHEBON**, declare that this is my own original work, except where duly acknowledged. It is being submitted for the degree of Doctor of Philosophy in Bioinformatics in the Faculty of Science of Rhodes University. It has not been submitted before for any degree for examination in any other university.



Date: 8th August 2022

DEDICATION

To my dear husband Nelson Kibichii Bore whose love, sacrifice, support, understanding and commitment carried me through this journey. I love you.

To my daughter Kloe Nandi Bore whose laughter, presence and smiley faces kept me going through the tough times. You have taught me that being a mom might mean that my hands are full, but so is my heart. I love you.

To my parents Gideon and Zipporah Chebon who, besides their love and encouragement, sacrificed their time and money to educate and support me in my career path. I love you.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank the Almighty God for His sufficient grace through this journey, good health, wisdom, knowledge, strength and peace of mind to carry out this research. Thank you, Lord, for the support you gave me through family, special friends and colleagues.

Also, I would also like to thank my supervisor Prof. Özlem Tastan Bishop for incredible support she gave me over the last few years. She has challenged and inspired me intellectually and inspired me always to give my best. For these reasons and more, I would like to say, ‘Thank you’.

I would like to thank my mentor LTC Edwin Kamau who believed in me and supported me in pursuing the Bioinformatics course. I thank him for his intellectual support and inspiration.

Most of the computational experiments including simulations were performed on the Centre for High Performance Computing (CHPC), Cape Town, South Africa.

The entire RUBi research group has been supportive, and their contribution and intellectual support has allowed me to grow in Bioinformatics research. To mention a few, thank you to Rolland Bantar Tata, Thomas Musyoka Mutemi, Afrah Khairallah and Allan Sanyanga for their fruitful discussions and support in script writing. Thank you to my friends and Grahamstown church family that have encouraged me all the way.

Finally, this work would not have been possible without family support, thank you to the Bore and Chebon family for their love, encouragement and prayers that kept me going in this career.

Funding acknowledgement

I would also like to thank my funders whose financial support ensured I completed my research work successfully: DELTAS Africa Initiative [grant 107740/Z/15/Z]. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)’s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa’s Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [DELGEME grant 107740/Z/15/Z] and the UK government.

TABLE OF CONTENTS

ABSTRACT.....	i
DECLARATION	iv
DEDICATION.....	v
ACKNOWLEDGEMENTS.....	vi
TABLE OF CONTENTS.....	vii
LIST OF FIGURES	xii
LIST OF TABLES.....	xv
LIST OF SUPPLEMENTARY FIGURES AND TABLES	xvii
RESEARCH OUTPUTS.....	xx
LIST OF WEB SERVERS	xxi
LIST OF ABBREVIATIONS	xxii
THESIS OVERVIEW.....	xxv
CHAPTER 1	1
LITERATURE REVIEW	1
1.1 Malaria: Scope of the disease burden.....	2
1.2 <i>Plasmodium</i> parasites.....	4
1.3 <i>Plasmodium</i> parasite transmission and life cycle.....	5
1.4 Malaria disease management	6
1.4.1 Diagnosis of malaria	6
1.4.2 Prevention of malaria.....	7
1.4.3 Treatment of malaria.....	8
1.5 Antimalarial drug targets.....	9
1.6 Antimalarial drug resistance.....	11
1.7 Atovaquone	12
1.7.1 Mechanism of ATQ drug action	13
1.7.2 ATQ resistance and mutations	14
1.7.3 Review on current ATQ alternatives	16
1.8 Problem statement and justification	17
1.9 Hypothesis.....	19
1.10 Research aim and objectives	19
1.10.1 Aim	19
1.10.2 Objectives	20

CHAPTER 2.....	22
IN SILICO METHODS AND TOOLS FOR DRUG DISCOVERY	22
2.1 Introduction to drug discovery	23
2.1.1 Antimalarial drug history and discovery.....	23
2.2 Drug discovery methods; molecular similarity and mapping techniques	25
2.2.1 Multiple sequence alignment	25
2.2.2 Phylogenetic tree calculations.....	26
2.3 Molecular protein modelling.....	27
2.4 Molecular docking.....	29
2.5 Molecular dynamic simulations	31
2.6 Dynamic residue network (DRN) analysis.....	32
CHAPTER 3.....	34
COMPARATIVE SEQUENCE AND STRUCTURAL ANALYSIS	34
3.1 Cytochrome <i>bc₁</i> complex.....	35
3.1.1 Structure and function.....	35
3.1.2 <i>Plasmodium falciparum</i> cytochrome <i>bc₁</i> complex.....	37
3.1.3 Electron transfer process.....	37
3.1.4 Protein structural features and location of point mutations	38
3.1.5 Sequence characterisation of <i>P. falciparum</i> cytochrome <i>b</i> protein	40
3.2 Methodology	41
3.2.1 Sequence retrieval and characterisation at residue and structure level.....	41
3.2.2 Multiple sequence alignment	42
3.2.3 SNP identification and prediction of their effect on cytochrome <i>b</i> modelled protein structure	42
3.2.4 Phylogenetic tree calculations.....	43
3.2.5 Protein modelling of the Qo oxidation site of the PfCytb-ISP	44
3.2.6 Motif calculations and structural mapping analysis.....	45
3.3 Results and discussion.....	46
3.3.1 Sequence structural and functional analysis	46
3.3.2 Modelling results for PfCytb-ISP protein	51
3.3.3 Structural analysis of <i>P. falciparum</i> cytochrome <i>b</i>	53
3.3.4 Prediction analysis on the effect of mutations on proteins' sequence and structure	55
3.3.5 Cytochrome <i>b</i> multiple sequence alignment.....	59
3.3.6 Evolutionary analysis using phylogenetic trees	61

3.3.7 Motif analysis and structural mapping of the conserved and unique motifs	65
3.4 Chapter conclusion.....	74
CHAPTER 4.....	76
FE²⁺ METAL PARAMETERISATION	76
4.1 Introduction	77
4.2 Methodology	81
4.2.1 Structure preparation; protonation of the protein.....	81
4.2.2 DFT Calculations	81
4.2.2.1 Geometry optimisation.....	81
4.2.2.2 Force field generation.....	82
4.2.2.3 RESP charge calculations.....	84
4.2.3 Force field validation by MD simulations	85
4.2.3.1 POPC: POPE lipid bilayer construction and topology generation.....	85
4.2.3.2 System energy minimisation, heating, equilibration, and MD production run	86
4.3 Results and discussion.....	87
4.3.1 Protonation states for PfCytb-ISP protein metal coordinating residues	87
4.3.2 Membrane and topology construction; MD simulations	89
4.3.3 Derived force field parameters; Bond distance and angle	91
4.3.4 RESP (Partial atomic) charges.....	93
4.3.5 Validation of the force field parameters using MD simulations.....	95
4.4 Chapter conclusion.....	98
CHAPTER 5.....	99
DOCKING AND MOLECULAR DYNAMICS STUDIES.....	99
5.1 Introduction	100
5.2 Methodology	104
5.2.1 Protein modelling; WT and mutant proteins (HsCytb-ISP).....	104
5.2.2 Allosteric pocket identification in PfCytb-ISP protein specifically interface pocket.....	104
5.2.3 Retrieval of drug compounds and Molecular docking approach	105
5.2.4 Molecular dynamics simulations and trajectory analysis	106
5.2.4.1 POPC: POPE Lipid Bilayer Construction and Topology Generation.....	106
5.2.4.2 Molecular dynamics production runs	107

5.2.5 Hydrogen bond calculations	108
5.2.6 Dynamic residue network (DRN) and Residue contact map calculations	109
5.2.7 Principal Component Analysis	111
5.3 Results and discussion.....	112
5.3.1 Structure and sequence conservation of PfCytb-ISP as an attractive drug target	112
5.3.2 Protein structure and modelling of the Qo active site; PfCytb-ISP and HsCytb-ISP	113
PART I	115
EFFECT OF MUTATION ON ATQ DRUG ACTIVITY.....	115
5.3.3 Global MD analysis of holo protein and ATQ-bound systems in a POPC: POPE phospholipid bilayer.....	115
5.3.3.1 ATQ docking results and trajectory analysis via MD simulations.....	115
5.3.3.2 Trajectory analysis and dynamics of the holo protein and ATQ-bound systems	117
5.3.4 Analysis of conformational entropy in holo and ATQ-bound proteins comparing WT and mutant proteins.....	122
5.3.5 Local residue interaction analysis using weighted residue contact maps	125
5.3.6 Effect of mutation on protein communication and residue accessibility	132
PART II.....	139
TARGET 1: Qo ACTIVE SITE.....	139
5.3.7 Docking analysis for Qo active site; Potential hits identified through virtual screening	139
5.3.8 Molecular dynamics trajectory analysis: Eight potential hits identified out of 29 selective compounds via MD simulations	140
5.3.9 General information on the potential hit compounds	144
5.3.10 Global protein dynamics analysis: Root mean square deviations (RMSD) and radius of gyration (Rg).....	145
5.3.11 Principal component and free energy analysis.....	152
5.3.12 Ligand stability during MD simulations	156
5.3.13 Effect of the eight potential hit compounds on protein residue flexibility	161
5.3.14 Hydrogen bond, vdW, and other interactions maintaining the stability of protein-ligand complexes.....	163
5.3.15 Dynamic Residue Network (DRN); Global Residue Network analysis	165

5.3.15.1 Five centrality measures identified key residues ‘hubs’ important in protein function.....	166
5.3.16 Effect of ligand binding and mutations on intra-protein communication in the eight protein-ligand systems	175
5.3.17 DRN; Local Residue Network Analysis	180
5.3.17.1 Residue-residue interactions within the active site residues	180
PART III	187
TARGET 2: PLASMODIUM INTERFACE	187
5.3.18 PfCytb-ISP interface pocket prediction analysis	187
5.3.19 Docking output: Thirteen selective compounds were identified to bind in the PfCytb-ISP heterodimer interface	190
5.3.20 Analysis of the stability of the interface selective compounds in the <i>Plasmodium</i> interface pocket via molecular dynamics simulations	192
5.3.21 Protein backbone flexibility and structural conformations in ligand systems and in the presence of mutations	201
5.3.22 DRN centrality metrics analysis to determine protein communication patterns and ‘hubs’ upon binding of compounds within the PfCytb-ISP interface pocket	203
5.3.23 Effect of mutation on residue-residue interaction changes on 13 interface ligand systems; Contact map analysis.....	206
5.4 Chapter conclusion.....	211
CHAPTER 6.....	214
CONCLUSIONS AND RECOMMENDATIONS.....	214
BIBLIOGRAPHY	219
SUPPLEMENTARY DATA	259

LIST OF FIGURES

Figure 1.1: Global estimate of malaria indigenous prevalence/incidence cases. Map adapted from [1].	3
Figure 1.2: Life cycle of <i>Plasmodium</i> in the host and vector stages.	6
Figure 3.1: The structural model of the dimeric <i>bc1</i> complex in bovine (PDB ID: 1PP9).	36
Figure 3.3: 3D cartoon representation of PfCytb-ISP protein complex showing the location of Qo active site and the location of point mutation Y268C/N/S.	39
Figure 3.4: Secondary structure prediction between i) <i>P. falciparum</i> cytochrome <i>b</i> and ii) human homolog using PSIPRED.	48
Figure 3.5: Graphical results of cytochrome <i>b</i> functional domains in 10 Cyt <i>b</i> protein sequences including <i>Plasmodium</i> spp., human homolog, and other species for comparative purposes. This was performed using ProDOM server [159].	49
Figure 3.6: Predicted protein model validation in the form of Ramachandran plots.	53
Figure 3.7: A superimposition of <i>P. falciparum</i> and <i>H. sapiens</i> cytochrome <i>b</i> structures to highlight their structural differences. A colour scale for key structural features is provided.	54
Figure 3.8: Illustration of interatomic interactions occurring around residue 268 in the WT (wild-type) and Y268C, Y268N and Y268S mutant proteins.]	58
Figure 3.9: Multiple sequence alignment of 10 cytochrome <i>b</i> protein sequences. Alignment and sequence visualisation were performed using PROMALS3D and Jalview, respectively.	60
Figure 3.10: (A) Circular phylogenetic tree of cytb protein from 74 species calculated using RAxML software and visualised using Dendroscope.	64
Figure 3.11a: MEME motifs of cytb proteins in the first set of protein sequences;	66
Figure 3.11b: MEME motifs of cytb proteins in the second set of protein sequences;	67
Figure 3.11c: MEME motifs of cytb proteins in the third set of protein sequences;	68
Figure 3.12: Structural mapping of nine conserved motifs onto the (A) <i>P. falciparum</i> and (B) human homolog e.	69
Figure 3.13: Motif analysis of 74 cytochrome <i>b</i> protein sequences including Plasmodial and human homolog.	73
Figure 4.1: Chemical structures of the metal cofactors showing their geometry and residue coordination; heme bL, heme bH and [2FE-2S] cluster. CN is the coordination number.	79
Figure 4.2: Structural representation of membrane construction with PfCytb-ISP protein.]	89

Figure 4.3: Graphical stick representation of the Quantum Mechanics optimised sub-structures of the three prosthetic groups:].	91
Figure 4.4: A schematic representation of the partial atomic charges (RESP method).	94
Figure 4.5: Metal bond distances monitored throughout 350 ns MD simulations..	96
Figure 4.6: Comparison of bond distances monitored during 350 ns MD simulations in the mutant systems (Y268C, Y268N and Y268S).	97
Figure 5.1: Overview of PfCytb-ISP protein structure.	112
Figure 5.2: 3D and 2D ligand interactions of ATQ drug in docked pose and across different time steps during 350 ns simulations.	116
Figure 5.3: Molecular dynamics trajectory analysis of the changes to protein root mean square deviation (RMSD) in both the holo and ATQ bound systems.	119
Figure 5.4: Principal component analysis (PCA) of the holo and ATQ bound proteins..	124
Figure 5.5: Residue contact heat maps for PfCytb-ISP active site residues..	127
Figure 5.6: Dynamic residue network (DRN) analysis.	133
Figure 5.7: Effect of the three mutations Y268C, Y268N and Y268S on the protein intra-communication in the PfCytb-ISP protein.	136
Figure 5.8: A surface representation of the interface residues in the HsCytb-ISP protein...	140
Figure 5.9: 3D ligand interaction network for eight PfCytb-ISP selective inhibitors in the Qo active site pocket: .	142
Figure 5.10: 2D structures of the ATQ (reference) MMV and SANCDB compounds..	144
Figure 5.11: RMSD Line graph presentation of eight ligand systems.	146
Figure 5.12: RMSD density distribution for PfCytb-ISP holo and 8 ligand systems in WT, Y268C, Y268N and Y268S states across 350 ns MD simulations.	148
Figure 5.13: (A) Histogram plot showing backbone RMSD density distribution for PfCytb-ISP holo and 8 ligand systems.	150
Figure 5.14: Radius of gyrations line plots for PfCytb-ISP holo and eight ligand systems in WT, Y268C, Y268N and Y268S systems. A colour scale has been provided.	152
Figure 5.15: 3D Principal component analysis (PCA) plot for holo and eight protein-ligand systems (MMV and SANCDB potential hits)..	154
Figure 5.16: Ligand root mean square deviation (RMSD) for eight ligand systems.	157
Figure 5.17: Ligand RMSD for eight ligands during 350 ns MD simulation performed in WT, Y268C, Y268N and Y268S systems.	159
Figure 5.18: Centre of mass (COM) distance between ligand and the Qo active site.	160
Figure 5.19: (A) Root mean square fluctuations comparing ligand-free and ligand-bound protein systems for 8 ligands.	167
Figure 5.21: A heat map representation of top 4 % <i>Averaged CC</i> .	168

Figure 5.22: A heat map representation of top 4 % <i>Averaged DC</i> .	169
Figure 5.23: A heat map representation of top 4 % <i>Averaged EC</i> .	170
Figure 5.24: A heat map representation of top 4 % <i>Averaged KC</i> .	171
Figure 5.25: Structural mapping of the top 4 % residues as depicted by five DRN centrality metrics (<i>Averaged BC, CC, DC, EC, and KC</i>).	173
Figure 5.26: Structural mapping of significant residues onto PfCytb-ISP protein in WT and mutant models (Y268C, Y268N and Y268S).	176
Figure 5.27: A heat map representation of significant residues affecting protein communication; change in <i>BC</i> .	178
Figure 5.28: Structural diagram of PfCytb-ISP protein showing the potential interface pocket (circled).	187
	188
Figure 5.29: Structural representation of the identified pockets in the PfCytb-ISP protein using FTMap.	189
Figure 5.30: Virtual screening workflow showing filtering criteria based on binding site (<i>Plasmodium</i> interface).	190
Figure 5.31: 2D structures of (A) MMV and (B) SANCDB compounds that bound in the PfCytb-ISP heterodimer interface with binding energies < -8.7 kcal/mol-1.	191
Figure 5.32: 3D ligand interaction network for 13 selective compounds bound in the PfCytb-ISP interface pocket as highlighted in the figure.	193
Figure 5.33: Kernel density distributions of (A) Ligand RMSD and (B) COM for 13 interface selective compounds monitored across 350 ns MD simulation.	195
Figure 5.34: A plot of the average number of inter-molecular hydrogen bonds formed between the seven MMV compounds and PfCytb-ISP protein	198
Figure 5.35: A plot of the average number of intermolecular hydrogen bonds formed between the 6 SANCDB compounds and PfCytb-ISP protein.	200
Figure 5.36: Violin plot presentation of (A) RMSD and (B) Rg for 13 ligand systems during 350 ns MD simulations.	202

LIST OF TABLES

Table 1.1: Classification of antimalarial drugs [30,31,36,37]	9
Table 1.2: Summary of validated target (VT) sites (VT3 and VT4) and their inhibitors [4] .	10
Table 2.1: Dynamic residue network centrality metrics used to identify residues crucial for communication in FPs systems. Taken from [137].	33
Table 3.1: Physico-chemical properties of the cytochrome <i>b</i> amino acid sequence of 10 species.	47
Table 3.2: Functional domains that were common across all the 10 cytochrome <i>b</i> protein sequences.	50
Table 3.3: Prediction of the effect of mutations (Y268C, Y268N and Y268S) on PfCytb-ISP protein stability using VAPOR and DynaMut.	57
Table 3.4: Conserved cytb motif sequences (n = 9) identified across 74 protein sequences using MEME suite [179].	70
Table 4.1: Computed pKa (1/2) values for ionisable groups (side chains) that are coordinated in the metal centre of the PfCytb-ISP protein complex (Internal dielectric of 10).	88
Table 4.2: Molecular composition of the simulation systems. Reproduced from [140].	90
Table 4.4: A table showing derived metal force field parameters. Comparison of Fe ²⁺ bond distance (between Fe ²⁺ metal and its interacting residues) in template structure, derived metal parameters and post-MD.	92
Table 4.4: Partial atomic charges derived before (formal charge) and after charge calculations.	95
Table 5.1: Chemical information and pharmacokinetic profile of the selective inhibitors derived from MMV and SANCDB compounds.	143
Table 5.2: Mean protein backbone RMSD values for WT and mutant (Y268C, Y268N and Y268S) ligand-bound systems. Values in Å.	147
Table 5.3: Eigenvalues in fractions representing PC1, PC2 and PC3 for each of the eight ligand systems in both WT and mutant systems;	153
Table 5.4: Significant residues common across eight ligand systems in each protein system.	174
Table 5.5: Active site residue-residue interactions observed to be either reduced, lost or gained in the eight ligand systems).	184
Table 5.6: Protein communication ‘hubs’ identified from holo, ATQ and 13 ligand systems across five DRN metrics in WT and mutant proteins.	204
Table 5.7: Residue-residue interactions observed to be either reduced/lost or gained/enhanced in the 13 ligand systems.	208

LIST OF SUPPLEMENTARY FIGURES AND TABLES

Table S3.1: Retrieved cytochrome <i>b</i> sequences from a total of 72 organisms including <i>Plasmodium</i> and human homolog. # names used for convenience.	259
Figure S3.1: Secondary structure prediction of the three mutations (Y268C/N/S) in comparison to the wild-type.....	263
Figure S3.2: Multiple sequence alignment results generated by PROMALS3D which uses ClustalX colouring scheme.	265
Figure S3.3: Regular expressions of predicted conserved motifs comparing PfCytb and HsCytb proteins.	266
Table S4.1: Titratable residues and pKa values in the PfCytb-ISP protein complex. Reproduced from [140].	267
Table S4.2. Parameter and coordinate files for both Heme & [2FE-2S] cluster model B3LYP/6-31G*:	269
Figure S5.1: HsCytb-ISP model validation in form of Ramachandran plots..	285
Figure S5.2: Rg comparison of WT and mutant both in the presence and absence of ATQ during the 350 ns MD simulation.	286
Figure S5.3: Line graph comparison of RMSF as observed in the WT and mutant PfCytb-ISP protein systems during 350 ns MD simulation	287
Figure S5.4: Residue contact heatmap between active site residues (n = 28) and other residue contacts..	288
Figure S5.5: Residue contact heatmap between active site residues (n = 28) and other residue contacts..	289
Figure S5.6: Residue contact heat map between active site residues (n = 28) and other residue contacts.	290
Figure S5.7: Residue contact heat map of Y268C, Y268N, Y268S mutant systems with reference to wild type yielded during the last 100 ns of MD simulations.....	291
Table S5.1: Protein residues having high normalized <i>Betweenness Centrality</i> (<i>BC</i>) values in holo WT and three mutant systems across the triplicate runs.	292
Table S5.2: Residues having high <i>Betweenness Centrality</i> (<i>BC</i>) values in ATQ-bound protein systems (WT, Y268C, Y268N and Y268S) to assess effect of mutation.....	293
Figure S5.8: Heat map comparison of <i>BC</i> values in PfCytb-ISP holo and ATQ-bound proteins. High <i>BC</i> values indicate regions of high residue connectivity. Reproduced from [140].	294
Figure S5.9: Density plot for delta <i>BC</i> and <i>L</i> values (N =503) for mutant systems (Y268C, Y268N, and Y268S) in three replicate runs..	295
Table S5.3: Table of residues having significant changes in <i>Betweenness Centrality</i> (ΔBC) in three holo mutant systems to assess effect of mutation].	296

Table S5.4: Table of residues having significant changes in <i>Betweenness Centrality</i> (ΔBC) in three ATQ-bound mutant systems to assess effect of mutation.....	297
Table S5.5: List of a total of 38 inhibitors (MMV and SANCDB) docked into the Qo active site	298
Figure S5.10: A schematic diagram of the super-imposition of the co-crystallized ligand stigmatellin (SMA) re-docked and cross-docked.	299
Figure S5.12: Analysis of average number of intermolecular hydrogen bonds formed between the MMV potential hits and PfCytb-ISP protein.	301
Figure S5.13: Analysis of average number of intermolecular hydrogen bonds formed between the SANCDB potential hits and PfCytb-ISP proteins.	302
Table S5.6: Hydrogen bond (Hbond) and van der Waals (vdW) molecular interactions formed in eight protein-ligand complexes.	303
Figure S5.14: Structural mapping of top 4% BC residues onto PfCytb-ISP protein in WT and mutant models (Y268C, Y268N, and Y268S)..	307
Figure S5.15: Structural mapping of top 4% CC residues onto PfCytb-ISP protein in WT and mutant models (Y268C, Y268N, and Y268S)..	308
Figure S5.16: Structural mapping of top 4% DC residues onto PfCytb-ISP protein in WT and mutant models (Y268C, Y268N, and Y268S)..	309
Figure S5.17: Structural mapping of top 4% EC residues onto PfCytb-ISP protein in WT and mutant models (Y268C, Y268N, and Y268S).	310
Figure S5.18: Structural mapping of top 4% KC residues onto PfCytb-ISP protein in WT and mutant models (Y268C, Y268N, and Y268S)..	311
Table S5.7: Residues having significant ΔBC for eight protein-ligand systems showing conformational/structural changes.	312
Table S5.8: Residues having significant ΔBC for eight protein-ligand systems showing conformational/structural changes ;.....	314
Figure S5.19: A heat map representation of RMSF in holo, ATQ and thirteen ligand systems during 350 ns MD simulations. White to dark orange represent the degree of protein flexibility going from more flexible to less flexible respectively.	316
Figure S5.20: A heat map representation of top 4% BC values representing significant residues affecting protein communication.....	317
Figure S5.21: A heat map representation of top 4% CC values representing significant residues affecting protein communication.....	318
Figure S5.22: A heat map representation of top 4% DC values representing significant residues affecting protein communication.....	319
Figure S5.23: A heat map representation of top 4% EC values representing significant residues affecting protein communication.....	320
Figure S5.24: A heat map representation of top 4% KC values representing significant residues affecting protein communication across 13 ligand systems in their WT and mutant states. Ligand systems 1-13 representing MMV020982, MMV023860, MMV024101, MMV024937, MMV676050,	

MMV676599, MMV687700, SANC00257, SANC00264, SANC00347, SANC00559, SANC00587 and SANC00595 respectively. Active site residues and metal interacting residues are displayed in purple and green dots respectively. 321

Table S5.9: Residue-residue interactions observed to be either reduced, lost, or gained/enhanced around the hubs 322

Table S5.10: Residue-residue interactions observed to be either reduced, lost, or gained/enhanced around the hubs 323

RESEARCH OUTPUTS

Publications and contributions

1. **Chebon-Bore L, Sanyanga TA, Manyumwa CV, Khairallah A, Tastan Bishop Ö: 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*. 2021; 22(4):2138. <https://doi.org/10.3390/ijms22042138>**

Contributions: The project was designed by Prof Özlem Tastan Bishop. I performed metal parameter, membrane MD simulations and data analysis for this article. Sanyanga TA set up membrane MD experiments, wrote scripts for trajectory analysis as well as guiding in metal parameter calculations, Manyumwa CV and Khairalla A contributed to establishing metal parameter calculations. I wrote the first draft manuscript under the guidance of Prof Özlem Tastan Bishop and co-authors contributed to the peer review process.

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

Contributions: I participated in metal parameterization of the iron metal of iron-sulphur clusters and data analysis of the metal bond distances monitored throughout MD simulation, image generation and peer review of the manuscript draft.

Conference poster presentations

1. **Lorna Chebon-Bore, Colleen Manyumwa, Özlem Tastan Bishop: “Reliable charge parameterization of cofactors for the *Plasmodium falciparum* cytochrome b and iron-sulfur protein complex for inhibitory drug design”. *ISMBECCB 2019, 21-25 July 2019, Basel, Switzerland*. Published 06 Sep 2019 (<https://doi.org/10.7490/f1000research.1117472.1>)**

LIST OF WEB SERVERS

<u>Abbreviation</u>	<u>Link</u>
SANCDDB	https://sancdb.rubi.ru.ac.za/
PDB	https://www.rcsb.org/
MDM-TASK	https://mdmtaskweb.rubi.ru.ac.za/
SWISS-ADME	http://www.swissadme.ch/
PIC	http://pic.mbu.iisc.ernet.in/
I-TASSER	https://zhanggroup.org/I-TASSER/
PROCHECK	https://servicesn.mbi.ucla.edu/PROCHECK/
HHpred	https://toolkit.tuebingen.mpg.de/tools/hhpred
NCBI-BLAST	https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins
FTMap	https://ftmap.bu.edu/serverhelp.php
PRIMO	https://primo.rubi.ru.ac.za/
H++	http://biophysics.cs.vt.edu/

LIST OF ABBREVIATIONS

<u>Abbreviation</u>	<u>Description</u>
2D	2-Dimensional
3D	3-Dimensional
ACT	Artemisinin-based Combination therapies
AMBER	Assisted Model building with Energy Refinement
ATP	Adenosine triphosphate
ATQ	Atovaquone
ATQ-R	Atovaquone Resistant
BC	Betweenness centrality
BLAST	Basic Local Alignment Search Tool
BLOSUM	BLOcks SUBstitution Matrix
CADD	Computer-Aided Drug Discovery
CC	Closeness centrality
COM	Centre of mass
DC	Degree centrality
DEL	DNA-encoded library
DFT	Density functional theory
DRN	Dynamic residue network
EC	Eigencentality
ELISA	Enzyme-Linked Immunosorbent Assays
EPR	Electron paramagnetic resonance
ETC	Electron Transport Chain
FRET	Forster resonance energy transfer
GA	Genetic algorithm
GPU	Graphics processing unit
GTS	Global Technical Strategy
GUI	Graphical user interface
GRAVY	Grand average of hydropathicity
IFA	Indirect immunofluorescence assay
IMM	Inner mitochondrial membrane

IRS	Indoor Residual Spraying
ISP	Iron-Sulphur Protein
KC	Katz centrality
L	Average shortest path
LAMP	Loop-Mediated Isothermal Amplification
LLIN	Long-Lasting Insecticide-treated Nets
MAFFT	Multiple Alignment using Fast Fourier Transform
MAST	Motif alignment and search tool
MCSS	Multiple copy simultaneous search
MD	Molecular dynamics
MEGA	Molecular Evolutionary Genetic Analysis
MEP	Molecular electrostatic potential
MMV	Medicines for Malaria Venture
MM	Molecular mechanics
MSA	Multiple Sequence Alignment
MUSCLE	Multiple Sequence Comparison by Log-Expectation
NCBI	National Center for Biotechnology Information
NMR	Nuclear magnetic resonance
OPM	Orientations of Proteins in Membranes
PCA	Principal component analysis
PC1	Principal component 1
PC2	Principal component 2
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
POPE	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine
PPM	Positioning of proteins in membranes
PROMALS3D	PROfile Multiple Alignment with predicted Local Structures and 3D constraints
QM	Quantum mechanics
RDT	Rapid Diagnostic Tests
RIN	Residue interaction network
RESP	Restrained electrostatic potential

RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
Rg	Radius of gyration
SANCDB	South African Natural Compounds Database
SARS-CoV2	Severe Acute Respiratory Syndrome Coronavirus 2
SNP	Single Nucleotide Polymorphisms
TIP3P	Three-Point Transferable Intermolecular Potentials
VAPOR	Variant Analysis Portal
VMD	Visual Molecular Dynamics
WHO	World Health Organization
WT	Wild-type

THESIS OVERVIEW

The main objective of the thesis was to identify novel selective compounds against *Plasmodium falciparum* cytochrome *b* drug target; compounds that can withstand atovaquone-linked point mutations. Additionally, the study sought to identify other selective compounds binding at the protein's heterodimer interface while having allosteric properties.

Chapter 1 introduces malaria burden (including but not limited to prevalence and mortality) worldwide, highlighting the *Plasmodium* parasite as the cause of disease. It describes the parasite's life cycle, diagnosis and antimalarial drugs used in the treatment of malaria disease. Also, insights on drug discovery efforts including known drug targets are discussed.

Chapter 2 discusses computational approaches and tools that have since been employed in structure-based drug designs. This includes but not limited to molecular similarity tools (sequence alignments, phylogenetic tree), molecular modelling and docking approaches and molecular dynamics simulations.

Chapter 3 focuses on cytochrome *b* comparative sequence and structure analysis to determine differences between *P. falciparum* cytochrome *b* and its human homolog. Using *in silico* approaches such as multiple sequence alignment, motif discovery, phylogeny and physico-chemical properties, differences were identified that could be explored in selective inhibitory drug design studies.

Chapter 4 details Fe²⁺ force field parameters derived using quantum mechanics/molecular mechanics (QM/MM) methods. The force field parameters were derived for two b-type heme (heme bH and heme bL) cofactors in the cytochrome *b* protein and the [2FE-2S] cluster found in iron sulphur protein subunit. Validation through MD showed that the metal parameters held

the cofactors within the coordination sphere, proving the reliability of the derived force field parameters.

Chapter 5 is divided into three parts: PART I, II and III. The first part sought to understand resistance mechanisms by highlighting the effects of ATQ-linked point mutations. This was performed in the reference protein (holo) and ATQ-bound systems, with findings published as shown in the publication section. Part II identified eight selective compounds out of 1,024 (SANCDB and MMV) as potential inhibitors of the *P. falciparum* cytochrome *b* Qo active site. Through 350 ns MD simulations, these compounds remained bound in the active site in WT and mutant systems. Part III screened for the same set of compounds bound at the protein's heterodimer interface, where thirteen selective hit compounds were identified. Through MD and other *in silico* approaches, these compounds were evidently observed to have distal effects thus could be explored as potential allosteric modulators.

Finally, Chapter 6 summarises the key findings of this entire study and offers recommendations for future research.

CHAPTER 1

LITERATURE REVIEW

Malaria is a communicable parasitic disease caused by Plasmodium spp. and it is mainly transmitted through mosquito bites. In 2020, the World Health Organization (WHO) considered malaria a global health challenge accounting for 229 million infections and about 409,000 deaths worldwide. The majority (94%) of the global infections (215 million cases) were reported to come from the African continent. Other regions affected by the disease include but are not limited to Asia and South America continents. WHO also estimates that those at highest risk of infection are children under 5 years of age, who contribute to 67% of the global deaths, followed by immunocompromised individuals. Out of the several Plasmodium species, only five species are infectious to humans: Plasmodium falciparum, Plasmodium ovale, Plasmodium malariae, Plasmodium vivax and Plasmodium knowlesi. P. vivax and P. malariae are mainly found in America and Europe, while P. falciparum is most abundant in the African region. Thus, this chapter expounds on malaria as a life-threatening health problem, detailing its mode of transmission and strategies harnessed in its eradication. Furthermore, antimalarial drugs and drug targets will be reviewed in general, as well as recent advances in drug discovery.

1.1 Malaria: Scope of the disease burden

Malaria is a life-threatening parasitic disease accounting for 229 million infections and about 409,000 deaths globally in 2019 alone, making it one of the leading causes of mortality [1]. It is most common in African countries and some parts of Asia, while only reported as imported cases from endemic areas in some developed countries [2]. The majority of the infections and deaths occur in poor tropical countries, especially the sub-Saharan African region [1,3,4]. The World Health Organization (WHO) malaria report indicates that the African region contributed 94% of the global cases (215 million cases) in 2019, with six countries – Nigeria (27%), the Democratic Republic of the Congo (12%), Uganda (5%), Mozambique (4%) and Niger (3%) – responsible for 51% of all cases globally. Other regions like South-East Asia, Eastern Mediterranean, Western Pacific, and America only contributed 2.8%, 2.2%, 0.7% and 0.4% respectively, while the European region has remained free of malaria since 2015 (**Figure 1.1**). The most vulnerable population group is children under the age of 5 years, contributing up to 67% of the global deaths in 2019 [1]. According to Schumacher and Spinelli (2012) [5], children are overall the most vulnerable due to loss of maternal immunity and underdeveloped specific immunity to infection. Generally, children, pregnant women and immunocompromised individuals are most vulnerable to malaria infection [2]. In pregnant women, the disease causes maternal anaemia resulting in low birth weight for babies, and is a major cause of childhood anaemia [3].

The morbidity and mortality differences across the regions are partially due to the prevalence of the disease, as well as substantial differences in the geographical and climatic conditions that give rise to the disease [3] (**Figure 1.1**). The malaria vectors, *Anopheles* mosquitoes, are not found in extremely cold regions [3]. According to Sachs and Malaney (2002) [6], there is a decrease in transmission rate when temperatures fall below 18°C and a total halt in parasite

development at temperatures below 16°C, indicating that seasonal temperature variations are the main factor in interpreting the disease geographical distribution. According to Gollin and Zimmermann [3], the extensive malaria burden in the sub-Saharan region is also due to the prevalence of two highly anthropophilic *Anopheles* species, *An. gambiae* and *An. funestus*, which favour human hosts over other animals.

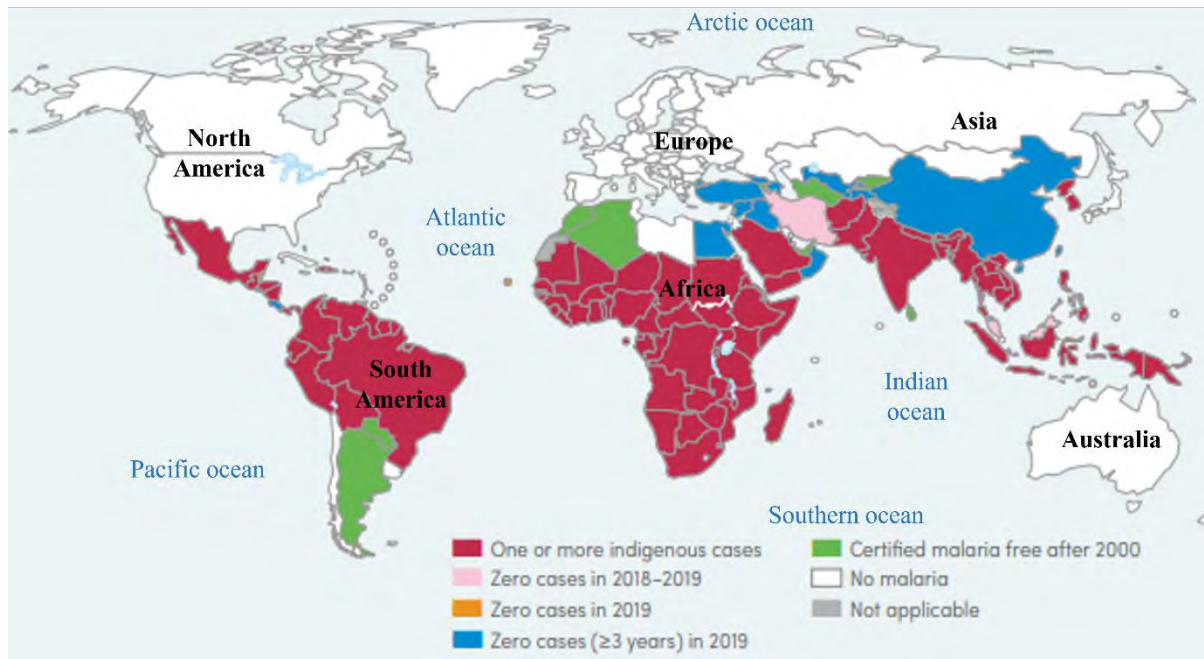


Figure 1.1: Global estimate of malaria indigenous prevalence/incidence cases. Map adapted from [1].

The Global Technical Strategy for Malaria 2016 - 2030 (GTS) set a goal to decrease the malaria case incidence and mortality rate by at least 40% by 2020, 75% by 2025 and 90% by 2030 from the 2015 baseline [1]. In comparison to the projected trends, as of 2020, the expected incidence cases and mortality rates were off track by 37% and 22%, respectively. WHO (2020) reported that failure to meet the expected 2020 target was because the designed projections did not account for potential disruptions, including the current severe acute respiratory syndrome coronavirus 2 (SARS-CoV2), causing the COVID-19 global pandemic. WHO (2020) also

reported that by April 2020, all malaria-endemic countries had been affected by COVID-19 which resulted in 22 million cases and 600,000 deaths by November 2020. To mitigate the spread of SARS-CoV2, the public was advised to stay home if they displayed any signs of a cold or fever. This strategy has disrupted patients from seeking treatment for other febrile diseases like malaria resulting in reduced access to effective antimalarial drugs and higher mortality rates despite all prevention campaigns being completed [1,7–9].

1.2 *Plasmodium* parasites

Malaria is caused by small protozoan amoeboid intracellular *Plasmodium* species that occupy the insoluble haemoglobin metabolite. The genus *Plasmodium* consists of several (~172) species, out of which only 5 infect humans [2]. These species include: *Plasmodium falciparum* (*P. falciparum*), *Plasmodium malariae* (*P. malariae*), *Plasmodium vivax* (*P. vivax*), *Plasmodium ovale* (*P. ovale*), and *Plasmodium knowlesi* (*P. knowlesi*) [2,10]. *P. knowlesi* has been recorded in South-East Asia [11], while in America and Europe, the most common species are *P. vivax* and *P. malariae*. *P. vivax* is mainly responsible for malarial relapse, impeding malaria treatment and eradication efforts [4]. *P. malariae* on the other hand causes a mild form of malaria although some of its early symptoms are similar to those observed in *P. falciparum* and *P. vivax*. *P. ovale* infecting humans, is the last to be described having insufficient information on its prevalence. Certainly, out of all the five species, *P. falciparum* is the most virulent [4]. It is abundantly found in Africa [2] and is the major cause of ill health, neuro-disability (cerebral malaria) and death in the tropical countries [12]. Thus, the *P. falciparum*. contributes significantly to the malaria disease burden and communicable diseases in general.

1.3 *Plasmodium* parasite transmission and life cycle

The *Plasmodium* parasite is transmitted through the malaria vectors, female *Anopheles* mosquitoes, to a human host during a blood meal [2,3]. In general, the *Plasmodium* life cycle involves two phases; the asexual phase in the vertebrate host and the sexual phase in the mosquito vector (**Figure 1.2**) [2,13]. The two phases are further categorised into exo-erythrocytic, erythrocytic and sporogonic cycles.

During the exo-erythrocytic stage, a person is infected with malaria when a mosquito carrying the sporozoite stage of the plasmodial parasite in its salivary glands injects this into the human host's bloodstream during a blood meal [3]. The sporozoites infect the liver cells and mature into merozoites which are then released into the bloodstream to invade the red blood cells [14], eventually leading to their rupture. The ruptured red blood cells release merozoites into the bloodstream to initiate another replication cycle.

The infective erythrocytic or blood stage, is initiated when a subset of the merozoites mature and develop into gametocytes, the sexual stages of the parasite. The gametocytes are ingested by mosquitoes biting the human host, after which they mate in the mosquito gut, beginning a multiplication cycle within the mosquito [3,14]. The blood stage parasites cause malaria symptoms. Typically, an infected person exhibits symptoms such as fever, chills, headache, muscle pains and fatigue. Moreover, symptoms of severe malaria include: anaemia, repeated seizures, coma, hypoglycaemia, metabolic acidosis or multiple organ failure [12].

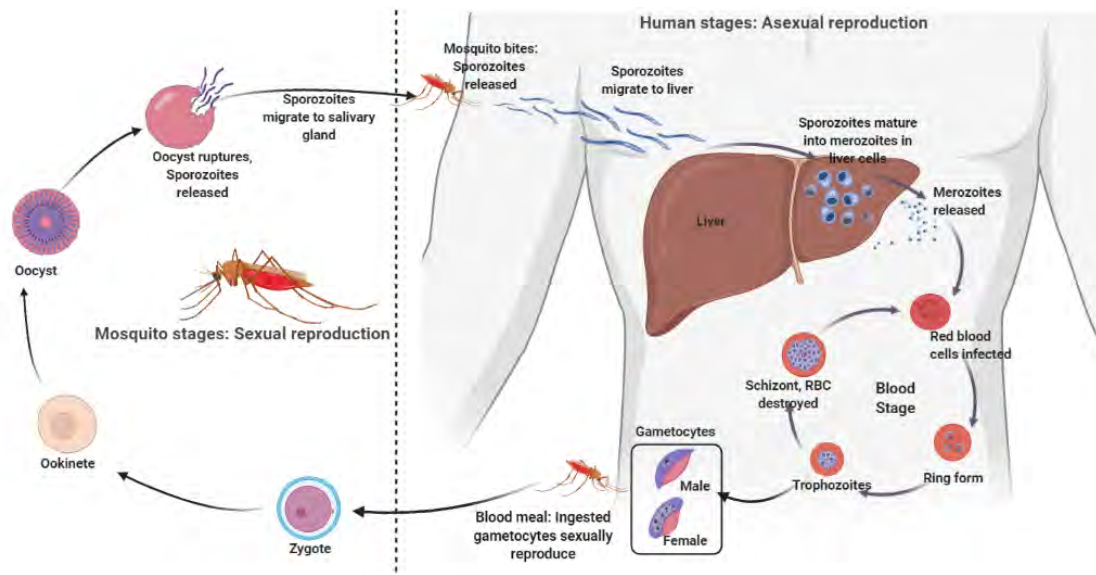


Figure 1.2: Life cycle of *Plasmodium* in the host and vector stages. Asexual reproduction occurs in the human body (host), mainly in the liver and red blood cells, while sexual reproduction occurs in the mosquito vector. Generated in BioRender, Adapted from [2,3,13].

The third and final stage is the sporogonic cycle. Here, the parasites (microgametes) ingested into the mosquito's stomach penetrate the macrogametes. This generates zygotes which develop into ookinetes and then oocysts after penetrating the mosquito's midgut wall. The oocysts then grow into sporozoites which move to the mosquitos' salivary glands, where the cycle repeats again when a mosquito takes a blood meal.

1.4 Malaria disease management

1.4.1 Diagnosis of malaria

Diagnosis of malaria can be performed using direct methods such as microscopic analysis, rapid diagnostic tests (RDT) and molecular tests [2]. RDTs for antigen detection in the blood have been approved as the first screening method across all malaria-endemic regions because they are fast and simple to perform. Nevertheless, the technique is less sensitive and accurate;

hence RDT test is recommended after which the results confirmed by either microscopy or Polymerase Chain Reaction (PCR) [15,16]. Specifically, the light microscopy technique of staining blood films with Giemsa stain is the standard method for malaria detection; however, it is not readily available in the sub-Saharan African region due to the lack of trained technicians and proper staining dye [17,18]. PCR is the most sensitive and specific conventional method which detects parasitic nucleic acids. However, in endemic areas, the technique is time-consuming thus, it is mainly used in identifying *Plasmodium* species after diagnosis has been performed using other techniques [2,19]. A simpler molecular technique, loop-mediated isothermal amplification (LAMP), can also be used in endemic regions [19]. There are some indirect methods that could also be used for malaria diagnosis, including the detection of antibodies in potentially affected individuals, such as Indirect immunofluorescence assay (IFA) and Enzyme-Linked Immunosorbent Assays (ELISA) [2]. Despite the high specificity and sensitivity of assays like IFA, they are not suitable for large sample sizes and interpretation of results is difficult.

1.4.2 Prevention of malaria

According to WHO (2020), there is only one vaccine, RTS,S/AS01 [20], that has demonstrated significant effectiveness against *P. falciparum* uncomplicated malaria (39%) and severe malaria (29%) in young children. The vaccine is currently being slowly administered in Ghana, Kenya and Malawi [4]. However, in terms of treatment and eradication, its efficacy is low and WHO did not recommend it for inclusion in the expanded programme of immunisations (EPI) [21,22].

Chemoprophylactic drugs like atovaquone (ATQ)/proguanil, chloroquine, doxycycline, mefloquine, primaquine and tafenoquine are also used for disease prevention (<https://www.cdc.gov/malaria/travelers/drugs.html>; accessed 19 August 2021). Other key

strategic preventative measures include the use of indoor residual spraying (IRS) and long-lasting insecticide-treated bed nets (LLINs) as well as controlling mosquito breeding sites by draining stagnant water bodies [23,24]. However, these interventions have faced a variety of setbacks, including vector insecticide resistance [25]. There have been changes in the vector biting behaviour, with mosquitos biting in the evening hours when individuals are not protected [26,27]. Furthermore, improper use, sub-standard quality and low supply of bed nets decreases the efficiency of malaria prevention [28,29].

1.4.3 Treatment of malaria

Generally, there are three broad categories of antimalarial drugs: i) Aryl amino alcohol compounds such as quinine, lumefantrine, chloroquine, amodiaquine, mefloquine and cycloquine; ii) Antifolate compounds like proguanil, pyrimethamine and trimethoprim; and iii) Artemisinin compounds like artemisinin, dihydroartemisinin, artesunate and artemether (**Table 1.1**) [30,31]. To date, the most effective antimalarial drugs derived from medicinal plants are quinine and artemisinin [32,33]. Besides the fact that natural products are often perceived to be safe, they present unique scaffolds that are very efficient in killing the parasite [34]. These natural products are advantageous due to their innate affinity for biological receptors and their easy accessibility across the globe. Thus, they can be used as a baseline to guide the development of novel antimalarials [35]. The use of natural products resulted in the synthesis of semi-synthetic artemisinin derivatives to counteract poor biopharmaceutical properties of artemisinin. However, the drugs' therapeutic effects declined due to parasite recrudescence, artemisinin's short half-life and complex 3-day regimens. This has paved the way for the synthesis of full synthetic drugs such as OZ277 and OZ439 to overcome artemisinin resistance [4]. Currently, the best available treatment for *P. falciparum* is artemisinin-based combination therapy (ACT) [1]. The most common ACTs are Artemether–lumefantrine, DHA-piperaquine,

artesunate-amodiaquine, artesunate-mefloquine and artesunate-sulphadoxine-pyrimethamine [36]. In the African region, the first-line treatments for *P. falciparum* malaria include artemether-lumefantrine (AL), artesunate-amodiaquine (AS-AQ) and dihydroartemisinin-piperaquine (DHA-PPQ) with consistent overall average efficacy rates of 98.0%, 98.4% and 99.4%, respectively [1].

Table 1.1: Classification of antimalarial drugs [30,31,36,37]

Category	Antimalarial Drug
Cinchaona alkaloid	Quinine and quinidine
4-aminoquinolines	Chloroquine, piperaquine, amodiaquine
Antifolates	Cycloguanil, sulfadoxine, pyrimethamine
Naphthoquinone	Atovaquone
Antibiotics	Doxycycline, clindamycin
Artemisinin	Artesunate + amodiaquine, Artemether + lumefantrine, Artesunate + mefloquine, Dihydroartemisinin + piperaquine
Non-artemisinins	Sulfadoxine + pyrimethamine, Quinine + doxycycline, Sulfadoxine + pyrimethamine + amodiaquine

1.5 Antimalarial drug targets

Most antimalarial drugs target the asexual erythrocytic stages (blood schizonticidal drugs) either as fast-acting (chloroquine, quinine) or slow-acting (pyrimethamine, sulphonamides) drugs [30]. Other antimalarial drugs target the liver hypnozoites (tissue schizonticidal drugs), while others destroy the sexual erythrocytic forms in the bloodstream (gametocytocidal drugs) or inhibit the formation of malarial oocysts and sporozoites (sporontocides) [30].

According to Yang *et al.*, 2021, the targeted sites are classified into four validated target (VT) groups; genetically targeted sites (VT1), chemically targeted sites (VT2), a combination of

VT1 and VT2 with a known resistance (VT3) and clinically targeted sites (VT4). Targeted sites from the VT1 group demonstrate that the gene encoding for that protein is essential for parasite survival, while the VT2 group identifies druggable sites. In VT4 sites, compounds acting on these sites are involved in eliminating malarial symptoms in humans. Examples of these validated target sites are Cytochrome *b1* and Phosphatidylinositol 4-kinase, and they are summarised in **Table 1.2**.

Table 1.2: Summary of validated target (VT) sites (VT3 and VT4) and their inhibitors [4]

Validated Target Class	Drug Target	Drug Inhibitors
VT3	Cytoplasmic prolyl-tRNA synthetase	Halofuginone
	Isoleucyl-tRNA synthetase	Thiaisoleucine
	Cytosolic lysyl-tRNA synthetase	Cladosporin
VT4	Cytochrome <i>b1</i>	Atovaquone; decoquinate; tetracyclic benzothiazepine
	Dihydroorotate dehydrogenase	
	Dihydrofolate reductase-thymidylate synthetase	Pyrimethamine; cycloguanil

From recent studies, the critical role played by mitochondria in cellular physiology identifies the associated organelles as potential targets for antiparasitic drugs [38]. Shibeshi *et al.*, [30] identified the most targeted biological pathway to be heme detoxification in the digestive vacuole. Other targeted pathways include biosynthesis of pyrimidine and folate and electron transport occurring in the mitochondria. *Plasmodium*'s liver and blood stages depend on the electron transport chain (ETC) found in its mitochondria. Since these stages are crucial for the parasite's survival, the enzymes involved in the ETC and the *de novo* pyrimidine biosynthesis pathways are desirable target sites for antimalarial drugs. Drugs such as decoquinate and ATQ target the ubiquinol-binding pocket of *P. falciparum* cytochrome *b* (PfCytb) [39]. In this study,

novel inhibitors against *P. falciparum* cytochrome *bc₁* complex will be investigated to identify those compounds that can withstand ATQ associated drug resistance mutations.

1.6 Antimalarial drug resistance

Despite the effective use of LLINs, IRS, prophylactic and ACT drug therapies in relieving the malaria burden, an increase in first-line antimalarial drug resistance and incidence rates have been reported [4,40]. Therefore, the development of novel control measures and medicines is necessary. According to previous studies [14,30], the high burden of malarial parasites in an infected person's blood and their genome's mutability contribute to malarial parasite resistance. On average, an infected human has a $10^3 - 10^9$ parasite load in the blood with approximately $1.0 - 0.7 \times 10^{-9}$ mutations per base per generation [41,42]. Therefore, there is high probability of a random survival mutation occurring in the presence of drug pressure during the replication phases that will lead to antimalarial resistance [14,30]. Some factors that facilitate the emergence of antimalarial drug resistance include poor treatment compliance, drug strength, improper dosing and poor adherence to malaria treatment [30]. The recent discovery in Madagascar of a tonic of *Artemisia annua* as a potential cure for SARS-CoV-2 will potentially increase the development of ACT resistance as the plant is being widely adopted to avert COVID-19-related deaths [8,43]. Additionally, the wide use of chloroquine and hydroxychloroquine to manage SARS-CoV-2 will potentially result in more resistant malaria parasites [44].

Cowell and Winzeler (2019) investigated the causes of resistance across some major antimalarial drugs like 4-aminoquinolines and aryl-amino alcohols, atovaquone and antifolate drugs [14]. Their findings concluded that 4-aminoquinolines and aryl-amino alcohols resistance is caused by mutations in the genes encoding transporters at the parasite's digestive

vacuole (DV) membrane. On the other hand, ATQ and antifolate drugs resistance is due to point mutations in enzyme encoding genes, resulting in a decrease in drug binding affinity. The changes in structure, function, or quantity of the targeted protein result in drug resistance. These changes are due to genetic alterations such as single nucleotide polymorphisms (SNP) or gene amplification. Due to this, resistance to most ACT drugs (amodiaquine, lumefantrine, mefloquine and sulphadoxine-pyrimethamine) has been reported [8].

1.7 Atovaquone

Atovaquone (ATQ) {2-[trans-4-(4'-chlorophenyl)cyclohexyl]-3-hydroxy-1,4-naphthoquinone}, a hydroxynaphthoquinone, is a broad spectrum antimicrobial [45] with activity against a variety of parasitic infections of *Plasmodium* spp., *P. carinii*, *Babesia* spp., and *Toxoplasma gondii* [46,47]. ATQ is known to be effective for the treatment and prevention of *Pneumocystis carinii pneumonia* (PCP), treatment of babesiosis as well as malaria treatment and prevention. It is administered in combination with the proguanil (1-(p-chlorophenyl)-5-isopropylbiguanide); a potent slow-acting antimalarial drug [47].

ATQ is an analogue of ubiquinone that acts as a selective inhibitor of *P. falciparum* cytochrome *bc₁* complex and is the leading antimalarial drug used in treatment and causal prophylaxis against uncomplicated *P. falciparum* malaria [38,39,48–50]. WHO approved it for use as an alternative treatment in first-line ACT resistant regions or in areas where ACT drugs are unavailable. It is also used as a preventative drug by tourists travelling to malaria endemic countries [38,51]. ATQ is the synthetic form of lapichol natural compound [34].

Previous studies have shown that administration and use of proguanil as a sole agent caused the parasites to be resistant to the drug more quickly, tolerating up to a thousand fold higher

ATQ concentrations [46,50,52]. The ATQ-proguanil (AP) drug combination (Malarone) has a synergistic effect that has been reported to be due to their biguanide mode of action, affecting cellular physiology [46,52]. According to Staines *et al.*, 2018 [38], proguanil alone does not directly affect the parasite; however, it enhances the ability of ATQ to collapse the parasite membrane potential by sensitising ATQ to the mitochondria. Clinical trials with the AP combination gave close to 100% cure rates [46], and it has been reported by Staines *et al.* in 2018 [38] to have a treatment efficacy of 89% - 98% for *P. falciparum* uncomplicated malaria.

According to Talapko *et al.* (2019) [2], the drug is used as an option for individuals that have failed therapy with the ordinary ACTs. However, it is not fully implemented for use in endemic countries due to its rapid development of resistance. The use of the drug concurrently with other drugs like antibiotics (rifampicin) or anticoagulants (warfarin) interferes with their activity, resulting in reduced ATQ plasma concentration and increased anticoagulant effects, respectively [47,52]. Some general side effects of the drug include maculopapular rash, nausea, diarrhoea, headache and elevated liver enzymes [47].

1.7.1 Mechanism of ATQ drug action

The inhibition of cytochrome *bc₁* is mainly achieved by ubiquinone antagonists that interfere with the electron transfer (ET) of ubiquinol oxidation or ubiquinone reduction steps [53]. ATQ inhibits the ETC via the Cyt *bc₁* complex where it collapses the mitochondrial membrane potential [38,54]. It is selectively toxic to the parasite's Cyt *bc₁* without affecting the host mitochondrial functions. This has been reported to be due to a discrete cavity on the Cyt *bc₁* complex that consists of subtle variations in hydrophobicity and amino acid side chain volumes [54]. It is a potent inhibitor of both the proliferating erythrocytic and sexual stages of *P. falciparum* parasites, as well as the liver stages [50,53].

The key features that contribute to the antiparasitic activity of ATQ drug are the presence of the cyclohexyl-chlorophenyl side-chain and the hydroxyl group on carbon-2 and carbon-3 positions, respectively [54]. Also, only the *trans* configuration on the side chain is active against the parasite's *bc₁* complex compared to the inactive *cis* configuration. ATQ's mechanism of action is based on its structural similarity to ubiquinone (also known as coenzyme Q), the mitochondrial protein that is an aerobic respiration electron flow component [49]. Coenzyme Q (coQ) passes electrons to cytochromes from the dehydrogenase enzymes. ATQ inhibits the electron passage from coQ to Cyt *bc₁* (complex III) which requires the binding of coQ complex III at the Cyt Qo domain [31,47,49]. The defined structure of the Qo binding site permits the drug's selective toxicity against parasitic mitochondria. This inhibition results in the collapse of the mitochondrial membrane potential. Previous studies report that the ATQ antiparasitic activity is also attributed to the activity of proguanil and its metabolite, cycloguanil, which inhibits folate synthesis and parasite replication [38]. Upon the binding of ATQ to the Qo site, the ISP domain movement for ET is blocked [55].

1.7.2 ATQ resistance and mutations

The primary goal of the AP treatment is to act as a causal liver-stage prophylaxis drug while also being effective against multidrug-resistant (MDR) malaria [56]. However, this treatment has shown rapid in-host clinical resistance because of mutation in the Cyt *bc₁* complex, although other mechanisms for resistance may be involved [57,58]. Mutation of the Cyt *b* gene has been reported to occur in atovaquone-resistant isolates of *T. gondii*, *Plasmodium* spp. and *Pneumocystis* spp. [47]. A study by Di Rago *et al.*, 1989 [59], reported that resistance of a variety of microorganisms, including *Saccharomyces cerevisiae*, to the Cyt *b* potential inhibitors was due to single point mutations in the Qo substrate binding site. Their study

showed that inhibitor-resistant mutants were due to mutations found on the Cyt *b* gene consisting of four exons (exons 1, 4, 5 and 6).

Generally, resistance to the ATQ drug in *P. falciparum* strains is associated with mutations called single nucleotide polymorphisms (SNPs) in the Qo and Qi sites in the Cyt *b* gene [53]. According to Baggish and Hill (2002) [47], the high rates of mutations in the Cyt *b* gene may be due to the gene being in the mitochondrial genome which is subjected to less efficient proofreading compared to the genes in the nuclear genome. In the *P. yoelii* species, a study performed on mice showed that ATQ resistance was attributed to treatment with suboptimal doses of the drug [50]. However, the authors still reported that all resistant lines were because of single or double amino acid mutations close to the Qo site in the Cyt *b* protein sequence. A study by Korsinczky *et al.*, 2020 identified point mutations in *P. falciparum* Cyt *b* that are associated with ATQ resistance. The mutations reduced parasite ATQ susceptibility from 25 to 9,354-fold, with most of the causative mutations clustered around the ATQ binding site. According to previous studies, the spontaneous resistance is mostly attributed to the Y268S mutation found in the Cyt *b* [53]. Staines (2020) also showed that other point mutations reduced ATQ activity in the same protein residue (268, *P. falciparum* numbering) of Cyt *b*, particularly Y268S/C/N, both *in vivo* and *in vitro*.

Studies on these mutations have shown that there is loss of residue contact and change in the active site pocket, resulting in ATQ resistance [58,60–65]. Failure of previous *in silico* and *in vivo* studies to provide a full account of ATQ resistance and offer new effective antimalarial drugs implies a need to identify cytochrome *bc₁* inhibitors with limited or no cross-resistance to the ATQ drug [39].

1.7.3 Review on current ATQ alternatives

One of the most effective antimalarial drug combination therapies is the use of ATQ and proguanil. However, the rapid development of ATQ resistance in both mono and combination therapies has rendered them less effective, hence the need to develop new drugs that are active against ATQ-resistant parasites [53]. According to Styka *et al.* [66], reducing AP treatment failure rates may be accomplished by including artesunate to the combination therapy.

One of the drugs that has shown effectiveness against ATQ-resistant parasites is decoquinate, which is a 4-hydroxyquinoline coccidiostat that interferes with the development of sporozoites by inhibiting mitochondrial electron transport [67]. A study by Nam *et al.* [39] showed that the drug decoquinate's primary mode of action is to inhibit the cytochrome *bc₁* complex. Their findings showed that there was limited cross-resistance across most of the mutations (M133I/V140T/L181V) which resulted in a huge potency reduction for ATQ compared to decoquinate. Mutations that caused decoquinate resistance were not found in atovaquone resistant (ATQ-R) lines, indicating that the two drugs have different and highly specific interactions within the ubiquinol-binding site of *P. falciparum* Cyt *b* [39]. On the other hand, acridinediones and quinolone derivatives have also been identified as *P. falciparum* cytochrome *bc₁* inhibitors [53,68]. Studies on these derivatives revealed that haloalkyl- and haloalkoxy quinolones have higher clinical antimalarial activity compared to some 4-amino-quinolines and antifolate drugs. However, findings also showed little to significant cross-resistance in ATQ-R isolates based on the chemical structure of the derivative compound [68].

A study by Dong *et al.* [53] discovered that tetracyclic benzothiazepine (BTZs) compounds also target *P. falciparum* Cyt *bc₁* complex and reported these compounds as a potent novel class of antimalarials. They reported that BTZ-based inhibitors retained activity against AQT-

R parasites, and they were more active against *P. falciparum* Cyt *bc₁* compared to *Hs* Cyt *bc₁*, a selective advantage appropriate for drug discovery. Thus, this class may be further studied and eventually included in malarial therapy as an alternative to atovaquone [53]. Inhibitors like stigmatellin and 5-n-undecyl-4,7-dioxobenzoxythiazol (UHDBT) are structurally similar to ATQ and have been reported to bind to the Qo Cyt *b* site and also inhibit the ETC [59]. Despite the identified ATQ alternative antimalarials, the rapid development of parasite resistance in most of these alternatives implies that there is an urgent need for novel ETC inhibitors with high distinct inhibitory effects that may withstand all the reported single point mutations [53].

1.8 Problem statement and justification

Malaria is a vector-borne disease caused by the *Plasmodium* species and remains one of the leading causes of mortality around the world [1]. Out of the five *Plasmodium* spp. that infect humans, *P. falciparum* is the most virulent cause of infections in the sub-Saharan African region. This greatly impacts the health and economic development of malaria-endemic countries within the region. Numerous strategies have been employed in malaria disease eradication, including indoor residual spraying, long-lasting insecticidal nets (LLINs), chemotherapeutic drugs and recently the malaria vaccine. These strategies have faced myriad challenges including non-compliance in drug dosage, insecticide resistance and drug resistance (for artemisinin combination therapy (ACT) and many other drugs) [63,69–71]. Also, the efficacy of the approved vaccine RTS,S/AS01 has unfortunately been reported to be low and was not recommended for inclusion in the expanded programme of immunisations (EPI) [21].

Specific to this work, recent studies investigating ATQ resistance mechanisms have shown that point mutations (Y268C, Y268N and Y268S) of *P. falciparum* Cyt *b* confer ATQ resistance.

The Global Technical Strategy for Malaria 2016-2030 has set a goal to achieve a 90% decrease in malaria mortality and incidence rate by 2030. However, WHO reports that the expected progress for 2020 was behind by 22-37%, and this was attributed to antimalarial drug resistance and the current disruptions caused by the COVID-19 pandemic. Resistance across all antimalarial drugs, including the first-line antimalarial drugs, has been reported, including but not limited to ATQ. ATQ, the leading *P. falciparum* antimalarial drug used in treatment and causal prophylaxis, is a selective parasite cytochrome *bc₁* complex inhibitor, administered in combination with proguanil drug [2]. However, the rapid emergence and dissemination of ATQ resistant parasites, mainly caused by Cyt *b* gene mutations, has reduced the efficacy of malarial treatment and compromised efforts to eradicate malaria disease [2,39,53].

The specific target on *P. falciparum* cytochrome *bc₁* complex for the ATQ drug is not only useful for malaria treatment but also as chemoprophylaxis, important in targeting travellers who are contributors to disease spread. Therefore, there is a need for research aimed at finding and testing novel antimalarial compounds using new strategies and principles for drug targets and discovery. The advancement of technology supports *in silico* studies to achieve quick structure-based screening of a variety of natural and synthetic drug compounds. Therefore, this study will screen and identify potential compounds for the *P. falciparum* cytochrome *b* drug target. We focused on *P. falciparum* only as it is the most virulent and is the parasite on which ATQ-linked mutations are implicated. Worthy to note, there is no computational study which has performed experiments incorporating the protein into the phospholipid bilayer. As such, this study presents reliable findings in investigating the resistance mechanisms of all three mutations and performing structure-based virtual screening with an aim to identify novel compounds with no cross-resistance to ATQ resistant parasites.

1.9 Hypothesis

The *P. falciparum* cytochrome *bc₁* complex is different from its human homolog and can be targeted to identify potential selective hit compounds derived from SANCDB database and MMV website.

1.10 Research aim and objectives

1.10.1 Aim

In contribution to drug discovery, the study seeks to identify new alternatives to atovaquone, a prophylactic drug that has already faced resistance. Therefore, the aim of this study was to first identify unique feature existing in *P. falciparum* comparing to human homolog described in chapter 3, incorporate force field parameters critical in setting up molecular dynamics in chapter 4. Also, the study sought to understand the resistance mechanisms to PfCytb-ISP protein associated to Y268C/N/S mutations. This study aims to explore natural compounds from the South African Natural Compounds Database (SANCDB) [72,73] and approved synthetic compounds from the Medicines for Malaria Venture (MMV) database to identify and repurpose potential antimalarial hit compounds with potential inhibitory activity against Qo active site *P. falciparum* cytochrome *bc₁* complex as outlined in chapter 5. Additionally, virtual screening was performed targeting PfCytb-ISP heterodimer interface to determine any possible allosteric modulation as outline in chapter 5.

1.10.2 Objectives

Using *in silico* approaches, the following specific objectives served as guidelines to achieving the aim of the study:

- 1) To further elucidate the techniques and the methodology utilized in this study including molecular similarity techniques (sequence alignment, phylogenetics tree calculations, protein modelling and mapping), molecular docking and dynamic simulations (Chapter 2).
- 2) Perform sequence and structural comparison of Plasmodial cytochrome *b* with other homologs (a total of 74 protein sequences), including the human homolog, to identify unique sequence features in the Plasmodium protein that can be explored for selective inhibitory drug design. In the absence of the *P. falciparum* and *H. sapiens* cytochrome *b* crystal structures, protein modelling of their 3D structures was performed (Chapter 3).
- 3) To derive force field parameters that can define the Plasmodial cytochrome *b* quinol oxidation metal centre and thereafter validate these through membrane molecular dynamics (MD) simulations. The developed parameters are key to describing the cofactors during the MD simulations (Chapter 4).
- 4) Use SANCDB and MMV drug compounds to identify Qo active site and *Plasmodium* interface novel compounds and analyse the effect of mutations on protein and ligand-bound systems (Chapter 5):
 - a) Understand the parasite's resistance mechanisms (Y268C/N/S) in holo and ATQ-bound systems (Chapter 5: PART I).
 - b) Identify potential inhibitors against *P. falciparum* cytochrome *bc₁* complex Qo active site and evaluate the stability of all the hit compounds using membrane

MD simulations. Further, the effects of mutations on the stability of the hit compounds and protein function were assessed (Chapter 5: PART II).

- c) Explore *P. falciparum* cytochrome *b* - iron sulphur protein (*PfCytb-ISP*) interface pocket as a potential allosteric site to identify selective compounds and evaluate their stability in wildtype and mutant states using MD simulations (Chapter 5: PART III).

CHAPTER 2

IN SILICO METHODS AND TOOLS FOR DRUG DISCOVERY

In drug discovery process, in silico approaches have been very critical in identifying hit compounds, in this case, compounds that selectively target the Plasmodium parasite. Thus, this chapter discusses computational approaches and tools that have since been employed in structure-based drug designs.

2.1 Introduction to drug discovery

Drug discovery is a process of identifying natural or synthetic chemical compounds with the potential of becoming therapeutic agents in the treatment of diseases. It involves discovering drug targets and validating, identifying and optimising lead compounds, followed by preclinical and phase 1 to 4 clinical studies as well as post approval monitoring [74]. The process is time consuming, costly, risky and requires research and technological expertise [75]. Alternative *in silico* approaches to identifying biochemical compounds are also being explored such as computer-aided drug discovery (CADD) approaches [76]. Specifically, the structure-based approach involves docking compounds on predicted target sites, essential for drug discovery [39]. In cell-based assays, the target-based drug discovery approach screens biochemical compounds on chemically validated target sites [4,39]. The use of structure (target)-based approaches enables visualisation of compound-target interactions (structure-activity relationship) leading to structure-based and fragment-based drug discovery approaches [4]. This approach is successful even in the absence of a structure, through counter screens against the human homolog, and can also lead to DNA-encoded library (DEL) technology methods. However, the lack of well-validated target sites impedes drug discovery using this approach. Therefore, discovering and validating of new target sites is crucial to further target-based drug discovery [4].

2.1.1 Antimalarial drug history and discovery

The treatment of malaria dates to the 2nd century BC where a sweet sagewort plant named Qinghai (*Artemisia annua*) was initially used in China. Further discovery of the quinine ingredient from the Cinchona tree (*Cinchona succirubra*) made it the first successful chemical in malaria treatment. Quinine was actively used in mainstream chemoprophylaxis and

treatment of malaria from the 1600s until the 1920s, when more efficient synthetic drugs like chloroquine were developed [2,77]. Rapid and frequent use of chloroquine resulted in *Plasmodium falciparum* becoming resistant to the drug across Southeast Asia and South America, and eventually across all regions by 1980. The increased chloroquine resistance left room for quinine to be used to treat severe malaria infections in second-line drug combinations and during the first trimester in pregnant women regardless of its adverse effects [77].

Screening for antimalarial drugs has mainly been done using traditional approaches [39] like phenotypic screening, which involves compound screening against blood and liver stage *Plasmodium* parasites by *in vitro* evolution, whole-genome analysis (IVIEWGA) chemoproteomics and chemoinformatics approaches [4].

Even though several drugs (**Table 1.1**) have been identified and developed for the treatment of malaria, most of these drugs may be excluded due to low *in vivo* efficacy, toxicity issues and high parasitic resistance. Therefore, there is a need for the identification and development of drugs that not only target the pathogenic asexual stage but those that are also involved in blood-stage treatment, hypnozoite anti-relapse, liver prophylaxis, blocking gametocyte transmission, and are mosquitocidal. Despite the identification of numerous effective drugs, the phenotypic approach is dependent on the presence of biochemical and cellular target assays to provide mechanism of action [4].

Also, although several essential proteins for *Plasmodium* survival have been identified, the potential target site needs to have a binding site for small molecules that can inhibit important parasite developmental stages and eventually lead to parasite death [4]. *In silico* studies that identify potential antimalarial drugs against essential parasite enzymes have been done [78–80]. However, there is a need to continuously develop novel antimalarials targeting other druggable sites to circumvent the growing ACT drug resistance [81]. In drug discovery, the

current time and cost-effective approach with little risks is drug repurposing. This approach uses existing approved or investigational drugs to treat other diseases and has been successful in repurposing drugs like sildenafil, minoxidil, efloornithine and zidovudine [81].

2.2 Drug discovery methods; molecular similarity and mapping techniques

2.2.1 Multiple sequence alignment

The multiple sequence alignment (MSA) is a technique that compares nucleotide or protein sequences by aligning the sequences to elucidate their molecular similarity. These methods include a tree-based progressive and global alignment method which are incorporated in most tools including T-Coffee [82], PROMALS3D [83] and ClustalW [84]. Additionally, there are tools such as MUSCLE [85], MAFFT [86] and Clustal Omega [87] that use iterative strategies to ensure convergence following realignments and tree re-estimation. In the study of protein sequences, amino acid substitution matrices such as PAM and BLOSUM [88] have been developed to determine sequence conservation and evolutionary events by introducing gap penalties. Due to the high computational cost demanded by other tools, the progressive alignment method has been advantageous in overcoming this challenge [89].

MSA can be performed using tools including: stand-alone MUSCLE, web servers MAFFT's FFT-NS-I, a progressive method employing iterative refinement approach, and PROMALS3D for comparison [83,85,86]. MAFFT utilises two algorithms (FFT-NS-1 and FFT-NS-2) and iterative refinement method (FFT-NS-i, NW-NS-i). The progressive methods first create a sketch distance matrix and build a guide tree, after which the sequences are aligned based on branching order [86]. To improve accuracy, a second progressive alignment is performed. The iterative refinement method is used to further improve the accuracy of the latter method. Here, the initial alignment is refined by subjecting the sequences to repetitive cycles of up to 1000

[90–92], performing both global and local alignment. By default, the BLOSUM62 matrix is adopted because of its higher accuracy values. Gap open and extension penalty parameters of 1.53 and 0.123 were used, respectively. For PROMALS3D, progressive methods are utilised to incorporate structural information during the alignment process. Here, similar multiple sequences were aligned together using BLOSUM62 scoring function. During the alignment process, an identity threshold of 0.6 (default value indicating sequences having an identity > 60%) was utilised. Identity threshold is a parameter that uses % sequence identity of the dataset in predicting quality and accurate alignments. Thereafter, second stage alignment was performed by subjecting the sequences to PSI-BLAST to search for homologs and PSIPRED to perform secondary structure prediction. A rigorous alignment having predicted secondary structures is performed where the first stage alignments are then merged to form final alignment of entire dataset [83].

Both these programs utilise the local alignment approach to identify conserved blocks of functionally relevant regions in the protein, e.g., conserved sequence motif. The structural alignments from these tools were compared to determine the alignment accuracy. Visualisation of the alignments was performed using Jalview software [93].

2.2.2 Phylogenetic tree calculations

To study the evolutionary trend of organisms/species, phylogenetic trees are calculated to predict relatedness of organism. From aligned sequences, phylogenetic trees are constructed using various methods including neighbor-joining, maximum likelihood (ML) and maximum parsimony. In this study, protein phylogenetic trees were created using RAXML [94] program which utilised the ML tree-building algorithm. For these phylogenetic tree calculations, amino acid models, incorporated in RaxML program, were used to determine evolutionary changes in multiple aligned proteins. These models included MTREV [95], JTT [96], RTREV [97], Le

and Gascuel (LG + G + I) [98] and WAG [99] that were used in calculations. Other parameters include (-f a) option which utilised rapid bootstrap algorithm to search for ML trees and -p option to permit reproducibility of runs (-p option). Gap deletions/missing data found in MSA, herein referred to as ‘gappy phylogenomic alignments’ (RaxML term), are factored in automatically during phylogeny construction using subtree equality vectors. The program considers gap deletions/missing data in its dataset where the amount of missing data ranges typically from 50 % to 90 %. As such, it screens, categorises, and represents alignments with gaps by 1 or non-gap sites by 0. That way, all gap sites are determined and factored in by calculating ancestral probabilities and thereafter calculate the branches for tree phylogeny [100]. Using 1000 bootstrap replicates in each calculated tree, the overall best phylogeny was selected from a model having strong branching based on bootstrap values. Numbers on tree branches represent posterior probabilities as percentages. Bayesian posterior probabilities have since been used to determine the phylogeny accuracy. Therefore, trees with bootstrap values cut-off > 70 % were considered 95 % accurate and reliable [101,102]. Visualisation of the best tree was performed using Dendroscope software [103].

2.3 Molecular protein modelling

In the absence of a crystal structure, proteins can be modelled and used in study of their structure and function through techniques such as molecular dynamics (MD) simulations. This study constructed its protein models using I-TASSER [104]. The I-TASSER program uses both threading and *ab initio* techniques. It first identifies proteins from the PDB through multiple threading approaches using Local Meta-Threading Server (LOMETS). LOMETS is an approach that uses both template-based and structure-based methods for protein structure prediction. This is performed using deep learning-based threading methods such as EigenThreader and profile-based methods such as HHpred [105]. Structural alignment is then

performed on these high-scoring templates derived from LOMETS for refinement purposes, and predicted models are calculated based on protein structures common across the templates. In addition, besides structural similarity, protein function is also considered in the structural analogues identified to the predicted final model. Full-length atomic models were then generated by iterative template-based fragment assembly simulations [106]. TM-score and confidence score (c-score) were used by I-TASSER to evaluate the quality of the modelled proteins. C-score scoring algorithm generates confidence estimates in the overall quality of the predicted model, while TM-score gives a global overview of the protein topology in terms of structural similarity between two proteins. In this case, TM-score analyses the structural similarity between the modelled protein and its respective template. The scoring algorithms are represented in the equation below (Equation 1):

$$TM - score = Max \left[\frac{1}{L} \sum_{i=1}^{L_{ali}} \frac{1}{1 + d_i^2/d_0^2} \right]$$

Equation 1

Where L is the target protein length, L_{ali} is the number of common residues between two proteins, and d_i is the distance between the i^{th} pair of corresponding residues in the two models, which is dependent on superposition matrix. The term ‘max’ maximises the sum after the superposition procedure is performed. For instance, in random protein pairs, the TM-score is normalised where the magnitude of the average score is independent of the size of the proteins by defining scale $d_0 = \sqrt[3]{(L - 15) - 1.8d}$ [107].

On the other hand, c-score evaluates the protein model accuracy based on the threading alignment and the consensus achieved after refinement process according to the equation below (Equation 2):

$$C - score = \ln \left[\frac{M}{M_{tot}} \times \frac{1}{\langle RMSD \rangle} \times \frac{1}{7} \sum_{i=1}^7 \frac{Z(i)}{Z_0(i)} \right]$$

Equation 2

Where M is the number of structure decoys in the cluster, M_{tot} is the total number of decoys, and $\langle RMSD \rangle$ is the average RMSD of the decoys to the cluster centroid produced during the I-TASSER simulations. $Z(i)$ is the Z-score of the best template generated by i^{th} threading in the seven LOMETS programs, and $Z_0(i)$ is a program-specified Z-score cut-off for discriminating between high and low quality templates [107]. After the scoring process was complete, geometric optimisation of the model was performed using the ChiRotor algorithm to improve its accuracy [108]. This was followed by energy minimisation before the quality of the model was evaluated.

2.4 Molecular docking

Molecular docking is a modelling technique employed in computer-aided drug discovery (CADD) field useful in predicting small molecules binding pose and affinity on its drug target. This technique is the starting point to selection of ideal protein-molecules complexes in not only biomedical research in general but specifically in drug discovery for hit compounds [109,110]. This approach has been widely used to address vast pathologies including malaria, cancer, and cardiovascular diseases.

There is a vast category of molecular docking programs useful in protein-protein modelling using quantum mechanics/molecular mechanics (QM/MM), solvent effect as well as protein-ligand modelling. On protein-ligand programs, there are three programs including GOLD [111], AutoDock [112] and GLIDE [113]. At an atomic level, molecular docking approaches are employed in target proteins to study the behaviour of small molecules in their binding site [114]. Prior to docking process, the small compounds undergo ligand preparation to add

hydrogen atoms and ensure whole chemical structure is optimised. Thereafter, docking will be performed following two steps: Firstly, the ligand conformation, position and orientation within the binding site usually referred to as ‘pose’ will be predicted using a sampling algorithm. The conformers of the ligands are generated via a computer-aided docking method which computes energetically favorable conformations based on the protein-ligand motion space. Depending on the number of rotatable bonds in the ligand and the search space, the number of conformers generated could range between 6-9 per ligand. This method can be performed in the case where the location of binding site is known referred to as targeted or unknown referred to as blind docking. The conformation sampling is performed on the entire protein following the blind docking strategy to obtain many hit compounds with no bias. Secondly, its binding affinity will be assessed based on an internal scoring function.

The protein-ligand binding mechanism adopts a lock-and-key ‘induced-fit’ theory that suggests that the protein and the ligand are flexible during docking process. This is because the nature of the binding site is altered with ever-changing interactions between protein and the ligand [115,116]. As previously mentioned, docking process involved sampling algorithms to obtain the best ‘pose’ and scoring functions that would present the optimal binding mode. Briefly, the sampling algorithms include matching algorithm, incremental construction, fragment-based methods for *de novo* design, multiple copy simultaneous search (MCSS) and LUDI, stochastic search algorithms such as Monte Carlos (MC) [117] and Genetic algorithm (GA) [117,118]. Out of these, AutoDock software used in this study used stochastic search algorithms where its earlier version adopted MC followed by GA.

Concerning scoring functions, force-field based, knowledge-based and empirical functions have since been used to estimate the protein-ligand binding affinity [119]. Our study uses force-field based scoring functions which factors in hydrogen bonds, non-bonded interactions, entropy contributions and solvations [120].

2.5 Molecular dynamic simulations

There are various reviews on the advance and applications of MD simulations in drug discovery [121–123]. MD simulation is a technique that models an actual protein system to study its structure, function and investigate protein stability. Essentially, through MD, the behaviour of protein molecules and their inter-atomic interactions are studied as a function of time. These forms different motions grouped into structural conformations that help in deciphering the protein structure and function. Another importance of conducting MD simulations is the ability to alter factors and parameters in the simulating system to determine the effect or contribution of a particular property examined [124]. The first protein MD simulation was performed in the late 1970s and is now widely used by different disciplines besides biomedical research [125].

To actualize the real behaviour of biomolecules in motion, various parameters must be incorporated to fit quantum-mechanical calculations. This parameterization would then allow chemical bonds and angles, atomic charges, electrostatic interactions, and energies to be defined, collectively called force-field parameters. Altogether, these parameters allow the successful implementation of MD simulations. The commonly used force-fields include AMBER [126,127], CHARMM [128] and GROMOS [129]. Additionally, MD simulations are often conducted in combination with other experimental structural biology techniques. These techniques include x-ray crystallography, nuclear magnetic resonance (NMR), cryo-electron microscopy (cryo-EM), electron paramagnetic resonance (EPR) and forster resonance energy transfer (FRET).

Concerning timesteps during simulations, a short simulation is often adopted because majority of biochemical events take place on timescale of nanoseconds, microseconds or longer when it comes to functionally important structural changes in the protein. Thus, to achieve the biological objective, most simulations are computationally expensive as simulations involve

millions of timesteps. To circumvent this limitation, the introduction of graphics processing units (GPU) has allowed simulations to be scaled up to conduct powerful simulations at an affordable cost [130,131].

2.6 Dynamic residue network (DRN) analysis

In a protein system, residue-residue interactions are important in defining its structure and biological activity [132]. As such, studying residue interactions is important due to the role they play in biological processes such as enzyme catalysis, protein folding and allosteric pathways [133–135]. In a protein residue interaction network (RIN), every residue belongs to a node where atoms are between two nodes. Because RIN is dependent on trajectory for analysis, combining its analysis with MD simulation which is time-dependent allows comprehensive elucidation of proteins structure and function. Previous studies have studied residue interactions networks in various proteins [136,137].

A total of five DRN centrality metrics were analysed. These were *Averaged betweenness centrality (BC)*, *closeness centrality (CC)*, *degree centrality (DC)*, *eigen centrality (EC)* and *Katz centrality (KC)*. *BC* accounts for the frequency of a node acting as a bridge to connect two other nodes along the shortest paths in a graph. *CC*, the inverse of L , measures the distance between a node and all other nodes, where those with the shortest distance have high closeness. *DC* computes the total number of edges present in a node. *EC*, also known as eigenvector centrality, measure a node's importance relative to the importance of its surrounding nodes. It does this by computing the vector for each MD frame and performing the average. *KC*, like *EC*, computes the influence of a node within a network. However, the metric is also inclusive of all other nodes in a network, considering the node's immediate neighbours. The procedure for calculating all five metrics, including the equations, is detailed in previous studies according to **Table 2.1** [138].

Table 2.1: Dynamic residue network centrality metrics used to identify residues crucial for communication in FPs systems. Taken from [138].

Centrality metric	Formula	Note
Averaged BC	$\overline{BC}(v) = \frac{1}{m} \sum_{i=1}^m \sum_{u=1}^{n-1} \frac{\delta(s_i, t_i v_i)}{\delta(s_i, t_i)}$	V is the complete set of nodes; m is the number of frames; $\delta(s, t)$ is the number of shortest paths connecting nodes s and t ; $\delta(s, t v)$ is the number of these paths passing through another node v ; and i is the frame number.
Averaged CC	$\overline{CC}(v) = \frac{n-1}{m} \sum_{i=1}^m \sum_{u=1}^{n-1} d(v, u)$	$d(v, u)$ is the shortest-path distance between v and u , and n is the number of nodes in the graph.
Averaged DC	$\overline{DC}(k) = \frac{1}{m(n-1)} \sum_{i=1}^m \sum_{j=1, j \neq i}^n A_{ijk}$	n is the number of nodes; A_{ijk} is the jk^{th} adjacency for the i^{th} frame.
Averaged EC	$A \cdot \overrightarrow{EC} = \lambda \cdot \overrightarrow{EC} \quad (\text{a})$ $\overline{EC}(i) = \frac{1}{m} \sum_{k=1}^m EC_{ik} \quad (\text{b})$	(a) EC is the eigenvector, and λ is the eigenvalue for the eigen decomposition of adjacency matrix A . In NetworkX, this is obtained by power iteration. (b) Averaged EC is computed for i^{th} residue by computing the vector for each MD frame and averaging.
Averaged KC	$KC(i) = \alpha \sum_{j=1}^n A_{ij} KC_j + \beta \quad (\text{a})$ $\overline{KC}(i) = \frac{1}{m} \sum_{k=1}^m KC_{ik} \quad (\text{b})$	KC is a modification of EC that employs a dampening coefficient and a constant in order to influence adjacency values.

CHAPTER 3

COMPARATIVE SEQUENCE AND STRUCTURAL ANALYSIS

With the existing atovaquone (ATQ) resistance, there is a need to identify other alternative and selective drugs which target the Plasmodial cytochrome b protein. The ATQ drug blocks the fundamental cellular respiration process by targeting the Plasmodial cytochrome b protein, leading to parasite cell death. Cellular respiration occurs in most organisms (eukaryotes and prokaryotes), including but not limited to Plasmodium parasites and the human host. Therefore, to selectively target the Plasmodium parasite in the drug discovery process, in silico approaches are important in identifying hit compounds that not only have good binding energy and withstand ATQ-linked mutations but also target the parasite without affecting the human host. As such, to begin the drug discovery process, this chapter examines sequence and structural analysis of the cytochrome b protein, highlighting key differences between P. falciparum and the human host that could be explored in selective drug discovery.

3.1 Cytochrome *bc₁* complex

3.1.1 Structure and function

Cytochrome *bc₁* complex is an enzyme in the inner mitochondrial membrane (IMM) involved in the transfer of electrons from ubiquinol to cytochrome *c* [139]. It is a key component of the electron transport chain (ETC) that is primarily responsible for pyrimidine biosynthesis. It is a homodimer with multiple subunits involved in stabilising the protein structure [139]. Each monomer has 10 subunits that consist of two cores, five low-molecular-weight and three respiratory subunits as shown in **Figure 3.1A**. The three respiratory subunits; cytochrome *b* (Cyt *b*), rieske [2FE-2S] iron-sulphur protein (ISP) and cytochrome *c₁*, form the protein's catalytic domain. Cyt *b* carries two hemes, high potential (bH) b562 and low potential (bL) b566, which play a central role in the electron transport circuit [59,140]. Heme b562 is located close to the inner membrane while b566 is close to the outer membrane. Both heme groups are intercalated in cyt *b* transmembrane helices, while the ISP chain embeds the [2FE-2S] cluster [59,139]. The schematic representation of the protein complex is illustrated from the *Bos taurus* (PDB ID: 1PP9) cyt *bc₁* structure complex, which is embedded in the IMM, as shown in **Figure 3.1**. The rest of the subunits in the enzyme complex, away from the catalytic domain, are structural units essential in stabilising the entire protein structure that exists as a complex [139].

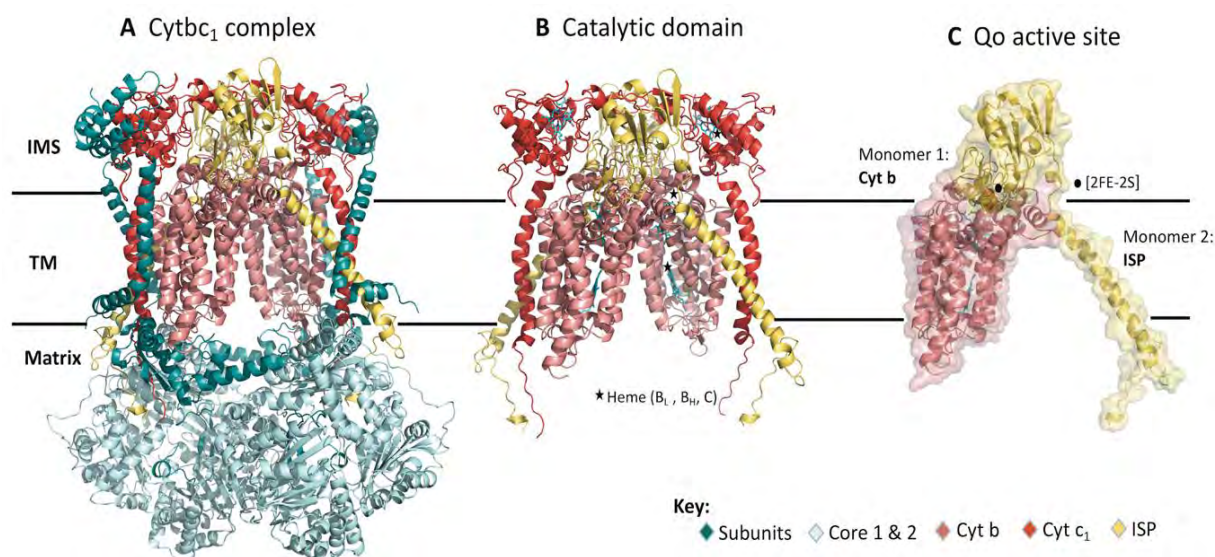


Figure 3.1: The structural model of the dimeric *bc₁* complex in bovine (PDB ID: 1PP9). (A) The crystal structure in cartoon representation shows the multi-subunit dimer. The different subunits are colour-coded (key provided) and labelled accordingly. (B) Catalytic domain composed of Cyt *b*, cyt *c₁* and ISP spans across the inner mitochondrial membrane. (C) Heterodimeric structure made up of Cyt *b* and ISP, both of which form the Qo active site. Black stars and circles represent the hemes and [2FE-2S] cluster respectively. (IMS-Intermembrane space, TM-transmembrane). Reproduced from [141].

Cyt *bc₁* complex has two discrete druggable reaction sites, Qo (outer side, upstream of b566) and Qi (inner side, downstream of b562), that are specific for ubiquinone oxidation and reduction reactions, respectively [39,53,59]. Compounds that bind at the Qo active site pockets are classified as class I inhibitors, while Qi binders are class II inhibitors (Chapter 1). The subclasses of Qo site inhibitors, subclass 1a, include β -methoxyacrylate (MOAS) compounds, while molecules that bind to class 1b include stigmatellin A and atovaquone (ATQ) compounds [142,143]. The Qo active site is made up of eight transmembrane helices, while ISP has one alpha-helix and extrinsic domain as shown in **Figure 3.1C**. The quinol oxidation (Qo) active site where substrates (such as stigmatellin and ATQ drug) bind involves cyt *b* and ISP heterodimer subunits (**Figure 3.1C**) which support the electron transfer (ET) process.

3.1.2 *Plasmodium falciparum* cytochrome *bc₁* complex

The schematic representation of the *P. falciparum* cytochrome *bc₁* (PfCyt *bc₁*) protein complex is also illustrated using the *Bos taurus* (PDB ID: 1PP9) cyt *bc₁* structure complex (**Figure 3.1**). Similarly, in *Plasmodium*, Cyt *b* subunit has eight transmembrane (TM) helices, while the ISP subunit has one alpha-helix and an extrinsic domain connected by an unusual neck or hinge region. This region is a unique feature found in *Plasmodium* spp. and most apicomplexan parasites as it is believed to have evolved to aid in catalysis of the protein complex. Additionally, the same region has been reported to be specifically useful in preserving the catalytic effectiveness of the Qo site of the protein complex [144]. The entire ISP subunit has been shown to undertake a specific structural position in the ET process. The extrinsic domain (also known as the head section) of the ISP domain undergoes an approximate 65° rotational displacement. This domain movement allows the [2FE-2S] cluster centre to move towards the Cyt *b* interface in support of the ET process [145,146].

3.1.3 Electron transfer process

Ubiquinol oxidation of cytochrome *bc₁* complex is an important step in cellular respiration and photosynthesis, yielding adenosine triphosphate (ATP) energy for the cells. This process is mainly driven by ET across the IMM, which sustains membrane potential and eventually leads to ATP production [147–149]. Briefly, the ET, also known as proton translocation, is mediated by a proton motive Q cycle, which occurs across the IMM in most eukaryotes or the plasma membrane in bacteria [150,151] (**Figure 3.2**). The mechanism carries out ubiquinol oxidation (Qo) and ubiquinone reduction (Qi) at the catalytic sites [145,152] of cytochrome *b* of one dimer and transfers electrons to [2FE-2S] cluster of the iron sulphur protein of the other homodimer. Thus, the ET process involves heterodimer pattern to complete the process. The

ET process occurs between redox centres which have metal binding features including heme and the [2FE-2S] cluster cofactors.

Q cycle

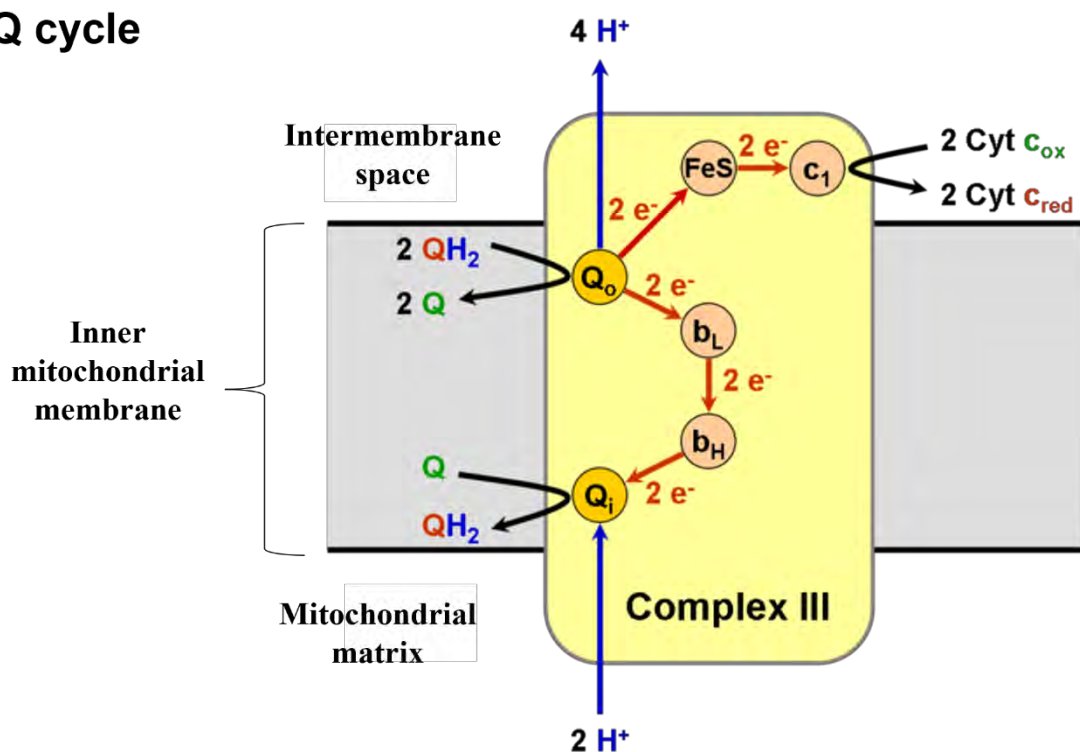


Figure 3.2: A representation of the Cyt *bc1* complex showing a section of the electron transfer pathway in the complex III, a part of the electron transport chain (ETC). The IMM is shown in the grey area where Q represents the ubiquinone form of CoenzymeQ while QH₂ represents the ubiquinol form (Adapted from https://en.wikipedia.org/wiki/Q_cycle).

One complete cycle of the redox reaction pumps a total of four protons from the matrix to the intermembrane space via the IMM, creating a proton gradient [152]. The gradient promotes the electron transfer process essential in producing ATP which is important to parasitic cellular activities. The function of the ETC has been inhibited previously, causing a collapse in its membrane potential.

3.1.4 Protein structural features and location of point mutations

Figure 3.3A shows the location of the point mutations on the PfCytb-ISP heterodimer forming the Q_o active site. The Q_o active site is comprised of Cyt *b* that has a total of eight TM helices,

while the ISP subunit has one alpha-helix and an extrinsic domain. The Qo active site is in the IMM towards the ISP chain interface, where a total of three cofactors are located. These include a [2FE-2S] cluster and two b-type hemes, heme bH and heme bL, where subscripts H and L refer to high and low potential respectively, as displayed in **Figure 3.3B, C**. The two b-type hemes are intercalated in the transmembrane helices of the Cytb subunit, while the [2FE-2S] cluster is embedded in the extrinsic domain of the ISP chain between two loops held together by a disulphide bridge [142,143].

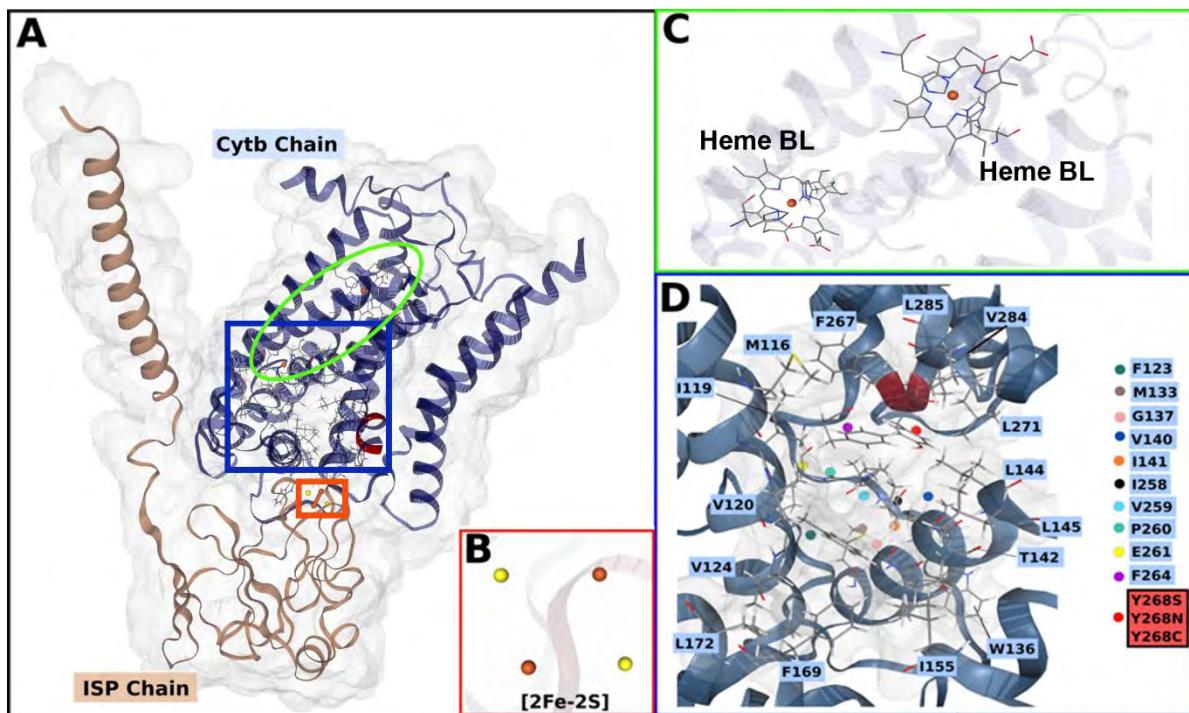


Figure 3.3: 3D cartoon representation of PfCytb-ISP protein complex showing the location of Qo active site and the location of point mutation Y268C/N/S. (A) PfCytb-ISP structure with different subunits colour coded (key provided) and labelled accordingly. The red, green, and blue delimitations indicate the [2FE-2S] cluster, the heme groups and the active site, respectively. (B) [2FE-2S] cluster, showing sulphur atoms in yellow and Fe²⁺ atoms in orange. (C) Heme bL and bH groups. (D) The zoomed in active site showing its contributing residues which are highlighted in blue. The point mutations at position 268 are shown in the red coloured box and indicated by a red sphere in the active site. This position is also coloured red on the cartoon structure. Reproduced from [141].

3.1.5 Sequence characterisation of *P. falciparum* cytochrome *b* protein

Comparative studies have been performed focusing on the cytochrome *b* sequence to gain insights on the evolutionary relationships that exist among different organisms. Such studies have sought to determine phylogenetic relationships of insects and apes/mammals at sequence level [153,154]. At the sequence level, physico-chemical properties can be determined including molecular sizes, nature, and length of amino acids, after which multiple sequence alignment (MSA) is calculated for comparison. There are various alignment methods that have been implemented using different algorithms.

With reference to sequence conservation, Cytb sequence patterns of functional importance have been previously identified. Mitochondrial PEWY motif, also known as Qo motif, has been reported to function in Qo substrate binding [149]. Specifically, the mitochondrial PEWY motif has been identified in many organisms, including but not limited to insects, bacteria, humans, parasites (for instance, *Plasmodium* spp.), and plants. However, there are other alterations of the PEWY motif that have been reported. PDWY is present in gram-positive bacteria, PVWY in proteobacteria, and PPWF has been identified in archaea organisms only [149]. These alterations in the PEWY motif have been attributed to the evolutionary process where other organisms have adapted better to different potentials in their ET systems. For instance, PEWY and PDWY have been reported to be catalytically efficient with high and low potentials, respectively.

This chapter characterises the *P. falciparum* cytochrome *b* comparing it to other related homologs including human cytochrome *b*. The study analyses were performed using computational tools. The intention of this comparative study was to identify significant and unique differences that can be explored for selective drug inhibition studies. Thus, their sequence and structure were analysed using MSA, phylogenetics, motif features and other physico-chemical properties.

3.2 Methodology

3.2.1 Sequence retrieval and characterisation at residue and structure level

Focusing on cytochrome *b* as the drug target, the protein sequences of PfCytb (UniProt accession: Q02768) subunit were retrieved from the Universal Protein Resources (UniProt) [155] and compared with those from PlasmoDB [156]. Thereafter, a total of 74 cytochrome *b* protein sequences from mammalian (human) host, parasites (*Plasmodium* spp. and other apicomplexans) and vectors were retrieved from NCBI database [157] and PlasmoDB (release 49) database [156]. The *P. falciparum* cytochrome *b* was used as the query and the BLASTp tool as the homology search program. These were utilised in multiple sequence alignment (MSA), phylogenetic tree and motif calculations. E-value and sequence coverage were considered for each sequence retrieved for further analysis.

Physico-chemical properties were derived from protein sequences based on atomic and amino acid composition, the number of amino acid (aa) residues, as well as a grand average of hydropathicity index (GRAVY). This was performed using PSIPRED program [158] which uses amino acid sequences to characterise the protein at a primary and secondary level including secondary structure predictions. Briefly, GRAVY is a parameter used to determine the hydrophobicity and hydrophilicity of a peptide based on its aa sequence using a hydropathy scale composed by Kyte and Doolittle [159]. GRAVY reports a value calculated from the sum of all the amino acids' hydropathy values divided by the length of the aa sequence. Also, the protein was subjected to ProDOM database [160] to predict the functional domains present in the cytochrome *b* protein (n = 10 sample dataset).

3.2.2 Multiple sequence alignment

MSA was performed using three tools: stand-alone MUSCLE, web servers MAFFT's FFT-NS-I, a progressive method employing iterative refinement approach, and PROMALS3D for comparison [83,85,86]. In the multiple sequences dataset, sequence from crystal structure (*Bos taurus* PDB ID:1CYO) was utilised as a standard reference. Gap open and extension penalty parameters of 1.53 and 0.123 were used, respectively.

Both these programs utilise the local alignment approach to identify conserved blocks of functionally relevant regions of the protein, e.g., conserved sequence motif. The structural alignments from these tools were compared to determine the alignment accuracy. Visualisation of the alignments was performed using Jalview software [93]. From the MSA results, conserved regions, insertions and unique residues in both *Plasmodium* and mammalian sequences were compared and highlighted. Out of these tools, the PROMALS3D tool produced accurate MSA; thus, its results were used in subsequent analysis.

3.2.3 SNP identification and prediction of their effect on cytochrome *b* modelled protein structure

The reported point mutations (Y268C, Y268N and Y268S) conferring PfCytb resistance to the ATQ drug were obtained from the literature [161–163]. Using WT (reference), Y268C, Y268N and Y268S modelled structures, brief analysis was performed using VAPOR (Variant Analysis Portal) [164], PhD-SNP [165] and DynaMut [166] to characterise the three SNPs to predict their potential effect on the structure, function and stability of the PfCytb-ISP protein. VAPOR, a tool integrate into the HUMA interface [164], utilises several tools to predict the effect of mutations, herein referred to as 'variants', on protein function as well as on its stability. Using protein sequence as its input and selected variants, the sequence is screened and characterised by six tools under VAPOR workflow. These tools include: I-Mutant 2.0 [167], MuPRO [168],

PolyPhen-2 [169], PROVEAN [170], PhD-SNP [165] and FATHMM [171]. These tools use sequence and structure-based features to compare wild-type and mutant alleles and correlate to either non-damaging versus damaging alleles known to cause disease (information obtained from UniProt [155]). PhD-SNP incorporated also in VAPOR, aligns wild-type and mutant protein sequence, and uses conservation index in the mutation position to classify the mutations as either disease causing (set to 0) or neutral (set to 1). On the other hand, DynaMut utilises normal mode analysis (NMA) to determine protein motions between the wild-type and mutant protein, in this case modelled structures. Thereafter, the program assesses the vibrational entropy changes between the two structures to determine the effect of the mutations on protein dynamics and stability.

3.2.4 Phylogenetic tree calculations

In this study, protein phylogenetic trees were created using RAxML [94] program which utilised the Maximum Likelihood (ML) tree-building algorithm for a total of 74 sequences previously aligned. For these phylogenetic tree calculations, amino acid models, incorporated in RaxML program, were used to determine evolutionary changes in multiple aligned proteins. These models included MTREV [95], JTT [96], RTREV [97], Le and Gascuel (LG + G + I) [98] and WAG [99] that were used in calculations. Other parameters include (-f a) option which utilised rapid bootstrap algorithm to search for ML trees and -p option to permit reproducibility of runs (-p option). Gap deletions/missing data found in MSA, herein referred to as ‘gappy phylogenomic alignments’ (RAxML term), are factored in automatically during phylogeny construction using subtree equality vectors. The program considers gap deletions/missing data in its dataset where the amount of missing data ranges typically from 50 % to 90 %. As such, it screens, categorises, and represents alignments with gaps by 1 or non-gap sites by 0. That way, all gap sites are determined and factored in by calculating ancestral probabilities and

thereafter calculate the branches for tree phylogeny [100]. Using 1000 bootstrap replicates in each calculated tree, the overall best phylogeny was selected from a model having strong branching based on bootstrap values. Numbers on tree branches represent posterior probabilities as percentages. Bayesian posterior probabilities have since been used to determine the phylogeny accuracy. Therefore, trees with bootstrap values cut-off > 70 % were considered 95 % accurate and reliable [101,102]. Visualisation of the best tree was performed using Dendroscope software [103].

3.2.5 Protein modelling of the Qo oxidation site of the PfCytb-ISP

The focus of the entire study was to inhibit the Qo active site of cytochrome *b*. Therefore, PfCytb and ISP, relevant subunits in the entire complex that form the Qo inhibitor binding pocket (for drug compounds such as atovaquone (ATQ) and its analogues) were incorporated in this study. Briefly, to model the heterodimer complex PfCytb-ISP, the Universal Protein Resources (UniProt) [172] was used to access and download *P. falciparum* cytochrome *b* (UniProt accession: Q02768) and iron-sulphur protein (ISP) (UniProt accession: Q8IL75) subunits [172]. Crystal structures of the mentioned protein subunits were not available, therefore PfCytb and ISP protein models were constructed using I-TASSER (Iterative Threading Assembly Refinement) [104,106,173] as detailed in Chapter 3. The *Bos taurus* crystal structure (PDB ID: 1PP9), retrieved from PDB [174] was used as the principal structural template in guiding the modelling process. It was selected due to its high sequence identity to the *P. falciparum* protein where they share 76% similarity and 43% sequence identity.

To obtain a complex between the two predicted models (PfCytb and ISP), the individually modelled subunits were separately superimposed onto corresponding chains of homologous *Bos taurus* cytb_{c1} complex (PDB ID: 1PP9) using PyMOL [175]. This approach was performed to obtain coordinates of the cofactors, as the I-TASSER web server does not model cofactors.

It was performed by first superimposing the modelled PfCytb subunit onto the crystal structure, after which the ISP chain was then superimposed onto the resulting complex to form the final PfCytb-ISP complex. This ensured that the model was in a similar orientation to that of the template, which would enhance the ease of transferring all the coordinates of the prosthetic groups. The coordinates for b-type hemes and [2FE-2S] cluster prosthetic groups on x-ray crystal structure 1PP9 were then directly transferred onto the modelled complex. The overall structural quality of the PfCytb-ISP complex was further evaluated using z-DOPE score, RAMPAGE [176] and PROCHECK [177].

For predicting the effect of mutation on static protein structures, and for future studies, mutant models were constructed. The reported point mutations (Y268C, Y268N and Y268S) conferring PfCytb resistance to the ATQ drug were obtained from the literature [161–163]. The crystal structures of the mutants (Y268C, Y268N and Y268S) were not available within the PDB; therefore, mutant protein models were generated using the wild-type (WT) protein as a template. This was performed by mutating the Tyr268 position to the appropriate mutant using Discovery Studio 2016 [178] followed by structure minimisation to remove any bad contacts and relax the structure. Structure minimization was performed using AMBERTools19 [179] program considering 20000 minimisation cycles: 10000 cycles of steepest descent followed by 10000 cycles of conjugate gradient minimisation. A non-bonded cut-off distance of 10 Å was applied. Thereafter, model quality for each mutant model was then evaluated and validated according to that of the WT protein. Thereafter, structural analysis was performed.

3.2.6 Motif calculations and structural mapping analysis

Further sequence analysis was performed on the PfCytb protein to identify functional patterns in the protein. Motif analysis was performed using Multiple Expectation Maximisation for Motif Elicitation (MEME) version 4.11.2 [180]. The following parameters were utilised in the

search for conserved motifs unique to PfCytb protein within the dataset of 74 proteins. The command line with the parameters was as follows; -nostatus -time 18000 -maxsize 160000 -mod zoops -nmotifs X -minw 6 -maxw 50. Briefly, a maximum sequence width of 6-50 protein residues was applied and -nmotifs X was varied until no more motifs were discoverable, from 100 to 50 then finally a total of 35 motifs were calculated in this dataset. Thereafter, MAST (Motif Alignment and Search Tool) was used to identify overlapping motifs [181] which will be eliminated to minimise redundancies in further analysis. MAST calculates match score for each sequence comparing with each of the motifs (residue by residue) in the group of motifs identified. Based on this pairwise correlation, the matrix would identify the motifs that are in proximity based on the match score (converted to p-values) to determine overall match. From the Pearson's correlation coefficients, the motifs having correlations < 0.60 are referred as accurate and significant while those > 0.60 are termed less accurate as they overlap with another motif hence, they are eliminated [181,182]. MAST and MEME files were further characterised using an in-house Python script (*motif_analyzer.py*) to generate a heat map that represents the distribution of different motifs based on the degree of conservation [101]. The heat map shows motif conservation across all the 74 sequences characterised. Using PyMOL [175], the unique motifs in *Plasmodium* cytochrome *b* protein were mapped onto the wild-type modelled protein structure.

3.3 Results and discussion

3.3.1 Sequence structural and functional analysis

The sequences ($n = 74$) retrieved were tabulated (**Table S3.1**), showing details such as their query coverage, sequence identity and E-values for reference. These sequences were used in MSA, sequence identity, motif discovery calculations and phylogenetic analysis. To

investigate sequence, structural and functional features unique to *P. falciparum* cytochrome *b*, 10 species were sampled for comparison purposes. These species comprised five *Plasmodium* spp., *Bos taurus* (*B. taurus*), *Toxoplasma gondii* (*T. gondii*), *Saccharomyces cerevisiae* (*S. cerevisiae*), *Gallus gallus* (*G. gallus*) and *Homo sapiens* (*H. sapiens*). Firstly, physico-chemical properties were derived from protein sequences based on atomic and amino acid composition, the number of amino acid (aa) residues, as well as a grand average of hydropathicity index (GRAVY) as shown in **Table 3.1**. Briefly, GRAVY is a parameter used to determine the hydrophobicity and hydrophilicity of a peptide based on its aa sequence using a hydropathy scale composed by Kyte and Doolittle [159]. GRAVY reports a value calculated from the sum of all the amino acids' hydropathy values divided by the length of the aa sequence. For interpretation purposes, positive GRAVY values indicate a hydrophobic protein while negative values mean a hydrophilic protein.

Table 3.1: Physico-chemical properties of the cytochrome *b* amino acid sequence of 10 species.

Species ID	Length (no. of aa)	Formula (atomic composition)	Grand Average of Hydropathy (GRAVY)
<i>P. falciparum</i>	376	C ₂₁₀₃ H ₃₁₂₇ N ₄₇₁ O ₄₉₄ S ₁₅	0.846
<i>P. malariae</i>	380	C ₂₁₃₅ H ₃₁₆₉ N ₄₇₅ O ₅₀₄ S ₁₃	0.772
<i>P. knowlesi</i>	376	C ₂₁₁₇ H ₃₁₄₁ N ₄₆₉ O ₄₉₈ S ₁₃	0.813
<i>P. ovale</i>	376	C ₂₁₀₃ H ₃₁₂₀ N ₄₇₀ O ₄₉₇ S ₁₃	0.802
<i>P. vivax</i>	382	C ₂₁₄₀ H ₃₁₉₁ N ₄₈₁ O ₅₀₅ S ₁₃	0.800
<i>H. sapiens</i>	378	C ₂₀₁₀ H ₃₀₇₂ N ₄₆₆ O ₅₀₈ S ₁₇	0.650
<i>T. gondii</i>	368	C ₁₉₇₄ H ₃₀₁₉ N ₄₆₃ O ₄₈₂ S ₂₀	0.877
<i>B. taurus</i>	379	C ₂₀₁₇ H ₃₀₇₂ N ₄₇₂ O ₅₀₂ S ₁₉	0.666
<i>S. cerevisiae</i>	385	C ₂₀₈₁ H ₃₁₂₂ N ₄₈₄ O ₅₀₄ S ₂₀	0.819
<i>G. gallus</i>	380	C ₂₀₂₀ H ₃₀₆₆ N ₄₇₈ O ₅₀₅ S ₁₃	0.591

The GRAVY values of all the 10 species were reported to be positive. Specifically, the cytochrome *b* sequence for *Plasmodium* and the human homolog were confirmed to be both hydrophobic in nature.

Secondly, the secondary structure of cytochrome *b* protein was investigated to decipher its structural composition; alpha-helix, beta strands/sheets and coils. The prediction was performed narrowing comparison to *P. falciparum* and *H. sapiens* cytochrome *b* protein sequence (Figure 3.4).

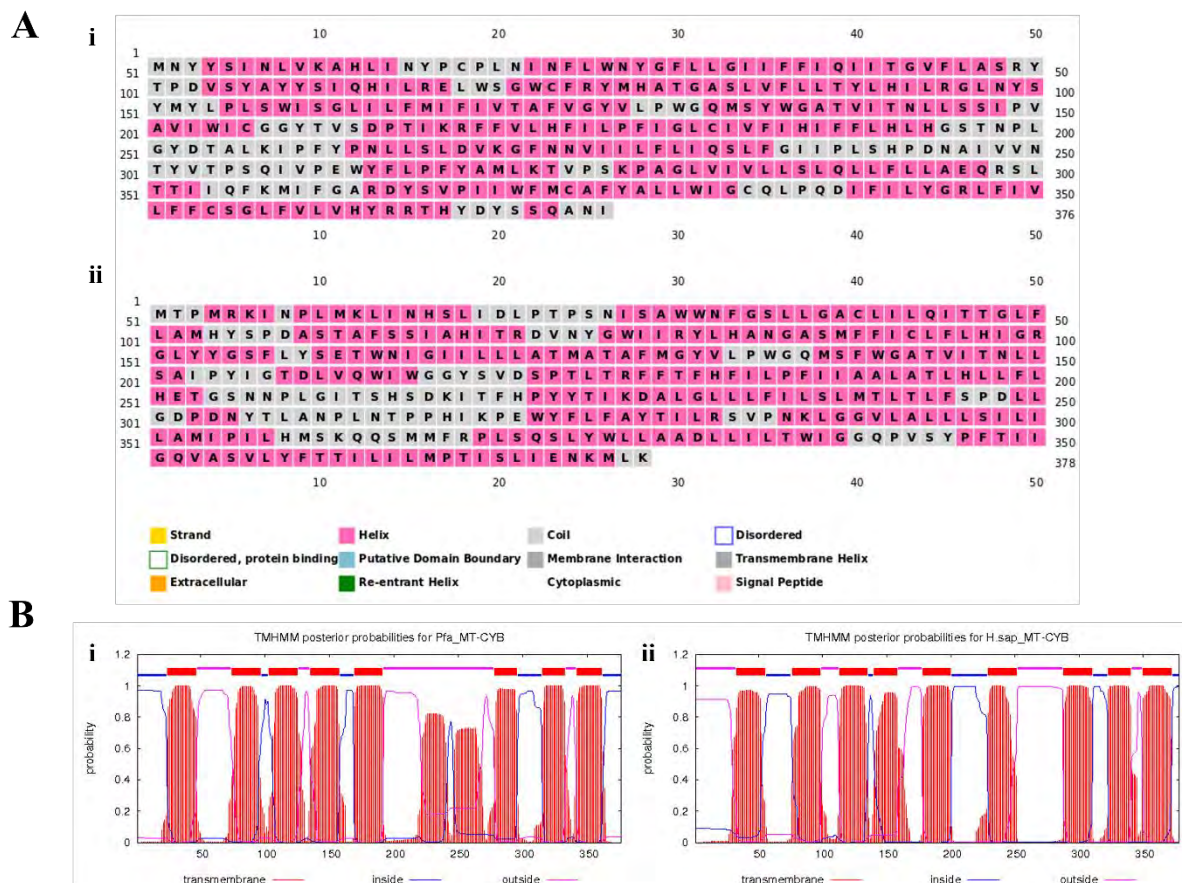


Figure 3.4: Secondary structure prediction between i) *P. falciparum* cytochrome *b* and ii) human homolog using PSIPRED. (A) Types of secondary structures, mainly composed of alpha helices and random coils ‘loops’; (B) Shows the type of alpha helices (transmembrane) that make up cytochrome *b* protein (based on location).

PSIPred results as shown in **Figure 3.4** predicted the secondary structures to be composed of majorly of alpha-helices with random coils/loops located sparsely within the protein. These helices were predicted to be transmembrane based on their location spanning across the double inner mitochondrial membrane, thus it is categorised as an integral membrane protein as calculated by the TMHMM 2.0 server [183].

The functional domain of cytochrome *b* was also investigated where ProDOM database [160] predicted the functional domains by clustering homologous segments across the 10 cytochrome *b* sequences. **Figure 3.5** shows the domain families predicted across 10 sequences, while common domains across these sequences were tabulated (**Table 3.2**).

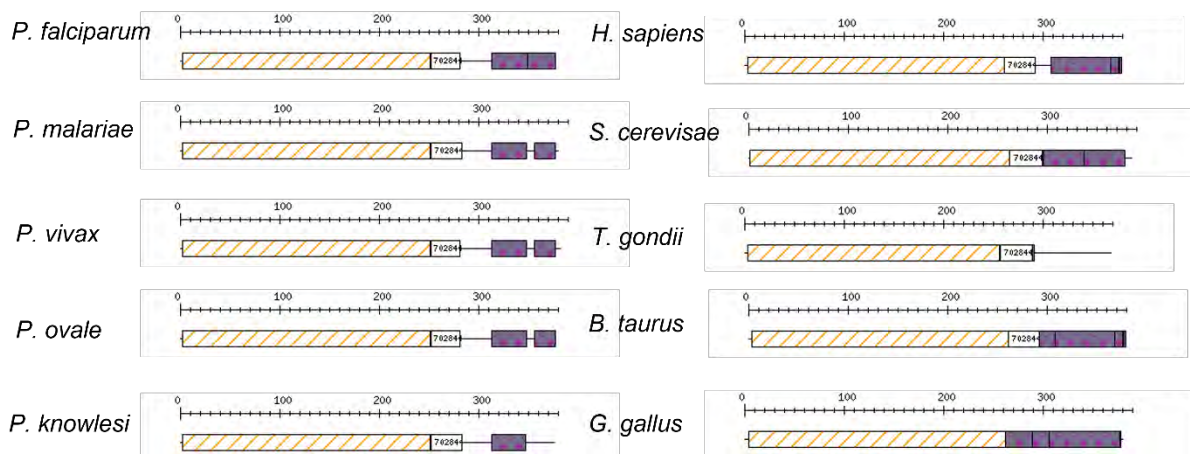


Figure 3.5: Graphical results of cytochrome *b* functional domains in 10 Cyt *b* protein sequences including *Plasmodium* spp., human homolog, and other species for comparative purposes. This was performed using ProDOM server [160].

ProDOM server [160] employs a Blast-P tool (NCBI) to search against consensus sequences provided by an amalgamation of databases including PROSITE [184], PFAM [185] or SMART [186]. Based on the comparative search and consensus achieved, the sequences were ranked

based on similarity scores, with the highest scores followed by lowest scores, as well as factoring an E-value of 0.01.

	PD000016		PD702844	
	Position	E-value	Position	E-value
<i>P. falciparum</i>	2-250	1e-162	252-280	3e-10
<i>P. malariae</i>	2-250	4e-157	252-282	4e-11
<i>P. vivax</i>	2-250	3e-160	252-280	4e-10
<i>P. ovale</i>	2-250	4e-157	252-280	4e-10
<i>P. knowlesi</i>	2-250	1e-160	252-282	4e-11
<i>H. sapiens</i>	3-260	2e-156	261-291	4e-12
<i>S. cerevisiae</i>	1-261	0	262-294	4e-12
<i>T. gondii</i>	3-255	2e-180	257-288	8e-12
<i>B. taurus</i>	3-260	2e-170	261-291	4e-12
<i>G. gallus</i>	4-261	0	262-288	4e-09

Table 3.2: Functional domains that were common across all the 10 cytochrome *b* protein sequences.

In **Figure 3.5**, showed predicted domains in rectangular bars that span across the ten sequences. Out of these, there was similar trend observed across the dataset that suggest a possibility of common domains thereby showing how conserved they are at a functional level. The domains commonly shared include PD000016 and PD70284, both of which were predicted to function in the transmembrane heme electron transport chain (ETC) [139]. Functionally, they were also predicted to have heme iron metal binding sites, confirming the two heme cofactors that bind in both Qo and Qi active sites in cytochrome *b* to facilitate the electron transfer pathway [187,188].

3.3.2 Modelling results for PfCytb-ISP protein

Due to the absence of the WT Plasmodial *cytb_{c1}* complex x-ray crystal structure, PfCytb and ISP subunits were modelled separately using I-TASSER [104,106,173]. The structural templates were 3CX5 [189], IBGY [190], IPPJ [191], 1EZV [192] and 1PP9 [191] from the Protein Data Bank (PDB). Using HHpred [193] and PRIMO [194], the aforementioned templates had respective sequence identities of 37%, 37%, 41%, 37% and 41% for the PfCytb subunit and respective sequence identities of 43%, 43%, 43%, 43% and 43% for the ISP subunit. PDB ID: 1PP9 (*Bos taurus*), having an experimental resolution of 2.1 Å, was selected as the main template and used to guide the modelling process [193]. The precision of a modelled structure has been suggested to be more refined upon the use of several templates, and this was utilised in this study to improve model quality and resolution of more mobile protein segments [195,196]. After modelling, the quality of the structures was further assessed using the in-built I-TASSER metrics, Ramachandran plot and z-DOPE score.

A TM-score > 0.5 indicates models with correct topology, as opposed to < 0.17 which signifies random topology [104]. C-score is calculated based on clustering structural density, a consensus on multiple threading templates and the convergence of structure assembly. It ranges from -5 to 2; higher values denote higher quality. The TM-score algorithm ensures the least coordinate difference in each modelled protein is measured [197]. TM-scores of 0.99 ± 0.04 and 0.74 ± 0.12 were reported for the PfCytb and ISP chains, respectively. A c-score of 2 was reported for both models, denoting good quality [104].

A Ramachandran plot of the proteins' psi (ϕ) and phi (ψ) main-chain torsion angles identifies the amino acids with acceptable torsion angles and the outlying residues [198]. The Ramachandran plot showed that 88.9 % of the total residues for PfCytb and 83.4 % of ISP

modelled subunits were in the most favoured regions, while 0.3 % and 0.7 % of residues, respectively, were in disallowed regions. These plots are presented in **Figure 3.6** [177].

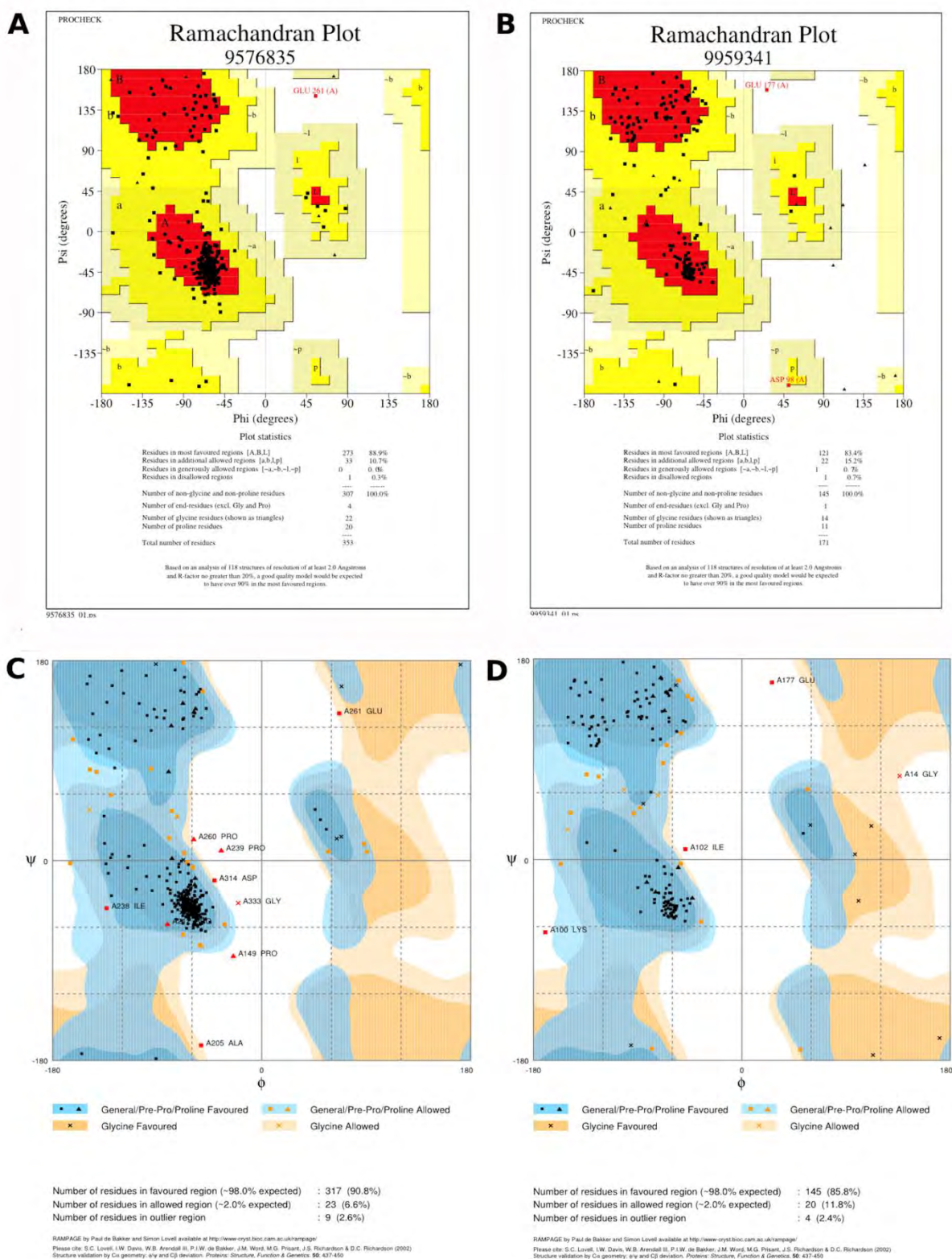


Figure 3.6: Predicted protein model validation in the form of Ramachandran plots; (A) and (B) representing WT PfCytb and ISP respectively using PROCHECK, as well as (C) and (D) representing PfCytb and ISP respectively using RAMPAGE. Most of the residues are observed to be in the most favoured and allowed regions. Overall, the model quality was good.

The z-DOPE score was then used to evaluate the modelled PfCytb-ISP protein complex. The calculated z-DOPE score of the complete PfCytb-ISP protein was -0.76. The score was close to -1.0, indicating that the model was of high quality. It has been reported that with careful model validation the use of homology models can be reliable and in good accordance with experimentally derived structures [199]. In the present study, the predicted structure quality was shown to be suitable for subsequent experiments using several validation metrics.

A comprehensive literature search of ATQ resistance mutations identified Y268C, Y268N and Y268S to be associated with drug resistance [60,107,161–163]. To generate these mutant proteins, the modelled WT protein structure was subsequently mutated using Discovery Studio Visualizer to generate the respective mutant protein models. Model evaluation determined Y268C, Y268N and Y268S to be associated with z-DOPE scores of -0.76, -0.75 and -0.75, respectively. These mutant models were validated to be reliable for further analysis of their structure and function.

3.3.3 Structural analysis of *P. falciparum* cytochrome *b*

Protein structural analysis was performed on the *P. falciparum* protein comparing it to its human homolog (**Figure 3.7**) to support future drug design studies. Key sequence features were highlighted in the *Plasmodium* structure by performing molecular structural super-imposition with *H. sapiens* cytochrome *b* structure. It was noted that the protein had a total of eight main helices with several intercalating loops. There were two heme cofactors (heme bL and heme bH) located in the cytochrome *b* protein near Qo and Qi active site pockets, respectively. These metal cofactors are responsible for redox processes that support the ET pathway between the active sites [200].

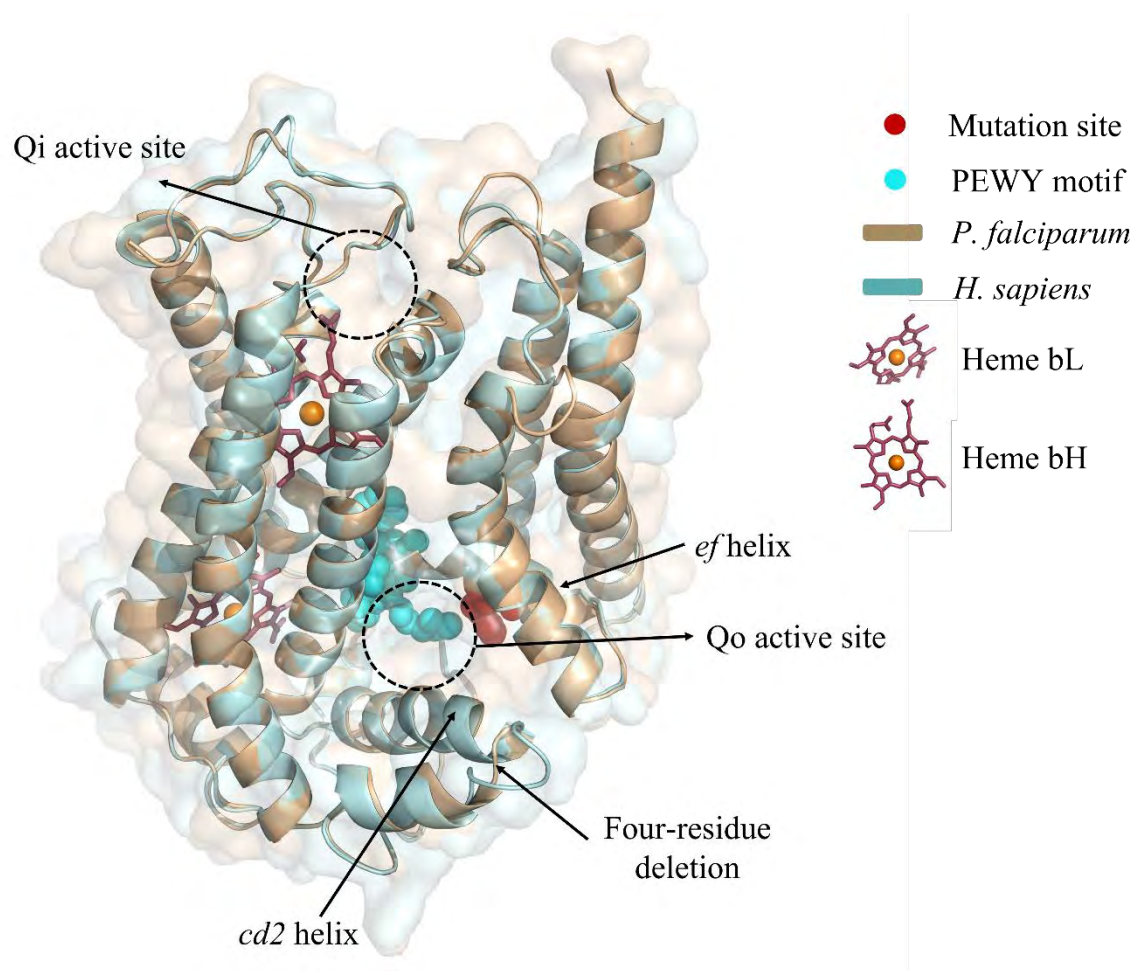


Figure 3.7: A superimposition of *P. falciparum* and *H. sapiens* cytochrome *b* structures to highlight their structural differences. A colour scale for key structural features is provided.

It is important to note the location of the *cd2* and *ef* helices which are essential in the catalytic function. After the structural overlay was performed, the four-residue deletion was noted to be located at the *cd2* helix, while the PEWY motif was located on the *ef* helix. The deletion caused a 0.83 Å displacement of the Qo active site as confirmed by previous studies as well as this study [107]. Based on literature, the PEWY motif is essential catalytically as it stabilises the substrate in its Qo active site [187]. This stability is provided by the hydrogen bond formed between the Glu261 residue (*P. falciparum* numbering) and the Qo substrate, such as the ATQ drug [201,202].

The point mutation location Tyr268 is an active site residue that participates in hydrophobic interactions in its wild-type state. When a mutation occurs there is a high likelihood of compromised residue interactions [141].

3.3.4 Prediction analysis on the effect of mutations on proteins' sequence and structure

Following ATQ treatment failures, clinical studies have linked resistance to single nucleotide polymorphisms (SNPs), also known as point mutations Y268C, Y268N and Y268S, located on an alpha-helix within the Qo active site vicinity [107,162,202,203]. Based on literature, the mutations occur through amino acid substitution from Tyr to Cys, Asn and Ser, which does not alter the secondary structure of the protein. However, these mutations have been implicated in altering the active site pocket volume. This increased pocket volume observed in all mutant systems compared to the WT is likely to lead to loss of crucial hydrophobic drug interactions, thus interfering with expected drug binding [203]. For instance, Tyr substitution to either Cys or Ser has been reported to decrease the hydrophobic interactions, whereby the aromatic group present in Tyr is replaced by nucleophilic side chains, thus decreasing the hydrophobic interactions. Clinical studies have linked resistance to the reported point mutations following ATQ treatment failures; out of which Y268S is more prevalent, followed by Y268C and Y268N [162,163,204]. The Y268S point mutation has been shown to compromise hydrophobic interactions between ATQ and amino acid residues as well as alter the active site pocket volume [203,205]. To provide more insight into the potential effect of mutations on the PfCytb-ISP protein, the location of these point mutations, shown in **Figure 3.3D**, was investigated with reference to protein secondary structure. Point mutations Y268C, Y268N and Y268S are located on an alpha-helix within the vicinity of the Qo active site. Secondary structure prediction was performed not only in WT but also in mutant proteins (Y268C, Y268N and Y268S). From the secondary structure prediction findings in mutant proteins (**Figure S3.1**),

the amino acid substitution from Tyr to Cys, Asn and Ser does not alter the secondary structure composition of the protein. However, the protein could possibly assume multiple structural conformations during MD simulations (further details on structural changes in Chapter 5: PART I - III). However, these mutations have been implicated to alter the active site pocket volume. Qo active pocket analysis using CASTp server [206,207] reported pocket volumes of 675.7, 726.6, 718.8 and 721.2 for WT, Y268C, Y268N and Y268S, respectively. The increased pocket volume observed in all mutant systems compared to the WT is likely to lead to loss of crucial hydrophobic drug interactions, thus interfering with expected drug binding [203]. For instance, Tyr substitution to either Cys or Ser has been reported to decrease the hydrophobic interactions, whereby the aromatic group present in Tyr is replaced by nucleophilic side chains thus decreasing hydrophobic interactions.

VAPOR (Variant Analysis Portal) [44] prediction tool was used to predict the effect of mutations on the function and stability of the PfCytb-ISP protein. From these analyses, three scoring metrics were calculated and reported, as shown in **Table 3.3**. The $\Delta\Delta G$ values obtained from both I-Mutant [167] and MuPro [168] were close, while slightly differing from DynaMut [166]. Nevertheless, all these metrics showed decreased protein stability in all three mutations. Data in **Table 3.3** also illustrates that the mutations were predicted to cause stability reductions within their respective proteins. Further, prediction of the effect of these mutations on the interatomic interactions occurring in their surrounding was calculated using DynaMut, and the results are shown in **Figure 3.8**.

Table 3.3: Prediction of the effect of mutations (Y268C, Y268N and Y268S) on PfCytb-ISP protein stability using VAPOR and DynaMut. The tools define $\Delta\Delta G > 0$ as stabilising and $\Delta\Delta G < 0$ as destabilising. Reproduced from [141].

Mutations			
Prediction Values			
	Y268C	Y268N	Y268S
DynaMut ($\Delta\Delta G$)	−1.06	1.45	1.49
I-Mutant ($\Delta\Delta G$)	−0.82	−0.98	−1.31
MuPro ($\Delta\Delta G$)	−0.91	−1.17	−1.37
Stability	Decrease	Decrease	Decrease
PhD-SNP	Disease	-	Disease

Various interactions formed in the WT PfCytb-ISP active site were shown in **Figure 3.8**. The Tyr residue in the WT was observed to form the highlighted residue interactions with its surrounding residues, whereas most of these interactions were seen to be lost in all three mutants (**Figure 3.8**). Interestingly, and consistent with previous findings [203], hydrophobic contacts coloured in green were seen to be lost, evidencing the detrimental effect of these mutations on protein function.

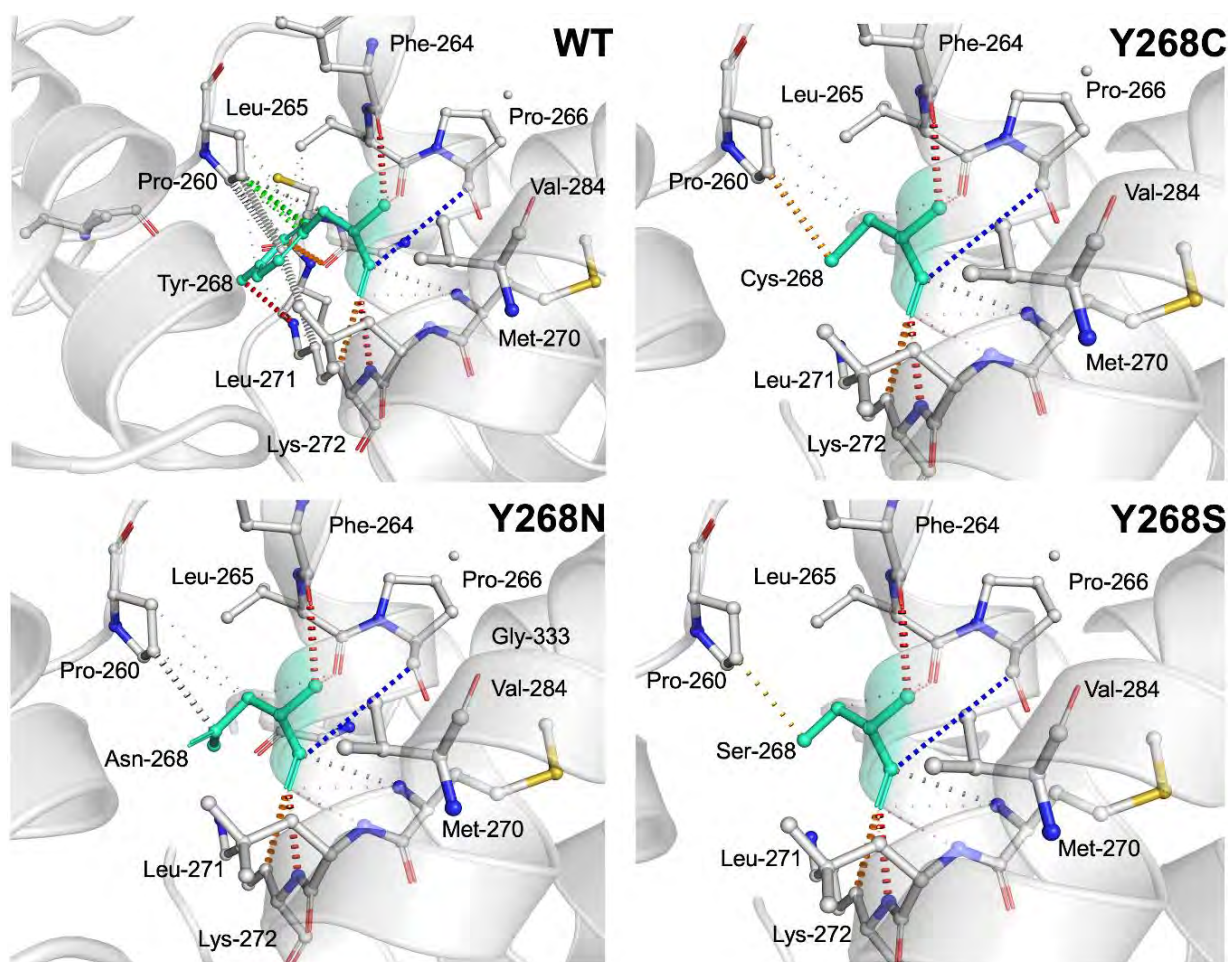


Figure 3.8: Illustration of interatomic interactions occurring around residue 268 in the WT (wild-type) and Y268C, Y268N and Y268S mutant proteins. Residue 268 in the WT and mutant proteins is coloured in light green and shown as sticks. The respective chemical interactions are labelled as dotted lines and coloured as follows: Hydrogen bonds - (red), weak hydrogen bonds - (orange), hydrophobic contacts - (green), amide-amide contacts -(blue) and ionic interactions - (gold). Amino acid residues are also coloured according to type, namely nitrogen (blue), oxygen (red) and sulphur (yellow). Reproduced from [141].

In the WT and mutant proteins (Y268C, Y268N and Y268S), the interactions are indicated by various colours. These interactions include weak van der Waals (vdW) interactions in grey (Asn268-Pro260/Met270, Cys268-Met270), hydrophobic contacts in green (Tyr268-Pro260), ionic interactions in gold (Ser268-Pro260), weak polar contacts in red (Tyr-Phe264/Lys272), water-mediated weak hydrogen bonds in orange (Tyr268-Leu271, Cys268-Pro260) and amide-amide contacts in blue (Tyr268-Val284). Overall, vdW contacts (Pro260, Leu265 and Leu271)

were observed to be lost in the three mutant proteins when compared to interatomic interactions in the WT protein,

3.3.5 Cytochrome *b* multiple sequence alignment

Multiple sequence alignment for a total of 74 sequences was performed and based on each alignment results profile, a more accurate alignment of loop regions, gaps and motifs was generated by PROMALS3D as displayed in (**Figure S3.2**). Out of the 74 cytochrome *b* sequences retrieved and aligned (**Figure S3.2**), 10 of them were further analysed in this section. These sequences were derived from five *Plasmodium* spp. affecting humans (*P. falciparum*, *P. malariae*, *P. vivax*, *P. ovale*, *P. knowlesi*), *B. taurus*, *T. gondii*, *S. cerevisiae*, *G. gallus* and *H. sapiens*. The analysis focused on a sample of the entire dataset to determine accurate information at the sequence level. After alignment was performed by multiple tools, best results from PROMALS3D were used for further analysis (**Figure 3.9**). This is because PROMALS3D produced a more accurate alignment guided by PEWY motif, the gaps, two inserts and well aligned functional residues that were identified across the entire dataset (**Figure S3.2**). For instance, functional residue Glutamic acid (E) also found in the PEWY motif is well aligned across the entire dataset. The residue is responsible for hydrogen bond formation with ligand in the Qo active site pocket hence stabilizing the ligand [201,203]. The tool uses both structural information from crystal structures (in this case *Bos taurus* PDB ID:1CYO was utilised) and sequence constraints in guiding a rigorous alignment process [83].

Performing MSA has been reported as a vital tool in *in silico* studies for defining evolutionary relationships. This has been made possible by factoring in evolutionary events such as deletions, insertions, and mutations. It has also been used in phylogenetic studies and motif analyses [208,209].

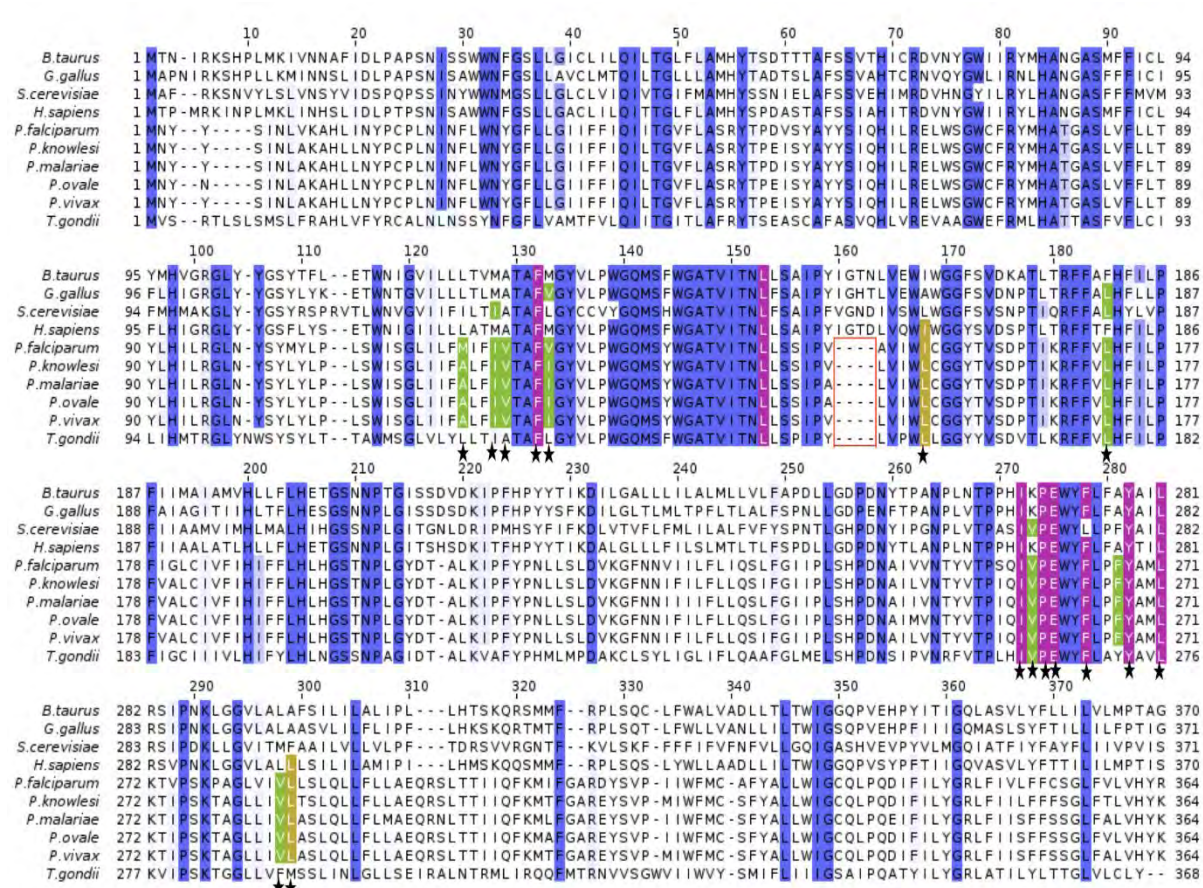


Figure 3.9: Multiple sequence alignment of 10 cytochrome *b* protein sequences. Alignment and sequence visualisation were performed using PROMALS3D and Jalview, respectively. Blue highlight shows conserved regions across the species, purple colour shows conserved active site residues, green colour shows active site residues unique to *Plasmodium* spp., while mustard colour shows active site residues commonly shared between *P. falciparum* and human homolog.

In this study, **Figure 3.9** highlights cytochrome *b* to be highly conserved as the highly conserved residues (100 %) were coloured in blue. This is in good agreement with previous studies [210]. These highly conserved regions were found to be in catalytic regions of the protein between residues 120-180 and 260-280 (*P. falciparum* numbering); residues indicated by the star symbols. A highly conserved PEWY motif (260-263: *P. falciparum* numbering) was reported in all protein sequences except in *S. cerevisiae* which had a PEWL motif. The PEWY motif is a distinct catalytic feature of the protein that is important in Qo substrate binding where the Glu261 residue forms a hydrogen bond with the substrate [201].

Important to note is the four-residue deletion that is highlighted in plasmodial sequences (between 146-149 residue position: *P. falciparum* numbering) compared to other non-apicomplexan species. This deletion is located on the *cd2* helix that forms the Qo active site. Previous studies report that the four-residue deletion is likely to alter the Qo active site pocket causing a 0.83 Å structural displacement [107]. The residues in purple were reported to be 100 % conserved, forming interactions with the Qo substrate. The residues in mustard colour represent residue positions that were identical between plasmodial protein sequences and the human homolog. Residues that were different comparing plasmodial and human homolog sequences were highlighted in green colour. This observation confirms sequence differences between *P. falciparum* and human cytochrome *b* that could be explored in drug design studies. Lastly, residue position Tyr268 (*P. falciparum* numbering) experiences point mutations substituting Tyr (Y) residue to either Cys, Asn or Ser. Such substitutions due to mutations have been implicated in affecting either substrate binding affinity or the protein's overall catalytic efficiency [211–213]. Also, on protein conservation, the *P. falciparum* and human *cytb_{c1}* proteins share 76 % similarity and 43 % sequence identity, showing that *cyt b* protein has some degree of conservation between the two species. Additionally, the identified sequence and structural differences could contribute to identifying selective inhibitors which will possibly serve as the best alternatives to the antimalarials already facing resistance.

3.3.6 Evolutionary analysis using phylogenetic trees

Phylogeny has been useful in understanding evolutionary relations in organisms. There are various methods used in phylogenetic tree calculation including distance-based (UPGMA, neighbour-joining) and character-based (maximum parsimony, maximum likelihood (ML) and Bayesian) methods [214]. Out of these methods, ML methods are usually exceptional because they are based on explicit evolutionary models where tree estimation is calculated using

statistical theory of evolution in calculating the probabilities of tree topologies. Such estimated trees have likelihood scores on them that boost confidence in the estimated tree. As such, bootstrap confidence levels for phylogenetic trees have been indispensable in assessing the accuracy of the statistical score generated [215] and have been shown to be unbiased [216]. There are multiple programs that implement ML methods, including RAxML, MEGA, Mesquite, Phargorn, PHYLIP, etc. A special focus on RAxML is due to its ability to infer evolutionary relationships of large datasets using ML [94,217]. It uses amino acid models such as MTREV, WAG, BLOSUM62, JTT and LG implementing the GAMMA model of rate heterogeneity [94].

A total of five ML rooted trees were generated including best tree, bootstrap tree, bipartitions tree, infotree and consensus tree. The best tree has an overall score across both branches and nodes, while the bipartitions trees have values on the nodes of a tree using the best tree. Bootstrap trees were generated after 1000 rapid bootstraps inferences were implemented. Infotree contained information/labels for each of the organisms aligned, and consensus tree presented trees with confidence values that guide the branching at the nodes. All these trees were generated in the five amino acid models (MTREV, JTT, RTREV, LGGI and WAG). Out of all trees, WAGGF (WAG + GAMMA + F) model produced the best tree with the highest bootstrap values.

Figure 3.10 shows that the clustering on the 74 cyt *b* protein sequences (entire dataset; MSA shown in **Figure S3.2**) and all versus all sequence identity were similar. The character-based tree (**Figure 3.10A**) showed all plasmodial species (those that infect humans and those that don't) and larger apicomplexan species clustered together as shown in red colour. These grouped together as they shared sequence identity above 70 % (**Figure 3.10B**). This indicates that the larger apicomplexan species are closely related. The rest of the apicomplexan species clustered together as highlighted in blue colour, while the bacteria, *Plasmodium* vectors and

human homolog branched and grouped together (mustard colour). Finally, the *Schistosoma* spp. grouped as indicated by cyan colour. In general, the clustering of the entire dataset based on bootstrap values branched with minimal difference in bootstraps, indicating the overall high conservation of the cyt *b* protein. Despite the high conservation of the protein, the *Plasmodium* spp. were observed to be located far from the human homolog, suggesting there could be some differences in these proteins.

Within all groups that clustered together, it was noted that parasite vectors and human sequences grouped closely together while the rest of protozoan sequences formed their own cluster. It was further noted that those sequences that had < 41 % sequence identity clustered the furthest in the tree, showing their distant relationship as compared to *Plasmodium* spp. These included the *Schistosoma* spp. (*S. japonicum*, *S. mansoni*, *S. haematobium* and *S. curassoni*) and *Trypanosoma* spp. (*T. rangeli* and *T. vivax*).

P. falciparum and human homolog were compared, and it was noted that although they clustered distinctly, there was some degree of sequence conservation as shown by MSA (**Figure 3.10A**), indicating that some differences were likely to exist between the two organisms' protein sequences. Previous studies have reported that *Plasmodium* spp. (specifically *P. knowlesi*) mitochondria use a different homolog of coenzymeQ (ubiquinone) than the human host [218,219]. This is also in agreement with the previous section on differences between the *P. falciparum* and human homolog.

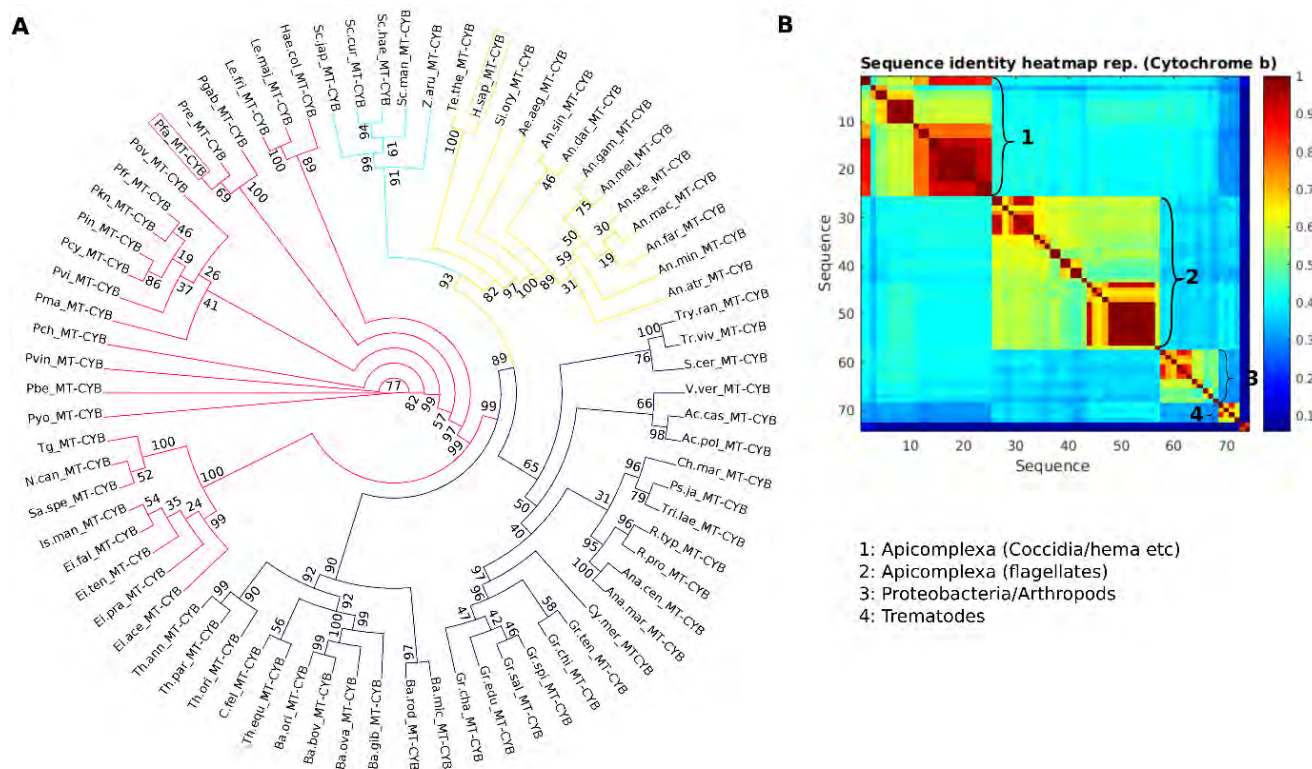


Figure 3.10: (A) Circular phylogenetic tree of cytb protein from 74 species calculated using RAxML software and visualised using Dendroscope. The maximum likelihood (ML) method was used to construct the tree. *P. falciparum* and human cytb are shown in red and chartreuse green colour, respectively. (B) A heat map showing all versus all pairwise sequence analysis where the retrieved cytb organisms grouped together based on their relatedness.

To determine sequence conservation, **Figure 3.10B** shows a heat map representation of all versus all pairwise sequence identity (sequences are displayed by both x and y axis) comparison of all the 74 sequences showing the degree of sequence conservation. The heat map was calculated using an in-house python script using *imagesc* function in Matlab software [101]. As previously published, sequence conservation was represented scaling from 0 to 1 indicating low (coloured blue) and high (coloured in red) conservation respectively [101]. According to **Figure 3.10B**, all the sequences were grouped into four main groups. Group 1 belonged to apicomplexans from coccidia/Hema sub-group including *Plasmodium* spp., *Theileria* spp., *Toxoplasma* spp. and *Isospora* spp. Group 2 included more apicomplexans, specifically flagellates which included *Babesia* spp. Group 3 included the arthropods and proteobacteria

such as malaria vectors (*Anopheles* spp.), *thermus thermophilus* etc. Finally, Group 4 included trematodes such as *Schistosoma* spp. The groupings were based on sequence per sequence similarity and agreed with those of phylogenetic tree clustering.

3.3.7 Motif analysis and structural mapping of the conserved and unique motifs

Comparative motif analysis was performed on a total of 74 protein sequences. **Figure 3.11** shows a total of 35 significant motifs of cytochrome *b* proteins from *Plasmodium*, apicomplexan, bacteria, insect, and human homologs. All the identified motifs ($n = 35$) were shown in a combined block of diagram for easy motif comparative analysis.

These motifs were shown based on their location in the protein sequence referred to as ‘site’, showing all the sites where each motif was located. All the significant motifs were analyzed in all 74 sequences after which the motifs sequence location was visualized in three groups for good visibility; **Figure 3.11a** ($n = 24$), **Figure 3.11b** ($n = 24$) and **Figure 3.11c** ($n = 26$). **Figure 3.11a** shows the first set which outlined motifs found in a *Plasmodium* spp. (that infects humans only) and the anopheles’ species, the primary vector for malaria disease caused by *Plasmodium* spp.

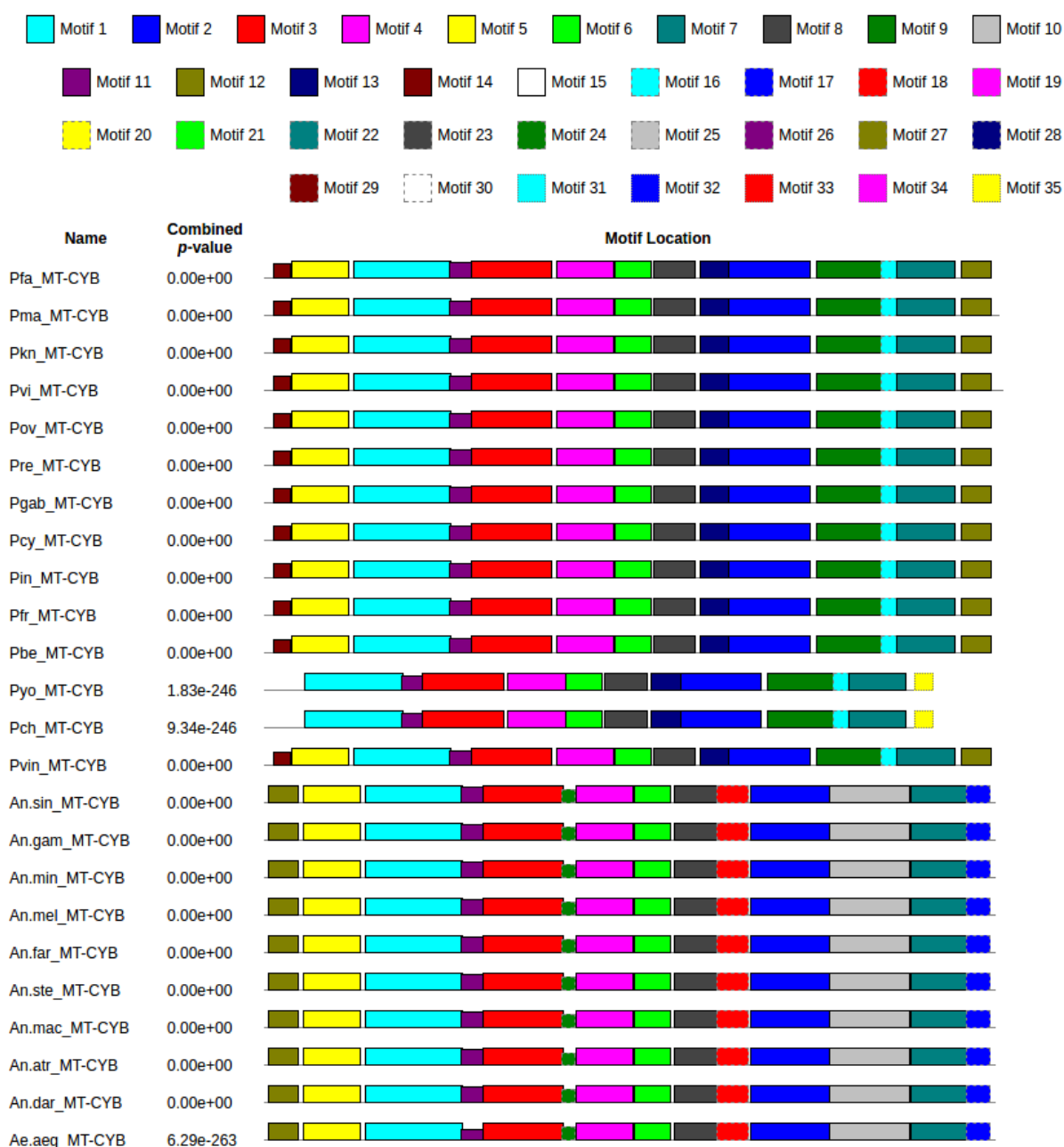


Figure 3.11a: MEME motifs of cytb proteins in the first set of protein sequences; 24 out of a total of 74 protein sequences/species (including *Plasmodium* spp. and parasite vector/insect). The motifs in boxes are highlighted in different colours following the key to differentiate them.

The second set shown in **3.11b** comprises apicomplexan and bacteria spp., while **3.11c** shows more apicomplexan spp., bacteria, and human homolog. Altogether, comparative analysis of these 74 protein sequences uncovered conserved motifs and highlighted each protein sequence having a unique motif.

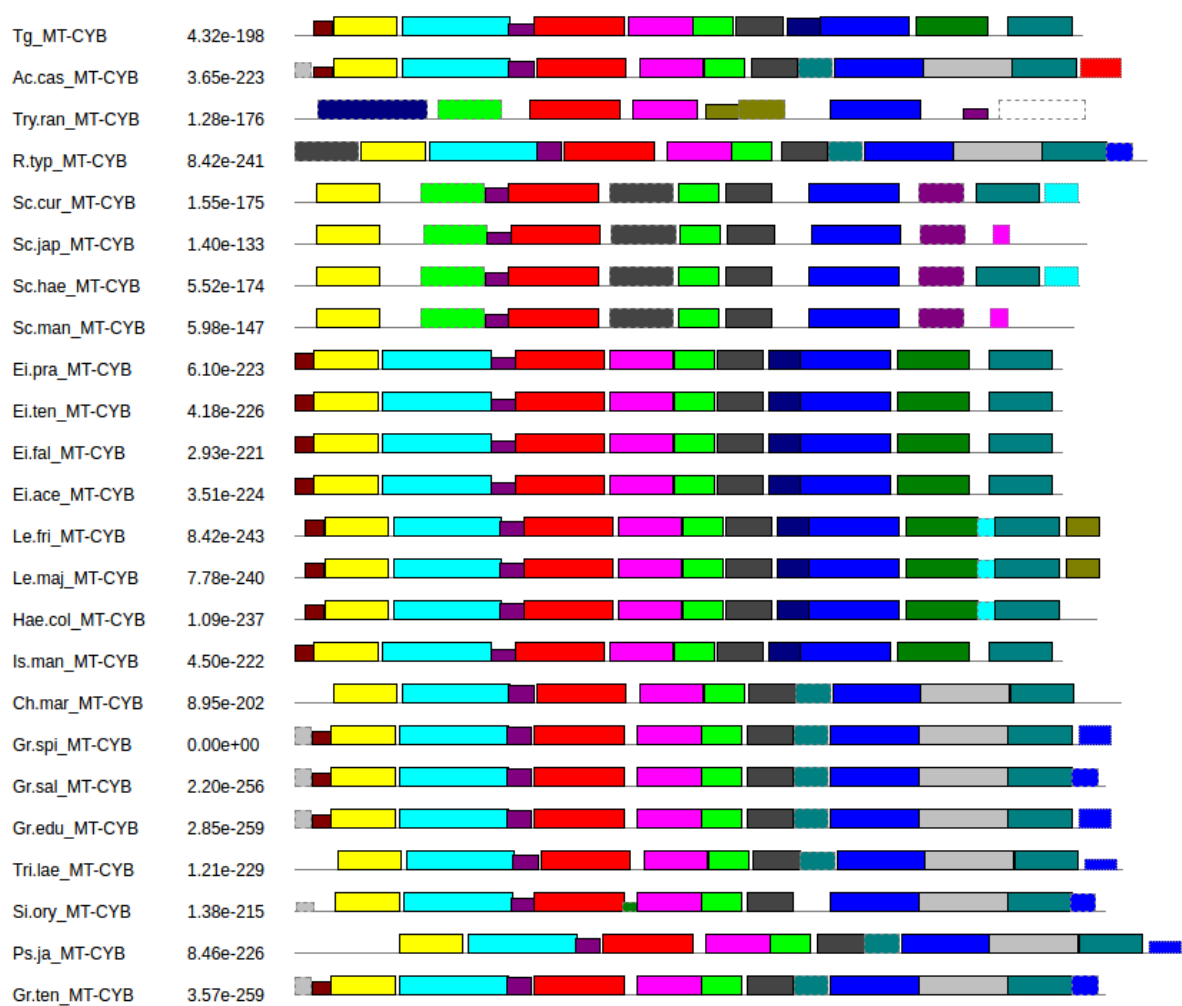


Figure 3.11b: MEME motifs of cytb proteins in the second set of protein sequences; 24 out of a total of 74 protein sequences/species (including apicomplexan spp. and bacteria). The motifs in boxes are highlighted in different colours following the key to differentiate them.

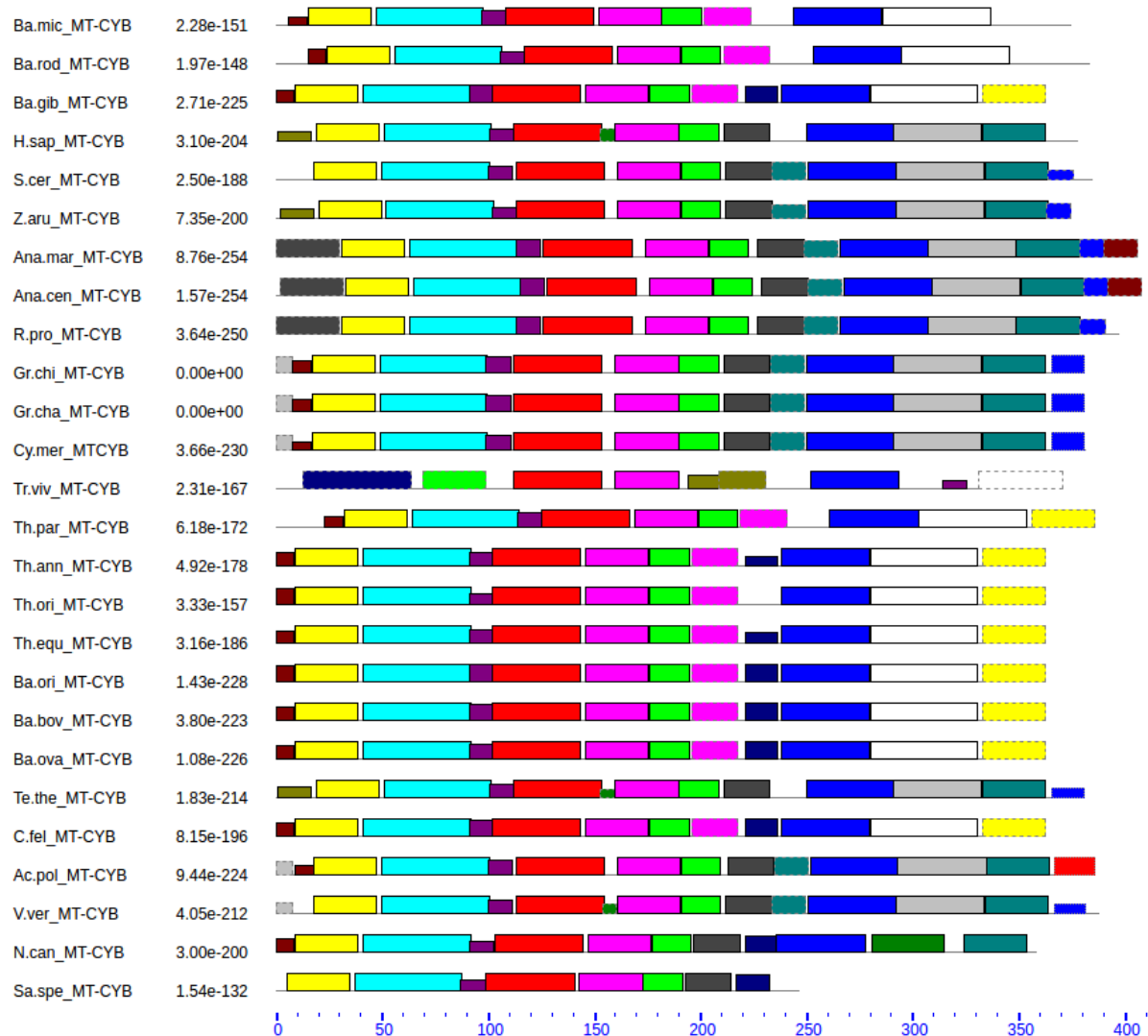


Figure 3.11c: MEME motifs of cytb proteins in the third set of protein sequences; 26 out of a total of 74 protein sequences/species (including apicomplexan spp., bacteria, and human homolog). The motifs in boxes are highlighted in different colours following the key to differentiate them.

All the 35 motifs were subjected to motif alignment and search tool (MAST) to calculate a similarity pairwise matrix to identify any overlapping motifs. Any motif that scored > 0.60 in the pairwise similarity matrix score was eliminated. This is because such motifs were found to be overlapping with other motifs, hence they were not significant for further analysis. In this case, Motif 35, located in two sites only, reported a similarity score of 0.64 when compared to

Motif 12. Therefore, it was eliminated, leaving a total of 34 motifs that were identified as significant motifs for comparative analysis with those found in the human homolog.

Motifs identified were grouped into two: conserved and unique motifs. Conserved motifs were identified as the protein sequence patterns in a group of related sequences. Firstly, based on the number of motif sites/total number of sequences, MEME software identified these conserved motifs following their position/location and then provided the probability of each letter occurring in a particular pattern. Across 74 protein sequences, a total of 9 conserved motifs were identified, as shown in **Table 3.4** and **Figure 3.12**. These conserved motifs, including Motifs 1-8 and Motif 11, were present in over 90 % of the 74 sequences analysed and occurred in similar locations, indicating their functional importance. The motifs were mapped onto *P. falciparum* (PfCytb) and its human host (HsCytb) models to assess motif conservation as shown in **Figure 3.12**.

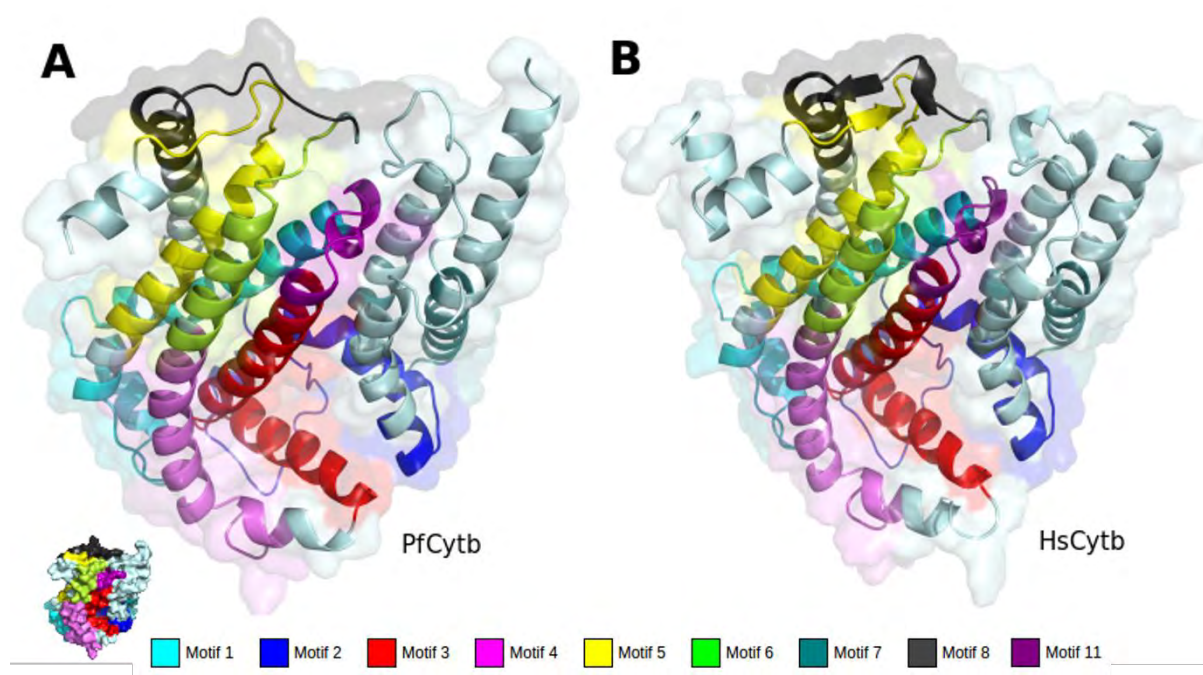


Figure 3.12: Structural mapping of nine conserved motifs onto the (A) *P. falciparum* and (B) human homolog cytochrome *b* protein models. Most of these motifs are in the protein's Qo active site. A colour scale representing conserved motifs is provided in the figure.

The motifs mapped onto similar locations in both models, indicating their degree of conservation. Of note is the Glu residue in the PEWY conserved motif (also known as Qo motif) [149], located in Motif 2, which has been reported to be catalytically important in forming hydrogen bonds in the Qo active site which in turn stabilise the substrate [62], thus supporting ubiquinol oxidation process [220].

Table 3.4: Conserved cytb motif sequences (n = 9) identified across 74 protein sequences using MEME suite [180].

Motif	E-value	Sites	Width	Motif location	Sequence motif
1	2.3e-2221	68	50	47 – 96	A[MS][HR]YT[PA][EDH][IV][SE]YA[FY][YAN]S[VI][QE]HI[LM]R[DE][VL][NW][SYN]GW[LCF][FL]RY[LM]H[AS][NT]GAS[FL][FV]F[ILF][CLV][LTV][YF][LI]H[IV][LFG]RG
2	1.2e-2192	73	41	241 – 281	[GS][HD]P[DE]N[YASF]IP[AV]N[PT][LY]VTPL[HQ]IVPEWYFLPFYA[IM]L[KR][ST]IP[SN]K[LT][GA]G[LV]
3	2.6e-2158	74	41	108 – 148	WI[ST]G[VL][IV]I[FL][FLI][LI][MFST][IM][AV]T[AG]F[LI]GYVLPWGQMS[FY]WGATVITNL[LF]SA[IV]
4	2.5e-1250	70	29	152 – 180	VIW[LI][WCL]GG[FY][ST]V[DS][ND][PA]T[LI][KN]RFF[VS][LF]H[FY][IL][LF]PF[IV][IA]
5	3.6e-1133	70	29	15 – 43	[DSN][YL]P[CTA]P[LS]N[IL][NS][YF]LWN[FY]G[FS]L[LA]G[IL]C[LF][VF][IL]QI[LI]TG
6	5.7e-657	72	18	182 – 199	[AC][LI]V[FV]IH[IL]F[FY]LH[LQ]HGS[TSN]NP
7	1.7e-768	58	29	328 – 356	[AL]L[GLT]WIG[AC][QR][PL][VP][EQ][DE][PI][FY][IV]L[YI]G[QR][LI][FAL][TI][IV]L[YF]FS[YS][FG]
8	6.3e-627	61	21	202 – 222	[YN][SD][TN][AV][DL]K[IV]PF[YH]P[YN][FL][LI][SY][KL]D[VLI][KFL]GF
11	0	73	11	97 – 107	[LV][YN]YS[SY]YY[LY]P[LR][ST]

To investigate how different the conserved motifs are at a residue level, a comparison between *Plasmodium* and human cytb proteins was performed. Here, the individual conserved motifs in one protein were compared to their corresponding motif in the other protein (**Figure S3.3**). The number of differences at the residue level in each motif and the amino acid substitution profile was assessed. The differences at a residue level were reported as follows: 38 % (Motif 1), 48 % (Motif 2), 26 % (Motif 3), 24 % (Motif 4), 51 % (Motif 5), 38 % (Motif 6), 69 % (Motif 7), 71 % (Motif 8) and 81 % (Motif 11). These findings evidently suggest that despite being conserved, significant differences exist where the impact of the accumulating changes would be a matter to investigate. On the other hand, amino acid properties showed the amino acid substitution was either polar uncharged to charged residue, hydrophobic to another hydrophobic residue, polar uncharged to hydrophobic or vice versa. Depending on the location, these amino acid changes could alter active site, support or hinder ligand binding affinity or affect protein interactions if located at proteins interface. For instance, Motif 2 having the PEWY motif has an amino acid substitution close to these residues between *Plasmodium* (P) and human (F) cytb, change from hydrophobic to another hydrophobic residue thus minimal effect on the active site pocket (**Figure S3.3**). As for Motif 3, located in the Qo active site, it reported 26 % differences yet one residue change such as S (big in size) to A (smaller in size) could alter the pocket size and possibly affect ligand binding. Overall, at a residue level there exists some differences however not huge to impact the protein function.

Secondly, out of the 74 protein sequences, comparative motif analysis was performed on motifs present in *P. falciparum* and *H. sapiens* to identify those that were unique in *Plasmodium* for further consecutive analysis (selective inhibitory drug design details further in Chapter 5) (**Figure 3.13**). Motif discovery and mapping (PyMOL script) of these motifs was prepared by an in-house python script according to previous studies [78,80,101,221]. Using MAST and MEME files, a heat map was generated by the script which selects significant motifs and orders

them based on their level of significance. The level of motif conservation scaling from 0 (blue) to 1 (red) was showed on the colour code representing low to high conservation, respectively. A total of five motifs were found to be unique to *P. falciparum cytb* protein sequence. These unique motifs include Motifs 9, 12-14 and 16, as indicated in **Figure 3.13A**. These were further mapped onto PfCytb modelled structure (**Figure 3.13B**). The mapping script obtains the residues that make up the motifs of interest, in this case unique motifs, and all together as a motif, it applies colour codes on the motifs and maps them onto the protein structure. This was performed to determine whether the motifs were in regions of functional importance.

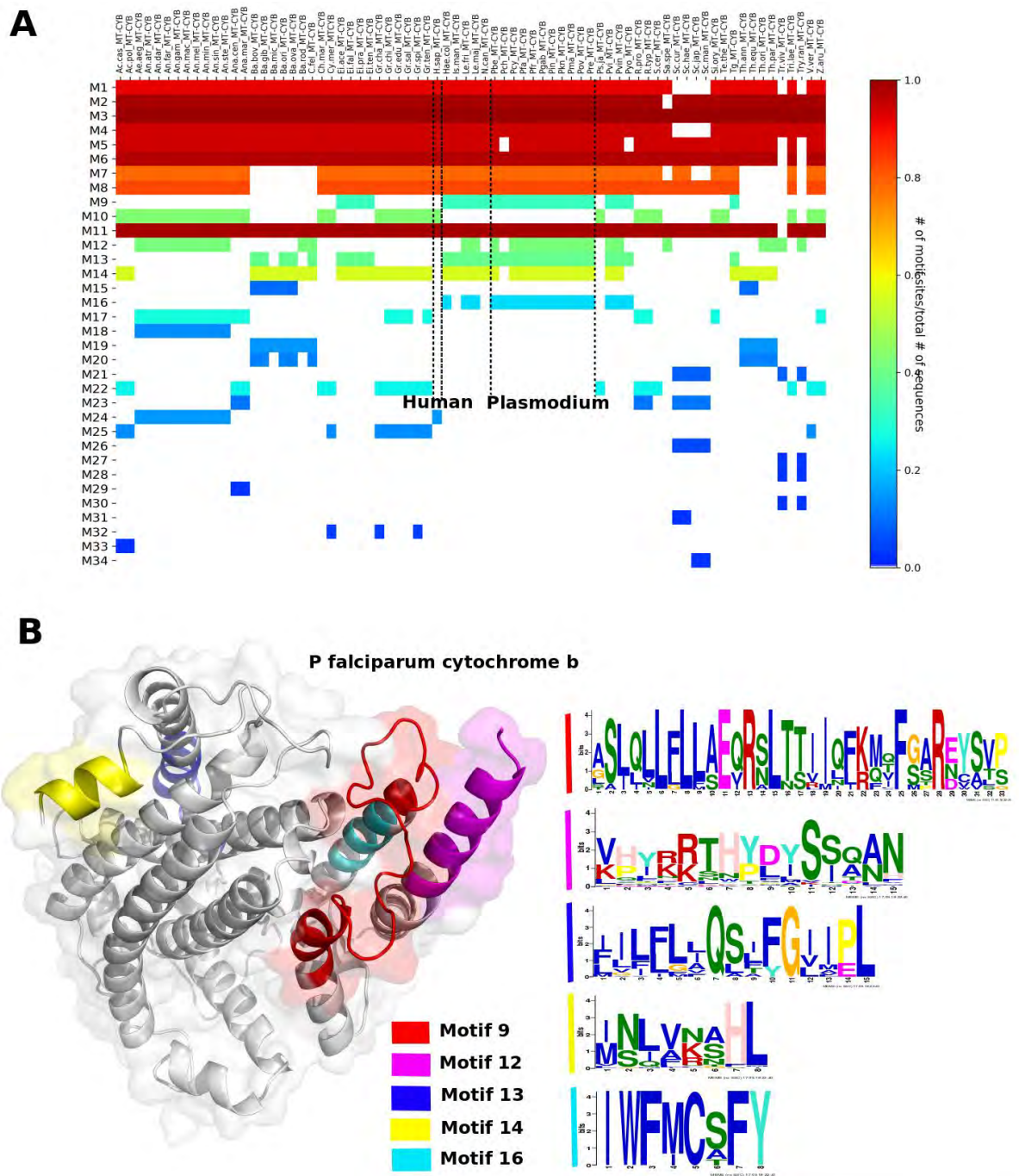


Figure 3.13: Motif analysis of 74 cytochrome *b* protein sequences including Plasmodial and human homolog. (A) Heat map representation of identified motifs based on their occurrence and conservation (number of motif sites/total number of sequences); dotted lines showed the *Plasmodium* and human homolog where the unique motifs in *Plasmodium* are absent in HsCytb (B) Shows structural mapping of the unique *P. falciparum* motifs onto the modelled PfCytb protein (wild-type). Some of these motifs are near the active site. The unique patterns are indicated alongside the protein structure.

Since the function of these unique motifs is unknown, the location of all five unique motifs was investigated to gain insights. Out of the five motifs, Motifs 9 (286-318), 12 (361-375), 14

(6-13) and 16 (320-327) were found to be located distal to both Qo and Qi active sites (*P. falciparum* numbering). Unlike these motifs, Motif 13 (226-240) is located near the Qi active site with a possibility of forming interactions with the ubiquinol substrate. These observations highlight the existence of significant differences between *P. falciparum* and human Cytb that can be explored in selective inhibitory studies.

3.4 Chapter conclusion

A comparative sequence and structure analysis of the cytochrome *b* protein was performed by studying various organisms (n = 74), analysing a sample dataset from the 74 sequences (n = 10), and finally focusing on *P. falciparum* and human homolog. Although cytochrome *b* protein sequence is highly conserved, this study sought to determine whether there were any differences between *P. falciparum* and human homolog at the sequence and structural level. Using computational tools, cytochrome *b* protein sequences were retrieved and analysed for significant differences. Analyses from physico-chemical properties, MSA and phylogenetics highlighted similar profiles and an overall high degree of conservation shared among the proteins. Specific to note is MSA analysis that identified several sequence differences between the two where a four-residue deletion was highlighted in *Plasmodium* spp. This deletion is a distinct feature that was previously reported to affect the Qo substrate binding site. It does that by shifting the pocket as compared to human homolog, thus there is some likelihood that compounds selective to *P. falciparum* could be identified in future drug design studies. Also, five unique motifs were identified in *P. falciparum* that distinguished it from human homolog. Even though their function is still unknown, the motifs' location was investigated. It was noted that four (Motif 9, 12, 14 and 16) out of the five motifs were located farther away from Qo and

Qi active sites, while Motif 13 was near Qi active site. This possibly suggests that Motif 13 could possibly participate in electron transfer process, an idea that is yet to be confirmed.

These findings highlight the existence of key differences between *P. falciparum* Cytb and the human host. As such, these differences could be explored in selective drug inhibitory studies to screen for compounds that could possibly withstand ATQ resistance.

CHAPTER 4

FE²⁺ METAL PARAMETERISATION

Accurate force field parameters are essential in the modelling and studying of molecular systems, such as in molecular dynamics simulations. Plasmodium falciparum cytochrome b and iron sulphur protein (PfCytb-ISP) complex have metal cofactors (heme groups and 2FE-2S cluster), coordinated by iron (Fe²⁺) metal, which are responsible for redox processes in the protein. Applying Fe²⁺ parameters at these metal centres is essential to defining all factors in these molecular systems, thus allowing an accurate study of molecular dynamics. This chapter gives a description of how Fe²⁺ metal ion force field parameters were derived and their validation in the PfCytb-ISP protein complex.

4.1 Introduction

Force fields define a set of parameters for interacting atoms, bonds, angles, and non-bonded interactions. They are essential in studying the behaviour of molecules in an entire molecular system [222]. Force field parameterisation calculates parameters such as the potential energy of a system of atoms in studies such as molecular dynamics (MD) simulations. The parameterisation process offers accuracy not only in *in silico* drug design studies but also other biomolecular studies. Specifically, proteins having catalytic metal ions, often referred to as metalloproteins, have since incorporated force field parameters in their modelled systems. These metal ions serve as coordination centres to support functional processes such as electron transfer (ET), protein structural maintenance, etc. For instance, Fe^{2+} metal is charged and is involved in the oxidation and reduction process, which ultimately is significant in cellular respiration [223].

Various methods are used in generating force field parameters based on quantum mechanics (QM), molecular mechanics (MM) or QM/MM, a combination of the previous two methods. Quantum mechanical (QM) approaches, based on wave functions, remain accurate in evaluating biomolecules' structure and thermodynamic properties. However, they are considered computationally expensive and slow, especially for large systems. As a result, the simulations are mostly combined with classical molecular mechanics (MM) energy functions. The MM energy functions utilise empirical parameters that are often derived from high-level QM calculations. Within the scope of the studied literature, three model approaches have been developed for metal ion parameterisation [224]. These comprise non-bonded, semi-bonded, and bonded. The non-bonded model describes the coordination environment of metals by their electrostatic and van der Waals (vdW) interactions [225,226]. The semi-bonded model involves the placement of dummy atoms around the metal. These atoms are covalently bound to the

metal, and the bonding is determined by the predefined coordination site geometry [227]. However, these two models are limited by the long-range nature of the electrostatic forces and could result in the metal escaping from its coordination centre or having anomalous geometry [224,228]. On the other hand, the bonded model overcomes the limitations of the other approaches as it can preserve the metal in its coordination environment by creating covalent connections to the metal [229].

Metalloenzymes are proteins containing metal ions in prosthetic groups that function to support either catalysis or maintenance of their structure. Specifically, in this study, the iron metals in the *cytb_{c1}* complex participate in redox processes, acting as electron donors or acceptors in the ET pathway [1]. These metals form various interactions with the surrounding residues, including either electrostatic or covalent interactions depending on the metal donor or acceptor atom. Prosthetic groups, also known as cofactors, constitute a non-polypeptide chemical structure and are known to have a biological function in some proteins. The catalytic domain where ET processes occur has a total of four cofactors: two b-type hemes (heme b_H and heme b_L, where subscripts H and L refer to high and low potential respectively) and [2FE-2S] cluster. These catalytic cofactors facilitate this process from the domain through to cytochrome c, a mobile electron carrier in the ETC. Metal binding sites in the cofactors contain Fe²⁺ metals, which act as redox centres where oxidation and reduction processes occur to support the ET process. Specifically, the two b-type hemes are intercalated in the transmembrane helices of the cytb subunit, while the [2FE-2S] cluster is embedded in the extrinsic domain of the ISP chain, between two loops held together by a disulphide bridge [230,231]. The [2FE-2S] cluster, the simplest form of the clusters, is involved in ET [232,233] at the Q_o active site which is ~ 20 Å away from cytb subunit.

Metal parameterisation in heme groups and [2FE-2S] cluster cofactors has previously been developed [234,235]. However, the existing parameters are not transferable, hence the need to

develop new force field parameters suited for this study. Firstly, concerning heme groups, cytochrome P450 is a common protein with b-type hemes in its metal centre. Briefly, these b-type hemes are essential in the P450 catalytic cycle involved in many important processes, including but not limited to oxygen transport and oxidation [236]. The P450 heme has a coordination number of five, assuming square pyramid geometry; that is, an Fe^{2+} metal forming a bond to four planar N atoms on its protoporphyrin ring and a sulphur atom (SG) of the cysteine dipeptide ligand.

On the other hand, parameters for b-type hemes that assume an octahedral coordination geometry having a coordination number of six have been described [237]. In this case, metal ions are coordinated by four nitrogen (N) donor atoms in the hemes' ring conjugated structure and two histidine (His) (through its NE2 atom) residues on each side of the metal. Pertaining [2FE-2S] clusters, various studies have successfully derived metal parameters for most classes of iron-sulphur clusters, including the simple [2FE-2S] cluster [234,238] and the complex [4FE-4S] [239]. The [2FE-2S] clusters have a tetrahedral molecular geometry, where the metal ions are coordinated by four residues: two histidine (His) and two cysteine (Cys) residues as shown in **Figure 4.1**.

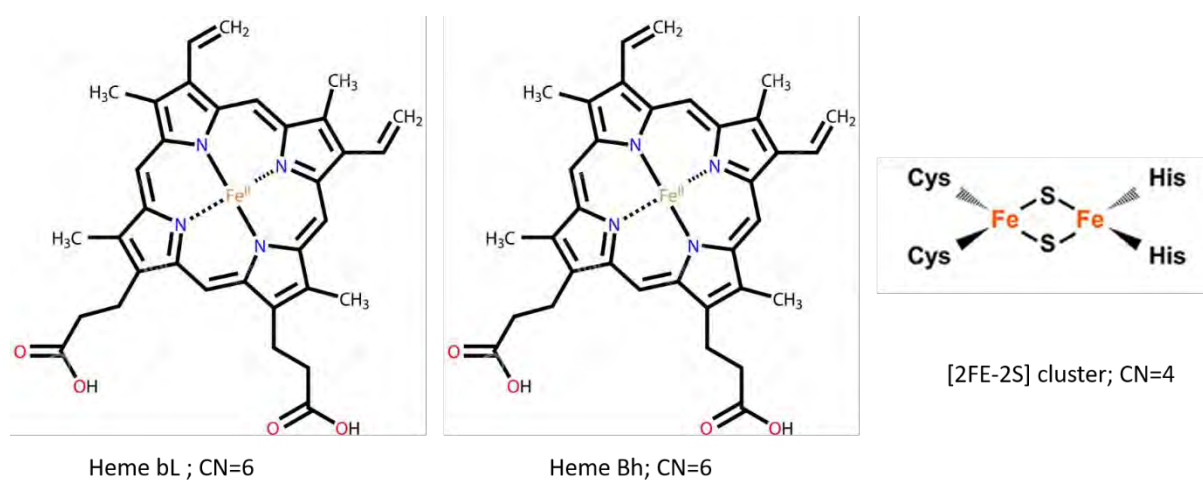


Figure 4.1: Chemical structures of the metal cofactors showing their geometry and residue coordination; heme bL, heme bH and [2FE-2S] cluster. CN is the coordination number.

During protein structure preparation, the protonation process is essential to deciphering its structure and predicting its protonation equilibria (pK_a) [240]. Given that H atoms hold only one electron, identifying their position by x-ray crystallography may prove difficult as they produce a deficient scattering factor. Furthermore, the density distribution around the only electron can either be displaced or pulled away by the dipole-dipole attraction [241]. Therefore, protonation is performed to determine the ionisable groups in the protein, and these are protonated to support proper protein structure and function.

The overall study focused on the PfCytb and ISP subunits because, at the heterodimer interface, the quinol oxidation (Qo) site inhibitor has been reported to bind and form a hydrogen (H) bond with the histidine residue of the ISP chain [201]. This implies that some part of the ISP chain forms a part of the ubiquinol oxidation pocket. Therefore, this chapter derives and validates Fe^{2+} force field parameters for both b-type heme and [2FE-2S] clusters appropriate for PfCytb-ISP complex set up into a phospholipid membrane using the bonded model approach. Using this approach, force field parameters were derived after which MD simulations were performed for validation purposes. Incorporating force field parameters and the phospholipid bilayer in this *in silico* study ensured that the PfCytb-ISP Qo active site was well described for the study of the PfCytb-ISP protein dynamics. It also supports structure-based virtual screening studies, where docking binding modes are seen to be strongly impacted when the calculated RESP charges are incorporated.

4.2 Methodology

4.2.1 Structure preparation; protonation of the protein

For the wild-type (WT) and mutant protein models, the H++ web server [242,243] was used to protonate all non-titratable and titratable residue groups to the correct state [240,244,245]. The titratable groups included residues such as His [246], Asp, Glu, Arg, Lys, Cys and Tyr residues. Each protein contained a total of 111 titratable sites. Protonation for the protein structure was performed at the physiological conditions of a neutral pH of 7.0, salinity of 0.15 M and an external and internal dielectric of 80 and 10, respectively. At the post-protonation step, AmberTools19 software incorporated in H++ was used to generate coordinate and topology files. The topology files for the protonated PDB files were then inspected using Maestro software [247]. It was noted that the H++ server had distorted the orientation of the distal His301 residue that coordinates one of the metal ions within the [2FE-2S] cluster. The His301 residue was manually rotated to bring it closer to and coordinate the Fe^{2+} atom to solve the problem. The resultant structure was then minimised using GROMACS v2018.2 software to relieve steric clashes [248]. Post-minimisation, the protonation states for all other metal coordinating residues were then evaluated for correct protonation and manually adjusted where necessary. This was performed because incorrect protonation states could influence metal ion parameterisation, protein behaviour and dynamics during MD simulation.

4.2.2 DFT Calculations

4.2.2.1 Geometry optimisation

Geometry optimisation for the molecular structures (heme bL and [2FE-2S] cluster) was performed using gaussian 09 software and visualised using GaussView 5.0.9 [249]. This was done to analyse bonded interactions within the chemical structures and check for any coordination or geometry distortion. The optimisation step utilised the B3LYP/6-31G* level

of theory. To confirm the accuracy of the structures, an evaluation of the presence of broken bonds in the structures was performed as their presence could potentially compromise the reliability of consecutive analyses. Specifically, these could affect the reliability of the parameters and their ability to maintain the cofactors within the protein. This was performed by visualising the lowest optimised structure and checking whether all its coordinating bonds were formed within estimated metal bond distances [250]. A total of 10 coordinate bonds were included in the calculations for all interacting atoms and residues for the [2FE-2S] cluster and heme bL cofactors. These included Cys299, His301, Cys317 and His320 for [2FE-2S] cluster and planar N atoms (FE₁-NA, FE₁-NB, FE₁-NC, FE₁-ND), His78 and His173 for heme bL.

4.2.2.2 Force field generation

Following the geometry optimisation process, the generation of force field parameters for the metal cofactors (heme bL and [2FE-2S] cluster) was performed using the WT PfCytb-ISP protein. The Metal Centre Parameter Builder (MCPB) [251], was used to generate input files for Gaussian09 quantum mechanical (QM) calculations. MCPB.py [251] is a script that forms a part of AmberTools19 that allows for generation of force field parameters [179]. The Seminario method incorporated in MCPB was used to derive bond lengths, angles and respective force constants upon ensuring accurate geometry optimisation. A cut-off distance of 2.8 Å was applied to describe all metal bond interactions. The Seminario method calculates these parameters from the submatrices of the Cartesian Hessian matrix [252]. Although Fe²⁺ force field parameters have previously been developed for the heme and [2FE-2S] cluster cofactors, re-parameterisation was performed to allow development of parameters in a format that is simple to use and modify for use in the drug discovery *in silico* studies (Chapter 5).

For the MCPB generated input files, GaussView 5.0.9, a compatible graphical user interface (GUI) and part of Gaussian software packages [249], was used to visualise all files to ensure

that all cofactors and coordination residues factored in for QM calculations were detected. All QM calculations were carried out using B3LYP/6-31G* level of theory, Gaussian09 software [253]. This basis set was applied as it has been shown to produce accurate results using inexpensive computational resources [254]. The 6-31G* is a split-valence double-zeta plus polarisation basis set particularly suitable for organic molecules. The polarisation functions on non-hydrogen atoms are indicated by the * (asterisk). The ** is used when the polarisation is further added to hydrogen atoms, such as in 6-31G**.

The energy profiles were then calculated using AMBER energy function where force field parameters for bonds, angles and dihedrals were factored according to the equation below (Equation 3):

$$V(r^N) = \sum_{bonds}^n k_b(l - l_0)^2 + \sum_{angles}^n k_a(\theta - \theta_0)^2 + \sum_{dihedrals}^n \sum_n^i \frac{1}{2} V_n [1 + \cos(n\omega - \gamma)] \\ + \sum_{j=1}^{N-1} \sum_{i=j+1}^N f_{ij} \left\{ \epsilon_{ij} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\}$$

Equation 3

Where the first three terms (bonds, angles and dihedrals) express harmonic oscillator approximations, while the last term describes the non-bonded interactions. Bond and angle terms l_0 and θ_0 are equilibrium bond lengths and angles, while ka and kb denote the bond stretching and angle bending force constants. In the dihedral torsional term, V_n represents the energy barrier of torsional motion; n , γ and ω stand for periodicity, phase shift and dihedral angle, respectively [126,255].

Heme bL and heme bH cofactors have similar coordination geometry, both forming bonds with two His residues on each side of the metal. The residues that participate in metal bond formation are as follows; heme bH axial ligands are His78 and His173, while heme bL axial ligands are His92 and His187. Therefore, parameter calculations were performed only for the heme bL and [2FE-2S] cluster in the WT protein model, after which the heme parameters were inferred onto the second heme (heme bH). The derived Fe^{2+} force field parameters for the PfCytb-ISP protein were validated using MD simulation and then compared to the *Bos taurus* (PDB ID: 1PP9) crystal structure. Also, the force field parameter set was inferred to all mutant protein models (Y268C, Y268N and Y268S) and ligand systems prior to MD simulations (further detailed in Chapter 5).

All QM calculations were performed at the CHPC (Centre for High Performance Computing) cluster using 240 CPU cores and 15,000 MB of memory.

4.2.2.3 RESP charge calculations

Electrostatic potential (ESP) is one of the many proposed ways for atomic charge derivation using *ab initio* QM calculations as well as the AM1-BCC approach [256,257]. This method is by far the most suitable and is applied in three-dimensional surfaces in many molecules of interest. ESP is best suited for describing inter-molecular properties in condensed phase simulations. Thus, the AMBER force field was utilised for the atomic charge calculations as it factors in molecular electrostatic potential (MEP). In the calculation of partial atomic charges, buried atoms are not well represented in the calculations. To cater for this drawback, some restraints have been applied on these atoms determined by ESP charges which do not significantly affect the overall fit. This approach is thus adopted as restrained ESP (also known as RESP) charge model [258].

Upon ensuring accurate geometry optimisation and force constant calculation, Merz-Kollman restrained electrostatic potential (RESP) charge calculation was performed. This was performed to describe possible electrostatic residue interactions occurring between cofactors and coordination residues. The partial atomic charges were derived from the protein model using the MCPB.py script. Gaussian09 and the hybrid density functional B3LYP with 6-31G* basis set were utilised as previously described [259]. This parameter set combination was applied as it has been shown to produce accurate results using inexpensive computational resources [260]. On this large model, QM was used for calculating the charges using Merz-Kollman RESP fitting algorithm, which factors in all the atoms that interact with the metal coordinating residues [258].

4.2.3 Force field validation by MD simulations

4.2.3.1 POPC: POPE lipid bilayer construction and topology generation

Since PfCytb-ISP is an integral membrane protein, prior to MD simulations, a POPC:POPE membrane of lipid ratio 1.3:1 was constructed, and protein topology was generated using PACKMOL-Memgen [261] as detailed in our previous work [141]. Here, the membrane intercalates the PfCytb-ISP protein acting as the IMM. Thus PACKMOL-Memgen, software integrated within AMBER, was used to build the IMM and insert the protein in order to simulate protein dynamics accurately [262]. The IMM membrane is composed of phosphatidylcholine (PC), phosphatidylethanolamine (PE), phospholipid cardiolipin (CL), Oleoyl (OL), palmitoyl (PA) and phosphatidylinositol (PI) lipid residues [263]. Of these, the lipid residues PA acyl chains as well as PC and PE head groups were available for POPC:POPE membrane construction using PACKMOL-Memgen [261]. PC and PE make up approximately 60–70% of the membrane, depending on the organism, with PC being more

abundant. PA is, however, not present in some organisms. As a result, the proteins were set up to be embedded in a POPC: POPE membrane of lipid ratio 1.3:1, respectively.

The membrane density was set to match the biologically relevant conditions while the simulation box size was adjusted to allow a 15 Å distance. The number of lipids to be constructed was then determined using the area per lipid and the desired lipid leaflet area. PACKMOL-Memgen was also set to solvate the protein-membrane complex using the TIP3P water model [264] to a thickness of 17.5 Å above and below each leaflet. After constructing the protein-membrane complex, protein orientation within the membrane was validated with *Bos taurus* PDB ID: 1PP9 using the Orientations of Proteins in Membranes/Positioning of Proteins in Membranes (OPM/PPM) server [265]. Visual inspection of the embedded protein was also performed to ensure that sufficient membrane lipids were packed between the protein and box boundary to investigate membrane effects accurately. The potential presence of water within the membrane was also investigated.

4.2.3.2 System energy minimisation, heating, equilibration, and MD production run

Topology and coordinate files for MD simulations were generated using LEaP program, where a simulation box size of 105 x 105 x 127 Å was generated and system neutralisation performed. After topology generation, the GPU version of the PMEMD program was used to prepare the system for MD simulation [179]. Here, the AMBER ff14SB and Lipid17 force field were utilised to provide parameters for the protein and lipid residues [266]. Firstly, the protein-membrane system was minimised using a combination of the steepest descent algorithm and the conjugate gradient method for a maximum of 20,000 steps. These included: 10,000 steps of steepest descent minimisation followed by 10,000 steps of conjugate gradient optimisation. This was performed at a non-bonded cut-off distance of 10 Å. After minimisation, temperature equilibration was then performed in two phases where the system was brought up to a

maximum temperature of 310K using Langevin dynamics at a collision frequency of 1.0 per ps and the canonical (constant T) ensemble. In the first heating phase, the system temperature was gradually increased to 310 K over 500 ps. Afterwards, the temperature for the system was maintained at 310 K for a further 500 ps, and then pressure equilibrated to 1.0 bar using anisotropic (x-, y-, z-) pressure scaling and the Berendsen barostat for 1 ns with a pressure relaxation time of 2.0 ps. During temperature and pressure equilibration, the membrane lipid residues numbering between 512 and 1501 were held fixed using a force constant of 10 kcal mol⁻¹ Å². All bonds involving hydrogen were also constrained using the SHAKE algorithm.

Prior to the production run, a hold step was performed to equilibrate the systems periodic boundary dimensions, considering MD was to be performed using the PMEMD GPU code. Each WT and mutant protein membrane system was subjected to 10 separate equilibrations of 500 ps each (5 ns in total) with the system temperature and barostat set to 310 K and 1 bar, respectively. Due to the changing nature of the system's periodic boundary conditions, the skinnb value of the system was set to 5 Å during every restart. Finally, a 350 ns MD production run was performed at the CHPC in Cape Town, South Africa, on 10 CPU cores and one Nvidia Tesla v100 GPU. The system was set to use a 0.002 ps timestep with coordinates written to file every 10 ps.

4.3 Results and discussion

4.3.1 Protonation states for PfCytb-ISP protein metal coordinating residues

Prior to force field calculations, the entire protein was protonated at an optimal pH of 7.0 at which the protein is active and stable. Protonation ensured that there were no misplaced H atoms where the bond formation is expected to form. After protonation, CYS residues that formed a bond with the metal ion within the coordination sphere were deprotonated on sulphur

atom (SG) and labelled as CYM, while those forming disulphide bridges were labelled as CYX. HIS residues that were protonated on the delta and epsilon N atoms (ND1 and NE2 respectively) were subsequently labelled as HID and HIE, respectively. It was also noted that structural restraints were applied to the [2FE-2S] clusters' HIE-301, which ensured that the coordinating residues were within cut-off distance and assumed proper geometry as well. The optimised structure with the lowest potential energy surfaces (reached local minima) for heme and [2FE-2S] cluster was then used for force field calculations.

The pK_a values of the ionisable groups (His and Cys) in the PfCytb-ISP metal centre (only the coordinating residues) are shown in **Table 4.1**.

Table 4.1: Computed pK_a (1/2) values for ionisable groups (side chains) that are coordinated in the metal centre of the PfCytb-ISP protein complex (Internal dielectric of 10). Reproduced from [141].

Deprotonation state(ϵ)		Deprotonation state(δ /SG)	
Site	pK_a values	Site	pK_a values
His78	4.7	His301	<0.0
His92	<0.0	His320	1.2
His173	2.6	Cys299	>12.0
His187	4.5	Cys317	6.6
C-Term	4.7	N-Term	5.1
Total charge at pH 7.0	4	Isoelectric point	8.64

Additionally, the protonation states of all ionisable groups were evaluated and inspected for correct protonation as they constitute essential parts of molecular surfaces, as reported in **Table**

S4.1. The protonation states reported were essential to ensure all interactions between protein residues were correctly formed.

The essence of protonation is that at specific protonation states, the atom is available for bond formation in aid of enzymatic catalysis [267]. Depending on the buried and surface atom groups factored in the calculations, the H⁺⁺ web server reports accurately predicted *pK* values as it employs various physiological conditions in its method; classical continuum electrostatics and statistical mechanics [240,268]. Specifically, the internal and external dielectric physiological conditions used for this protein ensured all kinds of atoms were factored, leading to accurate values that were comparable to the experimental values as previously reported [242].

4.3.2 Membrane and topology construction; MD simulations

Figure 4.2 shows an image of the PfCytb-ISP in the constructed POPC: POPE membrane. Additional details on membrane lipid composition are provided in **Table 4.2**.

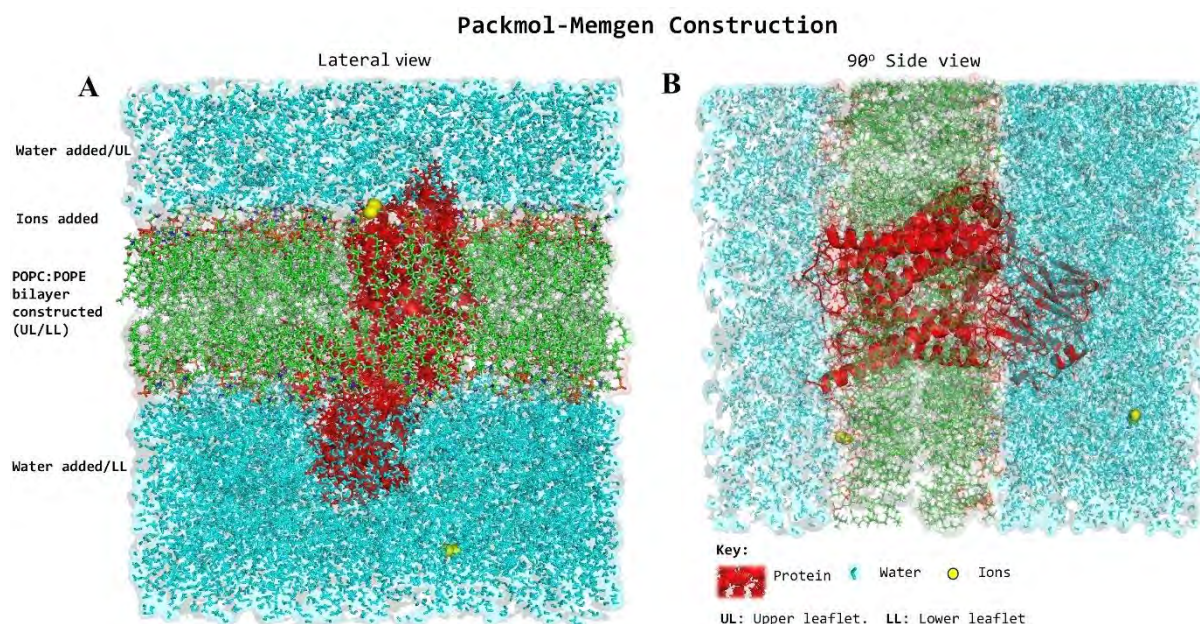


Figure 4.2: Structural representation of membrane construction with PfCytb-ISP protein. The protein is embedded in a POPC: POPE bilayer at a ratio of 1.3:1. (A) Shows a lateral view of

the system, while **(B)** shows the orientation of the protein as seen from the side view. Reproduced from [141].

Lipid residues composed of acyl chains Palmitoyl and Oleoyl accounted for 327 lipids each, while the head groups composed of PC and PE had 185 and 142 lipid residues, respectively. Further details for each system are provided in **Table 4.2**.

Table 4.2: Molecular composition of the simulation systems. Reproduced from [141].

System	Bilayer	Lipids	Water	Counter Ions (added after tleap)	
				Na+	Cl-
Wild-type	POPC:POPE	990	29497	3	4
Y268C	POPC:POPE	984	29574	3	4
Y268N	POPC:POPE	987	29710	2	3
Y268S	POPC:POPE	981	29504	2	3

The constructed protein membrane systems were loaded into the LEaP program for the generation of MD simulation topology and coordinate files. The AMBER ff14SB and Lipid17 force field were selected to provide parameters for the amino acid and lipid residues [266]. The AMBER ff14SB was further extended using the previously calculated metal parameters to incorporate the metal cofactors and coordinating residues. Although PACKMOL-Memgen solvated the protein membrane system, it was not set to neutralise it. Therefore, respective numbers of counter ions were incorporated, as shown in **Table 4.2**, to neutralise the system prior to energy minimisation. Finally, LEaP was set to generate the topology in a simulation box size of $105 \times 105 \times 127$ Å. AMBER topology files produced included: the coordinate (*.inpcrd*) file having coordinates for all atoms in the system, the parameter topology (*.prmtop*)

file containing atomic coordinates, and all the parameters that were used to calculate forces and energies.

4.3.3 Derived force field parameters; Bond distance and angle

The WT protein model was used to calculate force field parameters for the heme bL and [2FE-2S] cluster cofactors. Density functional theory (DFT) (B3LYP/6-31G*) calculations were performed to obtain all force field parameters, including but not limited to bond distance, angle, force constants and RESP charges as shown in AMBER parameter files in **Table S4.2**. Of keen interest is the metal bond distance which was the metric used to validate the derived parameters. A total of 10 coordinate bonds were included in the calculations for all interacting atoms, and residues involved in these bonds included Cys299, His301, Cys317 and His320 for the [2FE-2S] cluster and planar N atoms (NA, NB, NC, and ND), His78 and His173 for the heme bL. **Figure 4.3** shows the chemical structures of the cofactors fitted with the force field parameters.

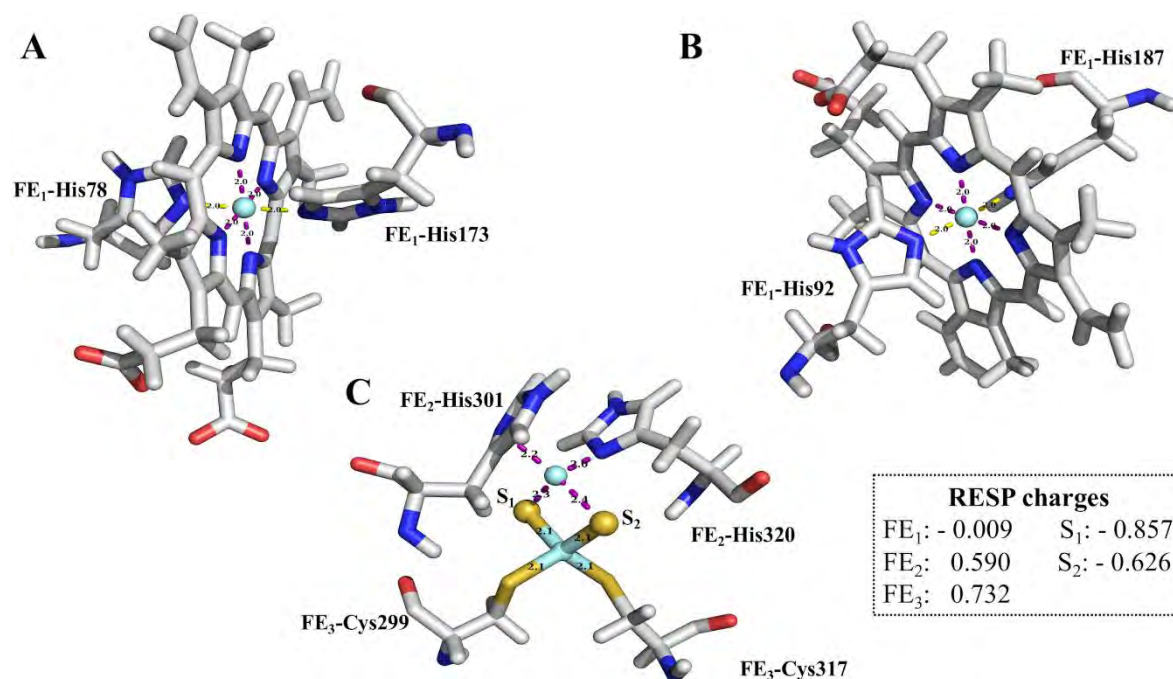


Figure 4.3: Graphical stick representation of the Quantum Mechanics optimised sub-structures of the three prosthetic groups: (A) Heme bL; (B) Heme bH; (C) [2FE-2S] cluster. Fe²⁺ metal,

nitrogen, sulphur, oxygen, and carbon atoms are shown in cyan, blue, mustard, red and grey colour, respectively. The dashed purple lines represent the chemical bonds. Reproduced from [141].

The derived bond distances and associated force constants for heme bL (whose calculations also represent heme bH) and [2FE-2S] cofactors calculated are shown in **Table 4.3**. Comparison of the mean metal and coordination atom distances data for template 1PP9 and QM derived (pre-MD simulation) show similar values, suggesting an accurate metal coordination environment in the heme bL and [2FE-2S] cluster. These results agree with previous studies [250,269–271]. Force constants in the FE₁-NE2 (His78) and FE₁-NE2 (His173) bonds were reported to be 42.7 and 43.0 kcal mol⁻¹ Å⁻² respectively, while FE₃-SG (Cys299) and FE₃-SG (Cys317) bonds had force constants of 131.0 and 118.6 kcal mol⁻¹ Å⁻² respectively. The similar trend observed in energy profiles in these bonds suggests similar binding properties to Fe²⁺.

Table 4.4: A table showing derived metal force field parameters. Comparison of Fe²⁺ bond distance (between Fe²⁺ metal and its interacting residues) in template structure, derived metal parameters and post-MD. (A) Heme bL and (B) [2FE-2S] cluster located in both PfCytb and ISP chains, respectively. Reproduced from [141].

A				
	Bond distance (Å): Heme bL			
	Template (1PP9)	QM derived parameters	Force constants (kcal/ mol⁻¹ Å⁻²)	Post-MD
FE ₁ - NA	1.970	2.0073	55.9	2.0387±0.00591*
FE ₁ - NB	2.032	2.0251	44.2	2.0479±0.00585*
FE ₁ - NC	1.927	2.0276	46.2	2.0745±0.00581*
FE ₁ - ND	2.007	2.0069	64.2	2.0275±0.00583*
FE ₁ -NE2 (His78)	2.005	2.0205	42.7	2.0546±0.00580*
FE ₁ -NE2 (His173)	2.003	2.0146	43.0	2.0831±0.00581*
B				
	Bond distance (Å): [2FE-2S]			
	Template (1PP9)	QM derived parameters	Force constants (kcal/mol⁻¹ Å⁻²)	Post-MD
FE ₃ -SG (Cys299)	2.279	2.1410	131.0	2.1199±0.00573*
FE ₂ -ND1 (His301)	2.139	2.2053	32.8	2.1507±0.00582*
FE ₃ -SG (Cys317)	2.225	2.1544	118.6	2.1234±0.00583*
FE ₂ -ND1 (His320)	2.147	1.9660	74.0	1.9311±0.00576*

*350 ns MD simulation values provided as an average ± 1 standard deviation

The bond angles in the [2FE-2S] cluster involving the sulphur atoms such as S1-Fe₂-S2, S1-Fe₃-S2, Fe₂-S1-Fe₃ or Fe₂-S2-Fe₃ were observed to be consistent with literature [269]. Angles of bonds coordinating the Fe²⁺ in heme bL showed some deviations compared with previous studies [270]. Structural features could explain the varied observation as calculations in this study used a modelled protein, while an x-ray crystal structure (PDB ID: 1YCC; yeast iso-1 cytochrome c) was used in previous studies. Also, this could be accounted for by the different axial ligands in their coordination geometry observed in the different heme groups [272,273]. Despite this observation, the derived values did not cause any geometry distortion during the optimisation process.

4.3.4 RESP (Partial atomic) charges

Restrained electrostatic potential (RESP) charges for the metal ion, sulphur atoms and interacting atoms are shown in **Figure 4.4**, while RESP charges for all coordinating atoms are provided in **Table 4.4**. **Figure 4.4** shows partial atomic charges in the protein model, consisting of not only metal interacting residues (in both heme and [2FE-2S] cluster) but also in the residues that surround the cofactors.

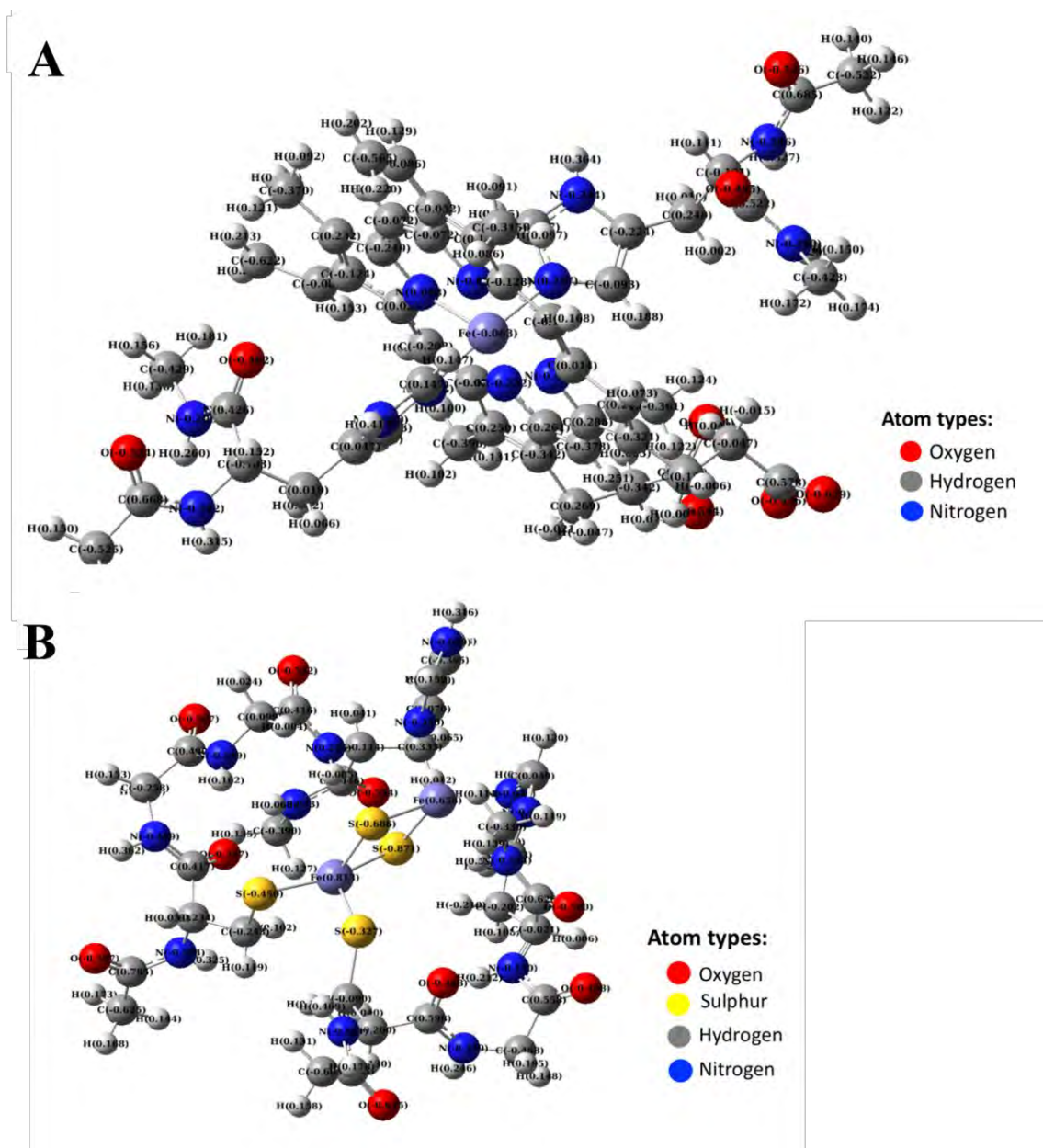


Figure 4.4: A schematic representation of the partial atomic charges (RESP method) as seen in all the atom types in the large models for the b-type heme and [2FE-2S] cluster as shown in (A) and (B) respectively. Colour representation is based on atom types as labelled.

The derived Fe^{2+} metal charges in the [2FE-2S] cluster were reported to have positive values of 0.590 (Fe_2) and 0.732 (Fe_3), which is consistent with previous studies [269]. However, Fe_1 in b-type heme bL was reported to have RESP charges of -0.009 , unlike previous studies that reported the RESP charge for Fe^{2+} to be 0.16 using the HF/6-31G* charge calculation method [274]. Additionally, other studies have reported a charge of 0.45 and 1.58 in a-type and c-type

heme structures, respectively [272,273]. Compared to the latter b-type charge reported, these values were relatively high.

Table 4.4: Partial atomic charges derived before (formal charge) and after charge calculations. The restrained electrostatic potential (RESP) charges were derived using gaussian09 software using B3LYP/6-31G* basic set at DFT level of theory. NB: RESP charges for only the coordinating residues are presented below. Reproduced from [141].

Atoms	Initial atomic Charge	RESP charge
	BEFORE	AFTER
FE1 (Heme bL)	2.0000	-0.0093
FE1 (Heme bH)	2.0000	-0.0093
FE2 [2FE-2S]	2.0000	0.5906
FE3 [2FE-2S]	2.0000	0.7323
SG1 CYS [2FE-2S]	-2.0000	-0.8579
SG2 CYS [2FE-2S]	-2.0000	-0.6260

Heme is axially coordinated by two His residues, while c-type heme is axially coordinated by His and Met residues. Despite this observation, it was noted that the low charges did not destabilise the metal coordination sphere of the entire protein. The relatively low Fe^{2+} charge in this study could be a result of the delocalisation of the electrons, which brings about a stabilising effect [275]. On the other hand, sulphur partial atomic charges reported in this study were similar to previous literature [273].

4.3.5 Validation of the force field parameters using MD simulations

The derived force field parameters were further evaluated via membrane MD simulation using the AMBER software package [179] to investigate whether the cofactors would be maintained within the protein. Additional standard deviation values were provided for the bond distances

representing all sampled points across 350 ns MD simulations. The average bond distances of all the 10 coordinate bonds involving His and Cys residues of the cofactors were 2.06 Å and 2.13 Å, respectively. These values are within the range of the reported experimental mean bond distances of 2.16 Å (FE₁-N) and 2.30 Å (FE₃-SG) [276]. Results evidenced that pre-MD and post-MD bond distances had similar values. In addition, MD trajectory analysis highlighted that the octahedral and tetrahedral coordination geometries of the cofactors were maintained within the protein. The bond lengths monitored during the 350 ns simulation were further presented in **Figure 4.5**.

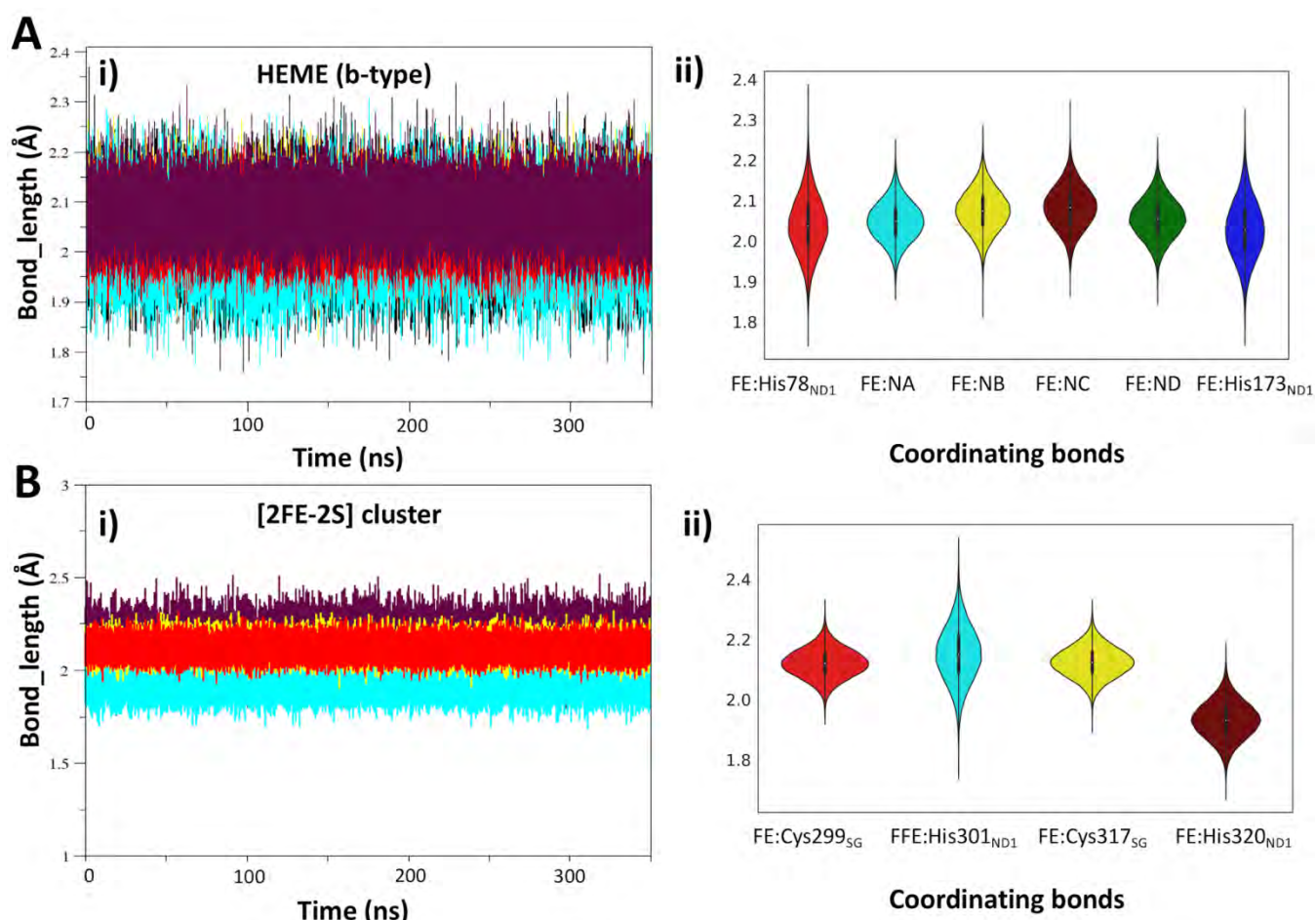


Figure 4.5: Metal bond distances monitored throughout 350 ns MD simulations. (A) and (B) represent i) Heme and ii) [2FE-2S] cofactors, respectively. Metal ion FE stands for FE₁ for the heme and two iron metals (FE₂ and FE₃) for the [2FE-2S] cluster. The various bond distances are also represented as violin plots for the distinction of some of the overlapping bonds present

in the line graph. Fe^{2+} bonds to His78, His173, NA, NB, NC and ND are shown in red, blue, cyan, yellow, maroon and green, respectively for Heme cofactor, while Fe^{2+} bonds to His301, His320, Cys299 and Cys317 are shown in red, cyan, maroon and yellow colours, respectively for [2FE-2S] cluster. Reproduced from [141].

The behaviour of all 10 Fe^{2+} coordinated bond distances was monitored throughout the MD simulations in the WT system and in all the mutant systems as represented in **Figure 4.6**. These bond distances were found to be comparable to those observed in the WT system. Additionally, intra-comparison of all protein systems showed that there was a subtle change in their bond distances. Overall, the bond distances were comparable (p -value = 0.7302) and reliable for use in subsequent MD simulations.

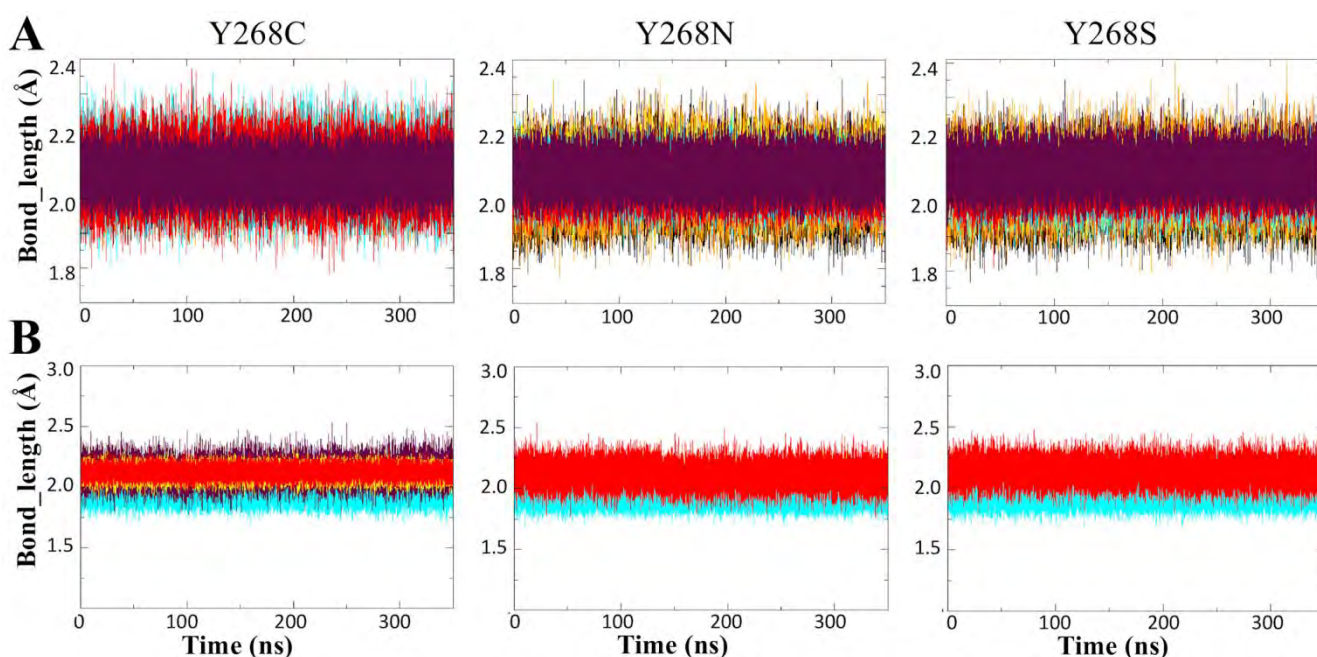


Figure 4.6: Comparison of bond distances monitored during 350 ns MD simulations in the mutant systems (Y268C, Y268N and Y268S). Graphs shown in (A) and (B) represent Heme and [2FE-2S] cluster cofactors, respectively. Fe^{2+} bonds to His78, His173, NA, NB, NC and ND are shown in red, blue, cyan, yellow, maroon and green, respectively for Heme cofactor, while Fe^{2+} bonds to His301, His320, Cys299 and Cys317 are shown in red, cyan, maroon and yellow colours, respectively for [2FE-2S] cluster. Metal parameters were successfully transferred to the systems prior to MD simulation. The bond distances are comparable to the WT system shown in Figure 3.5. Reproduced from [141].

4.4 Chapter conclusion

Reliable MD simulations depend on the use of accurate force field parameters describing the protein molecular dynamics. In the PfCytb-ISP active site, the cofactor Fe^{2+} metal binding sites act as redox centres where ubiquinone substrate oxidation and ubiquinol reduction occur to facilitate electron transfer processes. In this study, accurate force field parameters describing the b-type hemes and [2FE-2S] cluster were derived using AMBER force field and validated through all-atom 350 ns MD simulations. These parameters were found to hold the metal ions in the coordination sphere, as seen in their strong coupling. This depicted that accurate geometry was maintained as the coordinating distances were reported to be within the expected range. Thus, the derived force field parameters were shown to accurately describe and maintain the coordination environment of the heme groups (heme bL and heme bH) and the [2FE-2S] cluster.

The generated force field parameters were essential in providing a more reliable and accurate MD simulation of the plasmodial cytochrome *bc_L* complex catalytic domain. Also, these force field parameters are useful in other metalloenzymes that share similar coordination geometry to the PfCytb-ISP cofactors. Incorporating these force field parameters can promote efficiency in *in silico* studies such as future virtual screening and inhibitory drug design studies (further detailed in Chapter 5).

CHAPTER 5

DOCKING AND MOLECULAR DYNAMICS STUDIES

This chapter describes molecular docking and confirmatory studies such as molecular dynamics which have been useful in structure-based virtual screening to identify optimal compounds. After blind docking experiments, the identified hit compounds were further evaluated based on their binding energies, interactions formed as well as the binding site of interest. Through in silico studies, multiple studies have either targeted the active site pockets to identify conventional compounds or other alternative/potential drug binding pockets. Both methods have previously been used to screen for potential selective compounds to inhibit the canonical active site or for allosteric modulation of the targeted enzyme. Blind docking protocol was performed by screening both plasmodial (PfCytb-ISP) and human homolog (HsCytb-ISP) in search of selective inhibitory and allosteric compounds. In this study, both South African Natural Compound Database (SANCDB) and Medicines for Malaria Venture (MMV) compounds were used in docking, and MD studies were performed to investigate the dynamics and stability of the protein-ligand complexes throughout simulations. Protein stability is assessed whether it will assume its native folded or diminished conformation. Ultimately, the potential hit compounds targeting the PfCytb-ISP Qo active site and heterodimer were identified.

5.1 Introduction

Structure-based virtual screening has been used in drug design to search for novel antimalarial inhibitors as well as to understand drug action mechanisms [76,277,278]. It is important in molecular studies to obtain insights on crucial functional hypotheses. Also, molecular dynamics (MD) simulations have been useful in giving insights into protein dynamics and drug interactions. Targeting *cytb_c* has proven challenging, yet it remains an attractive drug target. One of the challenges includes the use of synthetic drugs that lack structural complexity to withstand parasite resistance. Currently, various drugs are in the pipeline to counter parasitic resistance. These include synthetic drugs such as ozonide OZ439 [279] and 4(1H)-quinolone derivatives like endochin-like quinolones (ELQ)-300 [280,281] and ELQ-400 [282] that have been tested on mouse models. Structure-based virtual screening has been used in drug design to search for novel antimalarial inhibitors as well as to understand drug action mechanisms [76,277,278]. It is important in molecular studies to obtain insights on important functional hypotheses. Also, molecular dynamics (MD) simulations have been useful in giving insights into protein dynamics and drug interactions. Targeting *cytb_c* has proven challenging, yet it remains an attractive drug target. One of the challenges includes the use of synthetic drugs that lack structural complexity to withstand parasite resistance.

Additionally, other studies have focused on developing drugs that act on a single active site and on multiple sites. For instance, ELQ-400 has a dual site inhibition, inhibiting both Qo and Qi active sites of cytochrome *b* [282]. However, the formulation of such a dual-site inhibitor may present a challenge during drug development due to the drug's hydrogen bond acceptors and donors' likelihood to have a similar affinity towards the two active sites. From such studies and more, different approaches and use of natural drug compounds that target *P. falciparum*

cytb_{ci} (PfCytb-ISP) are warranted to counter the alarming ATQ resistance. This can go a long way in eliminating malaria spread through travel.

ATQ is used as a fixed-dose combination therapy with proguanil drug having a synergistic effect as its mode of action [46,283]. ATQ drug administration is country-based, and it is not only used for chemoprophylaxis but also for the treatment of uncomplicated malaria mainly in sub-Saharan Africa and some parts of Asia. ATQ drug resistance has been reported [61,284] and is associated with treatment failures [58,60,285]. Literature reports parasitic mutations to be responsible for ATQ resistance to the long-acting drug in the combination therapy. This was directly linked to ATQ single-drug therapy that caused delayed parasite clearance rates as well as recrudescence parasites [38,286]. Previous studies have shown the point mutations to be located in the ATQ binding site of the cytochrome *b* gene [50,61,63,107]. In general, mutations in *Plasmodium* spp. have since been categorised as those that infect humans (*P. falciparum*) and those that do not, including *Plasmodium berghei* and *Plasmodium yoelii* (*P. berghei* and *P. yoelii*). In *P. falciparum*, an amino acid substitution of tyrosine to serine in the 268th residue of the cytochrome *b* gene of *P. falciparum* cytochrome *bc_i* complex (PfCytb) [63,65] was frequently reported, as well as two other mutations (Y268C and Y268N) [50,60,63]. In total, three point mutations in the same residue position have been described; tyrosine substituted with cysteine (Y268C), asparagine (Y268N), or serine (Y268S), out of which Y268S is the most prevalent [50,60,63]. In *P. berghei*, the point mutations include M133I, L144S, L271V, K272R, Y268C, Y268S, Y268N and V284F [283]. Lastly, point mutations I258M, F267I, Y268C, L271V and K272R have been associated with resistance in *P. yoelii* [287]. All these mutations are known to compromise essential hydrophobic contacts with ATQ thus leading to decreased substrate binding affinity in the active site. The ATQ drug interactions are either reduced or lost due to changes in the active site pocket [62,107].

Previous *in silico* studies on ATQ resistance mechanisms have failed to account for the embedding of the *P. falciparum* cytochrome *b* (PfCytb) protein in a membrane lipid bilayer. In addition, some of these studies have also only investigated the Y268S mutation using the yeast *cytb_{C1}* complex as a homologous model [203,288,289] and not incorporated the other two mutations (Y268C and Y268N). As such, there remains insufficient information describing PfCytb-ISP protein dynamics embedded into a phospholipid bilayer and within the parasite. As a result, the effects of all the reported mutations (Y268C, Y268N and Y268S) on this protein and ATQ drug activity are still not well understood. Thus, it is crucial to mimic *in vivo* conditions by incorporating the protein in a lipid bilayer in analyses such as protein molecular dynamics (MD) simulations to ultimately provide an accurate interpretation of the protein's structure and function. However, several challenges are faced during membrane MD calculations, including setting up lipid membrane residues that are more complex and make up a large portion of the simulation system. Also, the assembly of the lipids to form a membrane as well as embedding the protein within the membrane in the correct orientation is challenging [290,291]. Nevertheless, with the recent development of computational tools, MD simulations of membrane proteins have become more viable in understanding protein membrane dynamics and interactions [262,292,293].

For the first time, to the best of our knowledge, this study presents insights on the effect of mutations on PfCytb-ISP protein as well as on ATQ activity occurring at the Qo active site. Experiments were performed using *in silico* methods on the holo and ATQ-bound protein complexes embedded in monounsaturated 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) and 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE) bilayers via MD. Dynamic residue network (DRN) analysis was used to investigate the effect of these point mutations on protein structure and function, as proposed in previous studies [294,295]. Herein, the ATQ resistance mechanisms were elucidated. Point mutation Y268S was reported to affect

key active site residues, which were observed to lose communication. This was consistent with previous findings [141]. Likewise, similar resistance mechanisms were noted in both the Y268C and Y268N mutant systems. Ultimately, this approach and findings provide insights into the effect of point mutations Y268C, Y268N and Y268S on PfCytb-ISP protein and ATQ activity, hence their importance to future structure-based virtual screening studies and MD simulations.

For structure-based virtual screening, this study utilised both the South African Natural Compounds Database (SANCDB) and Medicines for Malaria Venture (MMV) compounds to identify compounds that selectively target PfCytb-ISP protein. Prior to structure-based virtual screening, the first part of this study ran experiments to understand the effect of mutations on ATQ drug activity and its resistance mechanism as reported in three mutations studied on both holo and ATQ-bound systems. In the second part of the study, molecular docking was performed to identify hit compounds targeting the PfCytb-ISP Qo active site. The compounds that showed good binding energy, were selective to *P. falciparum* protein and were stable in the presence of mutations were considered optimal. Finally, the third part of the study aimed to screen for selective hit compounds that bind on the protein's heterodimer interface pocket and are implicated to have allosteric modulation on the protein. To transcend the challenge multi-site drugs present, the study will focus on targeting the Qo active site and interface pocket of cytochrome *b* separately. Potential compounds with good structural complexity and no cross-resistance with atovaquone-resistant parasite strains are warranted to curb the looming ATQ resistance. Ultimately, these studies will shed more light on the inhibition of PfCytb-ISP protein and the effect of mutations on its function.

5.2 Methodology

5.2.1 Protein modelling; WT and mutant proteins (HsCytb-ISP)

The wild-type (WT) PfCytb-ISP protein was modelled successfully as detailed in Chapter 3. Here, the human homolog (HsCytb-ISP) will be used in selective structure-based screening for optimal compounds. Briefly, to model the heterodimer HsCytb-ISP complex, the Universal Protein Resources (UniProt) [172] was used to access and download human cytochrome *b* (UniProt accession: [P00156](#)) and iron-sulphur protein (ISP) (UniProt accession: [P47985](#)) subunits [172]. Crystal structures of the mentioned protein subunits were available. However, they had a poor resolution, therefore models were constructed using I-TASSER [104,106,173].. A structural template from *Bos taurus* PDB ID: 1PP9 was used as the principal template in guiding the modelling process. This was selected due to its high sequence identity to the human protein where the bovine and human *cytb_{c1}* proteins have been shown to share 79% sequence identity. Full-length atomic models were then generated by iterative template-based fragment assembly simulations [106].

To obtain a complex between the two predicted models (HsCytb and ISP), the individually modelled subunits were separately superimposed onto corresponding chains of homologous *Bos taurus cytb_{c1}* complex (PDB ID: 1PP9) using PyMOL [175] as described previously. The overall structural quality of the HsCytb-ISP complex was further evaluated using z-DOPE score and PROCHECK [177].

5.2.2 Allosteric pocket identification in PfCytb-ISP protein specifically interface pocket

Allostery is a term that describes structural conformational changes as a result of ligand binding at a distal site. To identify the PfCytb-ISP heterodimer interface binding pocket, the protein model structure was submitted in FTMap webserver [296], FTSite [297], Allosite [298] and

PARS [299] for consensus of the interface pockets identified. The predicted interface pocket was visualised using PyMOL to highlight the canonical pockets as well as potential interface pockets. The webserver used many organic molecules as probes to dock onto the whole protein structure, and those that contribute ligand binding free energy were marked to predict cavities in the protein. The interface pocket was explored for potential drug design studies and to understand allosteric and resistance mechanisms. The potential druggable site was selected as it was the only pocket identified at the interface by FTMap, which uses the LIGSITEs program to predict the cavities [300]. Residues within 4 Å were highlighted for further ligand interaction analysis.

5.2.3 Retrieval of drug compounds and Molecular docking approach

The ATQ, MMV and SANCDB 2D chemical structures were retrieved from the PubChem database, MMV website and SANCDB databases [72,73,301,302]. Prior to docking, the drug compound was converted to 3D then protonated to ensure accuracy in docking scores and poses generated. Both protein and ligand structures were prepared using *prepare_receptor4.py* and *prepare_ligand.py* protocols in MGLToolsPckgs of AutoDock4.2 software, respectively [112]. Using the Gasteiger-Huckel method in AutoDock tools, the partial atomic charges of the ligands were assigned. Thereafter, blind docking experiments were performed targeting the Qo site of the PfCytb-ISP and HsCytb-ISP WT protein model using AutoDock Vina v1.1.2 [303]. After docking protocol was optimized and established, the parameters used were as follows: a cubic box of grid points 48, 47 and 54 along x, y, z dimensions with a grid spacing of 1 Å, energy range of 4, cpu of 8 and exhaustiveness of 360. For the docked results, an RMSD cut-off of 2 Å was utilised to obtain accurate predictions of the bound structures.

A total of 10 orientations were generated for all the docked compounds, out of which ranking was performed based on binding energy and RMSD value. The ligand conformation with the lowest docking energy score and RMSD was selected. Thereafter, the docked ligand coordinates from the WT model were transferred to the mutant proteins (Y268C, Y268N and Y268S) for modelling MD simulations.

5.2.4 Molecular dynamics simulations and trajectory analysis

5.2.4.1 POPC: POPE Lipid Bilayer Construction and Topology Generation

Since PfCytb-ISP is an integral membrane protein, prior to MD simulations, a POPC: POPE membrane of lipid ratio 1.3:1 was constructed (membrane set-up designed by Dr Allan Sanyanga) [141] and protein topology was generated using PACKMOL-Memgen [261]. Here, the PfCytb-ISP protein intercalates the phospholipid membrane acting as the IMM. Thus, PACKMOL-Memgen software integrated within AMBER was used to build the IMM and insert the protein in order to simulate protein dynamics accurately [262]. The IMM membrane is composed of PC, PE, CL, PA and PI [263]. Of these, lipid residues Palmitoyl (PA) and Oleoyl (OL) acyl chains as well as Phosphatidylcholine (PC) and Phosphatidylethanolamine (PE) head groups. PC, PE and PA lipid residues were available for membrane construction using PACKMOL-Memgen [261]. PC and PE make up approximately 60 - 70% of the membrane each, depending on the organism, with PC being more abundant. PA is, however, not present in some organisms. As a result, the proteins were set up to be embedded in a POPC: POPE membrane of lipid ratio 1.3:1, respectively.

The membrane density was set to match the biologically relevant conditions while the simulation box size was adjusted to allow a 15 Å distance. The number of lipids to be constructed was then determined using the area per lipid and the desired lipid leaflet area.

PACKMOL-Memgen was also set to solvate the protein-membrane complex using the TIP3P water model [264] to a thickness of 17.5 Å above and below each leaflet. After constructing the protein-membrane complex, protein orientation within the membrane was validated with *Bos taurus* PDB ID: 1PP9 using the orientations of proteins in membranes/positioning of proteins in membranes OPM/PPM server [265]. Visual inspection of the embedded protein was done to ensure that sufficient membrane lipids were packed between the protein and box boundary to investigate membrane effects. The potential presence of water within the membrane was also investigated.

Thereafter, topology and coordinate files (protein embedded in a membrane) for MD simulations were generated using LEaP program, where a simulation box size of 105 x 105 x 127 Å was generated and system neutralisation performed. Here, the AMBER ff14SB and Lipid17 force field were utilised to provide parameters for the protein and lipid residues [266].

5.2.4.2 Molecular dynamics production runs

To explore the protein-ligand dynamics of the Qo active site and interface selective compounds on the PfCytb-ISP WT and mutant proteins, the GPU version of the PMEMD program was used to prepare the systems for MD simulation [179].

First and foremost, the membrane protein-ligand systems were minimised using a combination of 10,000 steps of steepest descent algorithm minimisation, followed by 10,000 steps of conjugate gradient optimisation. This was performed at a non-bonded cut-off distance of 10 Å. Temperature equilibration was then performed in two heating phases after energy minimisation. In this step, the system was brought up to a maximum temperature of 310K using Langevin dynamics at a collision frequency of 1.0 per ps and the canonical (constant T) ensemble. In the first heating phase, the system temperature was gradually brought up to 310K over 500 ps. The system's temperature was then maintained at 310K for a further 500ps in the

second heating phase. Thereafter, the pressure was equilibrated to 1.0 bar using anisotropic (x-, y-, z-) pressure scaling and the Berendsen barostat for 1 ns with a pressure relaxation time of 2.0 ps. Membrane lipid residues ranging between 513 and 1493 were held fixed using a force constant of 10 kcal/mol⁻¹ Å² during temperature and pressure equilibration steps. Using the SHAKE algorithm, all bonds involving hydrogen were also constrained.

Considering MD was to be performed using the PMEMD GPU code, a hold step was performed to equilibrate the system's periodic boundary dimensions prior to the production run. Each WT and mutant membrane protein-ligand system was subjected to 10 separate equilibrations of 500 ps each (5 ns in total). During equilibration, the system temperature and barostat were set to 310K and 1 bar, respectively. Due to the changing nature of the system's periodic boundary conditions, the skinnb value of the system was set to 5 Å during every restart. Finally, a 350 ns MD production run was performed at the Center for High-Performance Computing (CHPC) in Cape Town, South Africa on 10 CPU cores and one Nvidia Tesla v100 GPU. The system was set to use a 0.002 ps timestep with a coordinate file written every 10 ps.

The MD trajectory was visualised using VMD [304] and post processed with CPPTRAJ [305] to remove the solvent, membrane lipid residues and counter ions from the membrane protein-ligand complex. The CPPTRAJ program was further used to calculate root mean squared deviation (RMSD), root mean square fluctuations (RMSF), radius of gyration (Rg), and the centre of mass (COM) distance. All statistical tests were performed using Mann-Whitney U-test using R v3.6.1 where a p value of < 0.05 was considered significant.

5.2.5 Hydrogen bond calculations

Intramolecular hydrogen bonds are important electrostatic forces in molecular systems as they enhance structural and ligand stability [306,307]. The general rule of thumb in hydrogen bond

formation is such that a bond is formed between a donor and an acceptor atom such as F, O and N atoms which can be donors or can be acceptors.

Using CPPTRAJ commands, hydrogen bonds formed by the ligands were calculated using a distance of 3 Å and an angle cut-off 135° geometric considerations [305]. Hydrogen bond occupancy was represented in decimal values at the fraction column. The fraction column tabulates the proportion of MD simulation time a hydrogen bond was present across an entire 350 ns MD trajectory. The hydrogen bonds observed <10% of MD simulation time were considered short-lived and thus not significant in stabilising the ligand in the active site.

5.2.6 Dynamic residue network (DRN) and Residue contact map calculations

To assess intra-protein communication and changes occurring in the protein residues across the MD simulation, DRN calculations were performed. All the DRN and residue contact maps calculations were performed via the online platform MDM-TASK-web available at (<https://mdmtaskweb.rubi.ru.ac.za/>; accessed on 1 February 2021) [308]. Prior to DRN calculations, both the trajectory and topology files were reduced to consist of only C β atoms for each protein-ligand membrane system (C α for glycine) across all frames using VMD [304]. A cut-off distance of 6.7 Å was used on the reduced trajectories as the residue-residue contact distance, and a step size iteration of 100 frames was applied. There is a total of five DRN centrality metrics (*BC*, *CC*, *DC*, *EC* and *KC*) as described in chapter 2. Firstly, averaged *BC* and *L* were calculated for PART 1 section. Briefly, averaged *BC* identifies residues most important in communication within a protein while *L* identifies those accessible to the protein. This essentially provides more information on ligand binding and how it affects protein structure and function. The averaged *BC* was calculated to identify the residues in the network that are important in protein communication. The change in *BC* in protein residues was

calculated over the last 100 ns of the MD simulation, representing the equilibrated phase (250-350 ns). For easier comparison of *BC* values across WT and mutant proteins of the eight ligand systems, the data was normalised onto a scale of 0 - 1 using unity-based normalisation. Delta *BC* was then calculated by subtracting average values in mutant from WT proteins. Using the normalised output data, the significant residues within ± 2 SD were mapped onto the protein structure using PyMOL Molecular Graphics System v1.7.2.1 [175].

To determine communication ‘*hubs*’ comparing all ligand systems (PART II and III) using DRN five centrality metrics (*BC*, *CC*, *DC*, *EC* and *KC*), the top 4 % residues were calculated from script prepared by Rolland Bantar Tata as there were no ‘*persistent hubs*’ as previously published [137]. The residue hubs in each system were computed by first sorting the data points in each system in descending order. This was performed to obtain nodes having highest centrality. Thereafter, the top 4 % of these data points were obtained per system and using a heat map, comparison of the hubs with 90 % occurrence was performed across all the ligand systems. The significance level of 90 % was applied resulting in communication hubs with least chance of inaccuracy. A comparative analysis was performed between holo and ligand-bound systems in WT, Y268C, Y268N and Y268S states of the eight ligand systems in PART II and 13 ligand systems in PART III. Residue contact maps and heat maps were also calculated using in-house scripts and altered bash scripts prepared by Dr Thommas Musyoka for the equilibrated phase of the trajectory (last 100 ns) for all systems (WT, Y268C, Y268N and Y268S states of the eight ligand systems in PART II and 13 ligand systems in PART III), based on pairwise residue distance between all C β (or C α for glycine residue). Briefly, a residue network graph is generated targeting a residue of interest, in this case around all the 28 Qo active site residues. Contact frequencies are thus generated in decimal values ranging from 0 to 1 equivalent to percent (%) MD simulation time. A contact cut-off distance of 6.7 Å and an iteration step size of 100 frames were applied to obtain all residue-residue contacts around the

active site and subset residue positions. The residue-residue contacts were represented as the percentage fraction of the selected MD frames. All these calculations were performed to assess the effect of ligand binding as well as mutations on the PfCytb-ISP active site pocket.

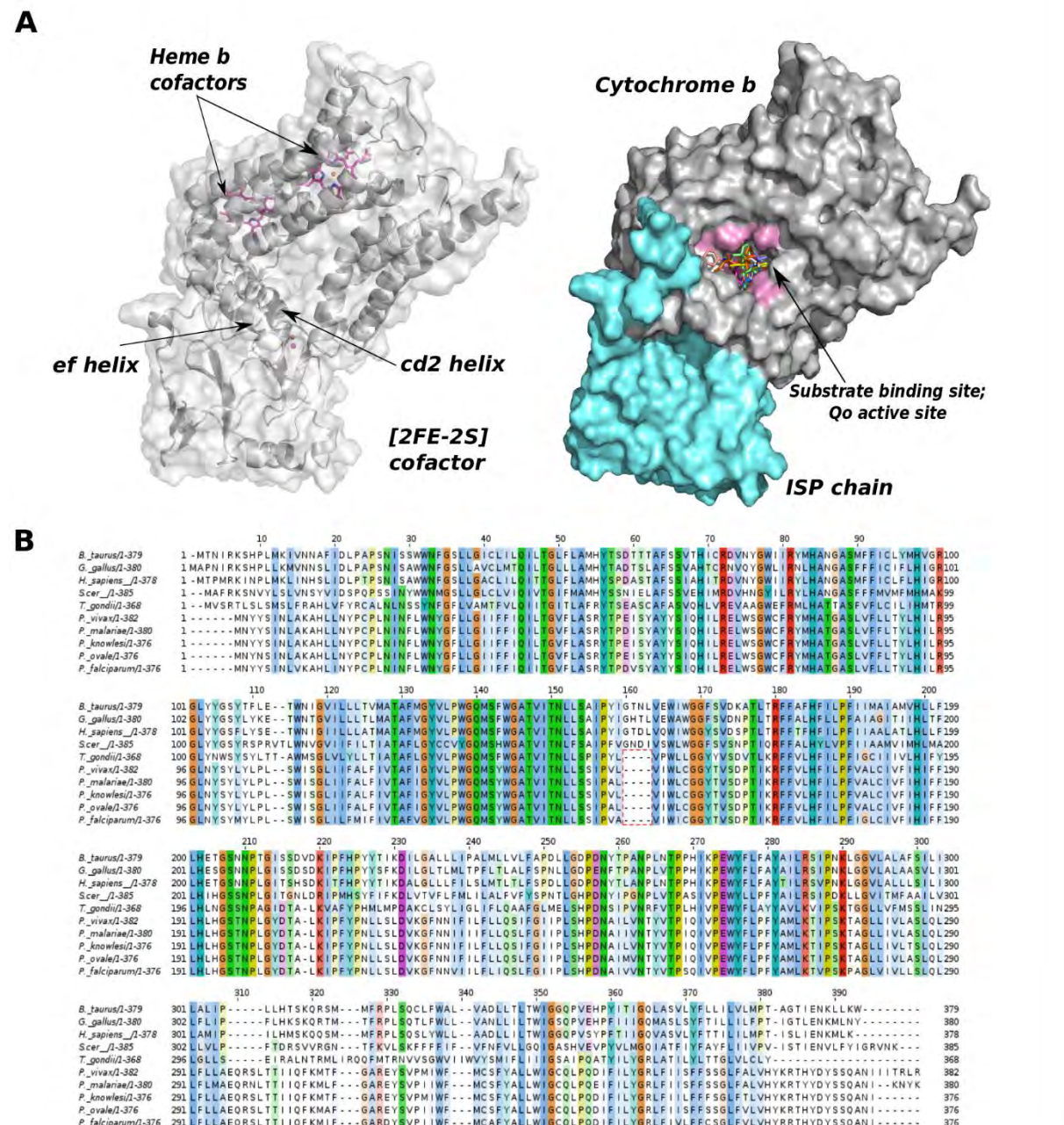
5.2.7 Principal Component Analysis

Principal component analysis (PCA) is a method used to explain variance occurring in a dataset representing conformations sampled in a system as principal components (PCs) [309]. PCA was performed on the holo and eight protein-ligand systems, assessing the conformational entropy of their WT and mutant states (Y268C, Y268N and Y268S) according to a previous protocol [310]. Concisely, the coordinate covariance matrix for the C α and C β atoms of each system was calculated by first performing a coordinate RMS best-fit of the trajectories to the reference structures. To obtain the eigenvectors and eigenvalues for each system, the generated matrix was diagonalised to allow projection of the protein coordinates, and then the normalisation process was applied to all systems. Respective free energies for each system were then calculated, after which PC1 versus PC2 displaying free energies per system were plotted to visualise the 3D conformational sampling across all protein systems.

5.3 Results and discussion

5.3.1 Structure and sequence conservation of PfCytb-ISP as an attractive drug target

An overview of the PfCytb-ISP structure highlighting the location of the Qo substrate binding site is shown in **Figure 5.1A**.



cytochrome *b* protein coloured by amino acid properties highlighting unique features (4-residue deletion) to be explored for inhibitory drug design.

From this, the Qo active site is shown to be located lateral to the transmembrane helices E and F (*ef* helix), whose tail loop carries the conserved 'PEWY' (Pro260-Glu261-Trp262-Tyr263) motif [311]. **Figure 5.1B** highlights the substrate binding pocket where selective inhibition analysis was performed. While the PEWY motif functions to preserve the entire catalytic site, residue Glu261 plays an essential role in formation of hydrogen bonds for ubiquinol substrate binding and stability [220,312]. The multiple sequence alignment in **Figure 5.1B** shows high conservation (residue regions coloured across all proteins) sites in cytochrome *b* protein across *Plasmodium* and its homologs. Also, as noted in Chapter 3, active site residues that the starred (Chapter 3; **Figure 3.9**) are key active site residues across the 10 species, expected to form interactions with the substrate. The residues coloured across both *Plasmodium* spp. and human host, showed high sequence similarity indicating 100 % sequence conservation.

5.3.2 Protein structure and modelling of the Qo active site; PfCytb-ISP and HsCytb-ISP

Both the plasmodial (PfCytb-ISP) and the human (HsCytb-ISP) protein models were successfully generated prior to performing molecular docking and MD simulation experiments. PfCytb-ISP was modelled, and its accuracy evaluated as detailed in our previous study (refer to Chapter 3) [141]. The HsCytb-ISP protein was modelled successfully, and its quality was evaluated using both the Ramachandran plot and z-DOPE score. A Ramachandran plot displaying the proteins' psi (ϕ) and phi (ψ) main-chain torsion angles was constructed which identified residues in the acceptable versus those in the outlier region. The plot, presented in **Figure S5.1**, showed that 89.5 % of the total residues for HsCytb and 79.2 % of ISP modelled subunits were in the most favored regions, while 0.0 % of the residues in both subunits were in disallowed regions. Additionally, the z-DOPE score of the modelled wild-type (WT) HsCytb-

ISP protein complex was -0.72, where a score toward -1.0 indicated that the model was of high quality thus suitable for further calculations. The mutant HsCytb-ISP proteins were generated by mutating the WT models using Discovery Studio Visualizer, and the z-DOPE scores for mutant models Y268C, Y268N and Y268S were -0.76, -0.75 and -0.75, respectively.

The mutations occurring at the Qo active site have been reported to alter the active site pocket volume affecting substrate binding affinity. This was confirmed in previous work where the pocket volume of the mutant proteins was reported to have increased compared to the WT [62,141]. A closer look at the mutations found in the active site revealed that the individual amino acid substitutions have possible implications on substrate binding affinity and stability. In terms of structure, tyrosine (Tyr) residue has an aromatic group on its side chain that maintains its structure and forms hydrophobic interactions to stabilise the substrate [62]. The lack of aromatic groups in the substituents of the point mutations Y268C, Y268N and Y268S reduces the capacity of these residues to form ligand stabilising interactions. For instance, previous studies confirmed this fact, showing that the mutation from tyrosine to cysteine at position 268 affected ATQ binding affinity [63].

PART I

EFFECT OF MUTATION ON ATQ DRUG ACTIVITY

The main aim of PART I was to investigate the effect of mutations on ATQ drug activity. Thus, the MD simulations of both holo and ATQ-bound systems in this section were performed in triplicate to validate the findings observed prior to performing structure-based drug discovery in the subsequent sections (PART II and III).

5.3.3 Global MD analysis of holo protein and ATQ-bound systems in a POPC: POPE phospholipid bilayer

5.3.3.1 ATQ docking results and trajectory analysis via MD simulations

Docking of the ATQ drug into the Qo site of the PfCytb-ISP protein was performed as detailed in the methodology section. The best docked pose and interactions of ATQ are shown in **Figure 5.2A**. The ATQ drug exhibited binding energy of $-8.7 \text{ kcal/mol}^{-1}$. The final docked pose showed ATQ bound in the Qo active site with its cyclohexyl group facing upward and away from the heterodimer interface of the protein complex.

Figure 5.2B shows 3D (3 dimensional) and 2D (2 dimensional) ligand interactions in different time steps (50 ns, 247 ns and 350 ns) during MD simulation of the WT system. Here, the ligand interactions within 5 Å around the drug were investigated across the three time steps (**Figure 5.2B**). Additionally, the trajectory was visualised and ATQ's chlorophenyl tail was shown to keep flipping throughout the MD simulations in both WT and mutant systems. **Figure 5.2** displays this flipping which is illustrated by the red circle (in 3D) as well as rotation exchange of atomic positions C20 and C21 (in 2D).

5.3.3.2 Trajectory analysis and dynamics of the holo protein and ATQ-bound systems

To determine the conformational sampling in the holo proteins during MD and understand the effect of the point mutations (Y268C/N/S) on the protein, a global comparative analysis of the holo protein and ATQ-bound systems was performed in triplicate. Studying these dynamic changes in PfCytb protein will be useful in future *in silico* drug design studies [313].

Prior to MD simulations, the lipid bilayer was inspected for any anomalies. The inspection showed that the ratio of POPC to POPE membrane lipids remained approximately 1.3:1 in both lower and upper leaflets. The presence of all the membrane elements including lipids was confirmed, thus the system complex was confirmed to be accurately constructed. Also, for accuracy in the experimental setup, the membrane packing process was checked where, during MD simulation, the water and lipids were monitored and none of the water molecules entered the membrane.

The protein dynamic changes to RMSD and Rg between the WT and mutant proteins during MD simulations were investigated (**Figure 5.3A, B**). From **Figure 5.3A**, it is observed that both WT and mutant protein systems showed less than 2 Å deviations in RMSD, showing that the systems equilibrated during simulations. Due to the complexity of the POPC:POPE bilayer complexed with the protein, MD simulations for all systems were performed over 350 ns to

obtain sufficient conformational sampling in monitoring protein dynamics [314]. Backbone RMSD distributions showed the Y268S mutant systems equilibrated at a higher value compared to the WT. Similarly, these findings corroborate previous *in silico* studies [18] (performed without heme and a phospholipid bilayer) that included protein dynamics for only the Y268S protein [315]. The higher RMSD values reported in mutant ATQ-bound systems may indicate instability due to ligand binding. The trend is more evident when ATQ is bound to the proteins across the triplicate runs, except for the holo Y268C systems.

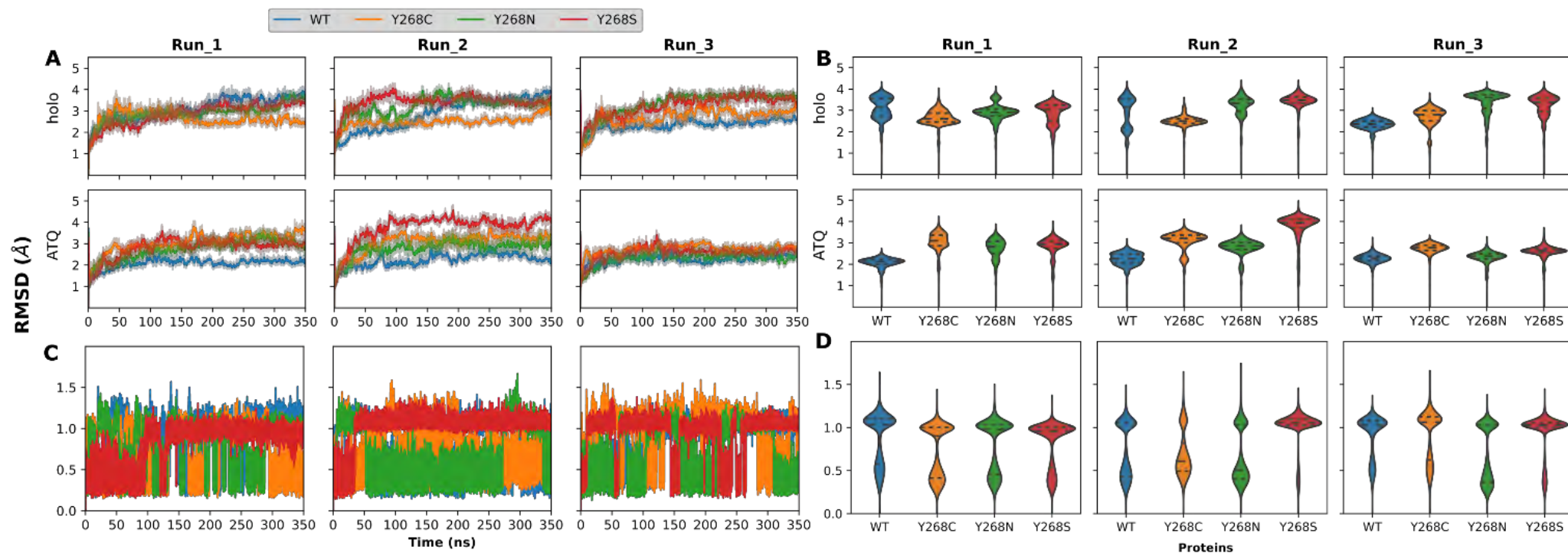


Figure 5.3: Molecular dynamics trajectory analysis of the changes to protein root mean square deviation (RMSD) in both the holo and ATQ bound systems. (A) Protein RMSDs over the 350 ns simulation; (B) Protein RMSD violin plot showing conformational distribution and sampling; (C) Ligand (ATQ drug) RMSD over the 350ns simulation; (D) ATQ RMSD violin plot presenting conformational sampling and distribution. Reproduced from [141].

RMSD conformational sampling distributions of the holo proteins (**Figure 5.3B**) reported multi-modal distributions for the WT and mutant proteins, suggesting that the proteins sampled multiple distinct conformations during MD simulations. Although protein distributions and RMSDs of the most sampled conformations may differ between WT and mutant proteins, the median structures of proteins across all runs are within 1.0 Å of each other. With regards to the presence and number of flexible loops in the protein, the simulated PfCytb-ISP protein complex accounts for a total of 33 % loop regions, which could explain the differences observed in the RMSDs. This highlights some degree of conformational similarity in the core region of the proteins. Since the mutations confer drug resistance, the relative degree of conformational similarity suggests that there is an insignificant impact of the mechanism of resistance globally, although more insights may be gained at local residue level.

We report ATQ-bound systems to sample unimodal distributions predominantly as compared to the holo proteins, suggesting ATQ drug binding may have a stabilising effect on the protein, unlike the first two runs of the holo WT protein that show multiple conformations (**Figure 5.3B**).

Out of all the ATQ-bound mutant proteins, Y268S showed different behaviour among all three runs, with first, second and third runs exhibiting similar median, higher median and lower median, respectively. This could indicate similar mechanisms of resistance between Y268C and Y268N. However, local analysis will further investigate the resistance mechanisms to confirm or refute the global analysis.

Additionally, to determine which system deviated the most from holo WT protein, average median RMSD values and their differences were calculated for all triplicate runs in each protein. The average median difference between WT and mutant systems was performed by

subtracting the mutant average median RMSD from that of the WT protein. Average median values of 3.06, 2.66, 3.35 and 3.38 Å were observed for WT, Y268C, Y268N and Y268S systems, respectively, while the average median differences were 0.40, -0.29 and -0.32 for mutant systems Y268C, Y268N and Y268S, respectively.

Out of all the mutant proteins, Y268C exhibited the largest deviation from WT after it sampled lower RMSD values in all triplicate runs. A marked difference in previous studies [316] which featured Y268S mutation only was that, in this study, all protein systems were embedded in a lipid bilayer. Here, the results presented protein dynamics for all the three mutant systems in comparison to the WT. Overall, the RMSD differences observed in these mutant systems showed the varied impact of the point mutations on the protein's structure and behaviour.

The RMSDs of the ATQ ligand with respect to protein in WT and mutant proteins over the 350 ns simulations are presented in **Figure 5.3C, D**. In each case, the entire trajectory was assessed, where the ATQ ligand was observed to adopt different poses in the mutant proteins as compared to the WT. Various pose changes are highlighted by shifting RMSD values. Pose changes to ATQ, with the exception of Y268N (second run) and Y268S (first two runs), were observed. This could be indicative of increased ligand stability. Due to these pose changes, it is possible that ATQ drug loses the ability to block ISP domain movement within the mutant proteins, which is a mechanism of action reported in previous literature [55]. Here, we observe that the ligand behaviour agrees with the previous section (**Section 5.3.3.1**), where RMSD jumps were observed due to flipping structures across all the simulation systems while the drug remained bound in the Qo active site.

Analysis of the protein compactness was performed as shown by the Rg (**Figure S5.2**) where the Rg values ranged from 26.5 to 27.5 Å in all systems indicating relatively similar compactness across all protein systems in all triplicate runs. Comparing WT and mutant

proteins, Y268C and Y268S mutant proteins might be more compact than the WT and Y268N systems due to their relatively lower Rg values. Upon ATQ binding, this trend was however not observed, which suggests that the presence of ATQ has a probability to influence the compactness of the proteins. Unlike the RMSDs of the holo proteins, the Rg violin plots showed unimodal distributions, suggesting that there were no major changes to the protein centre of mass during MD simulation. Even though RMSD and Rg portrayed differences of less than 1.0 Å indicating similar structures, these differences will further be analysed as protein structure and function is further investigated throughout this thesis.

The effect of mutations on residue flexibility was also analysed (**Figure S5.3**). Here, it is evident that the higher flexibility (greater RMSF values) was located towards the C-terminal of the protein at the intersection of the two chains as well as the loop regions of the protein, in accordance with previous studies [316]. The loop regions include PfCytb chain: 21-27, 187-180, 207-213, 218-232, 265-285, 294-315, 334-339 and ISP chain: 218-235, 238-243, 270-273, 277-282, 297-306, 317-343. No major differences were highlighted comparing the residue flexibility of WT and mutant proteins.

5.3.4 Analysis of conformational entropy in holo and ATQ-bound proteins comparing WT and mutant proteins

Using PCA, the first two principal components that show the largest variation (PC1 and PC2) and associated free energy for holo and ATQ-bound systems were analysed (**Figure 5.4**). The WT holo proteins were reported to adopt one major conformation cluster of low energy, supporting their global stability. A similar pattern was also noted in the mutant process, except for Run-3 which lacked low energy conformations. The WT holo proteins have been reported to have similar conformational sampling across the triplicate proteins with the low energy wells reported by PC1 (**Figure 5.4A**). Mutant holo proteins identified Y268C and Y268S to have

similar conformations within the first two runs, suggesting that the two mutants are likely to share similar resistance mechanisms. Also, the proteins have been reported to adopt low energy conformations even though the mutations are associated with ATQ resistance. These low energy conformations are expected as any instability would be detrimental to protein function.

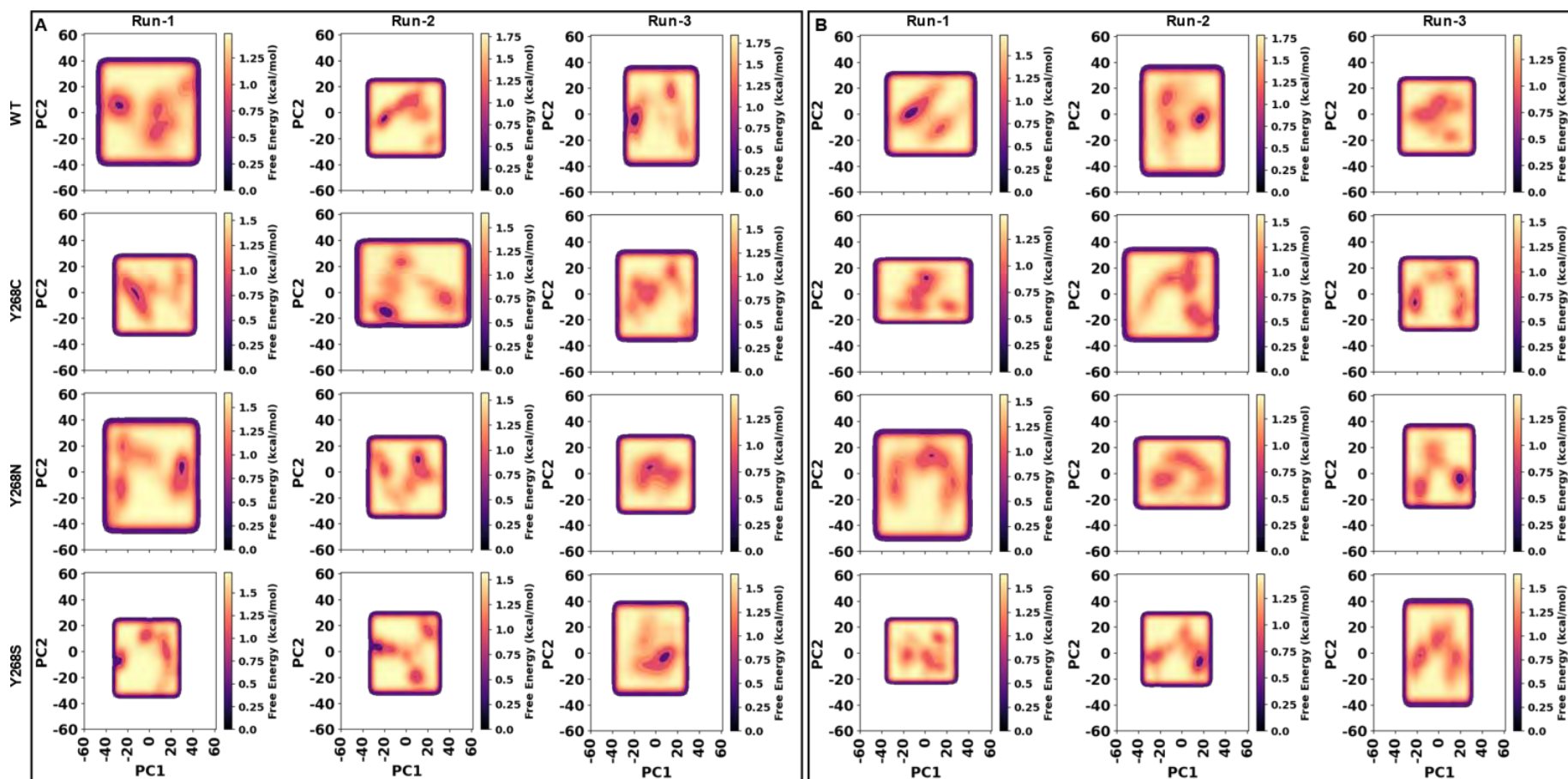


Figure 5.4: Principal component analysis (PCA) of the holo and ATQ bound proteins. (A) Holo; (B) ATQ bound proteins. The plots show a 3-dimensional (3D) plot of PC1 versus PC2 of the WT and mutant holo proteins as a function of free energy. All protein systems are shown in three replicate runs (Run-1, Run-2, and Run-3). Each scale corresponds to each individual plot where the black colour at 0 kcal/mol represents the lowest free energy conformations. Reproduced from [141].

Analysis of the Y268N mutation with respect to the WT protein highlighted two possible resistance mechanisms. Firstly, in the presence of mutations, the proteins are likely to adopt a different conformation capable of function and resistance to ATQ. Secondly, the same mutant system has shown higher energy conformations that coincide with the low energy wells of the protein (according to PC1 of -20 to -30) which suggests that the mutant can occupy a functionally active protein conformation like that of the WT but that is less energetically favourable. This could mean that the conformation is functional yet neither catalytically active nor abundant, which could have a direct implication on the activity of the ATQ drug. As such, further research should be performed to ascertain this finding because experimental analysis has only been performed on the Y268S mutant.

Upon ATQ binding (**Figure 5.4B**), lower energy conformations were observed in WT as compared to mutant proteins apart from Run-3, which lacked low energy well. These observations suggest that ATQ binding may possibly result in conformations that are not energetically favourable. For instance, ATQ-bound proteins in WT state (PC of 15) showed their conformations to be like that of the holo protein in Y268N state, which supports the previously suggested resistance mechanism. In general, the comparison of ATQ-bound proteins with holo proteins evidently illustrated an increase of free energy in the ATQ-bound proteins.

5.3.5 Local residue interaction analysis using weighted residue contact maps

To determine the potential effect of mutations (Y268C, Y268N and Y268S) on the protein, intra-protein network analysis was performed. The protein interaction network was investigated focusing on residue-residue interactions around a total of 28 Qo active site residues

(especially 268 mutation residue position) in both holo and ATQ-bound proteins. The active site residues included: Met116, Ile119, Val120, Phe123, Val124, Met133, Trp136, Gly137, Val140, Ile141, Thr142, Leu144, Leu145, Ile155, Phe169, Leu172, Ile258, Val259, Pro260, Glu261, Trp262, Tyr263, Phe264, Phe267, Tyr268, Leu271, Val 284, Leu285 [54]. This was inspired by the fact that one of the effects of point mutations on drug binding and PfCytb-ISP active site is the destabilisation of the hydrophobic interactions [203,205]. The active site residues are important in substrate binding and stabilisation [288]. Contact frequencies ranging from 0 to 1 were calculated, representing the degree of the presence of residue interactions during simulation (0 - 100 % of MD simulation time) where interactions such as hydrogen bonds and van der Waals (vdW) were analysed [317]. The residue interaction changes were compared between the WT (reference) and mutant systems where residue interactions within a cut-off distance of 6.7 Å were monitored throughout the equilibrated phase of MD simulation. In the WT holo protein, the residue-residue contacts were observed to be maintained 100 % of MD simulation time, hence used as a reference. These contact frequencies (0 - 1) allowed for assessment of reduced, lost, gained, or enhanced residue interactions. This method has been employed in various studies investigating the effects of ligand binding and mutation on proteins [80,310,317].

Like global analysis in previous sections, the residue interaction changes were performed in triplicate runs for consensus, with the local residue analysis giving consistent results, unlike RMSF results. Comparison was performed between both holo and ATQ-bound proteins of WT (**Figure 5.5**) and each of the mutant proteins (**Figure S5.4 - S5.6**).

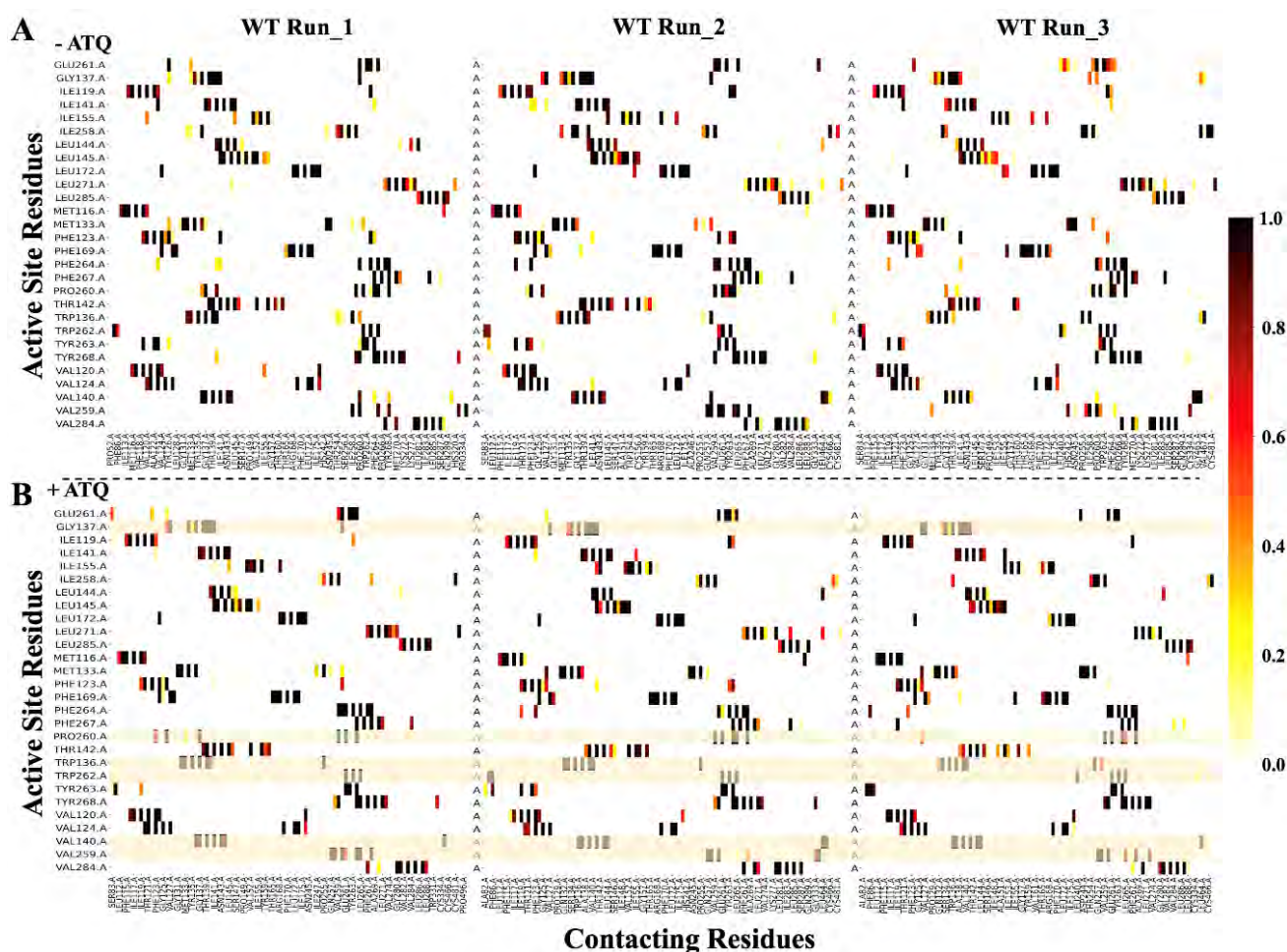


Figure 5.5: Residue contact heat maps for PfCytb-ISP active site residues. Heat map representations of the active site residues forming contacts with other residues (contacting residues) in (A) WT (wild-type) system across triplicate runs in holo protein structure and (B) WT system across triplicate runs in ATQ-bound protein structure. The residue contacts were monitored over the last 100 ns of the simulation time. The yellow lines across the heat maps in (B) highlight residues of either reduced or lost contacts relative to the WT. Colour intensity is proportional to the occurrence of residues contact throughout the simulation, i.e., 0 - 1 on the colour bar is equivalent to 0 - 100 % MD simulation time. Zoomed figures comparing each of the mutant systems to WT in both holo and ATQ-bound forms are shown in Figures S4.4–S4.6. Reproduced from [141].

According to **Figure 5.5**, hydrophobic interaction Val140-Ile258 was lost in the ATQ-bound system across triplicate runs while Val259-Cys334, Pro260-Phe264, Trp262-Ser83, Gly137-Pro260 and Trp136-Ile258 residue-residue interactions were reduced/lost in at least two runs. Motif analysis, outlined in chapter 3 showed that these residues formed part of conserved motifs 2 (259, 260, 262) and 3 (136, 137, 140), referring to conservation between *P. falciparum* and human homolog. Out of these interactions, Pro260 and Trp262 form a part of the PEWY

motif, a conserved motif important in substrate binding [311]. Hence, residue contact frequencies were observed to be compromised in these ATQ-bound systems, possibly affecting the substrate binding function. Worthy to note, the residue-residue interactions at position 268 were maintained in the WT ATQ-bound protein.

In mutant analysis, four active site residues (Trp136, Gly137, Pro260 and Phe264) were reported to have compromised (reduced/lost) interatomic interactions in the holo protein in at least two mutant systems (**Figures S5.4 - S5.6**). Comparative analysis of mutant systems Y268C, Y268N and Y268S residue contact map reported reduction/loss in seven residue-residue contacts. The first contact (Trp136-Val140) was reduced in mutant systems (Y268C: 26 %, Y268N: 79 % and Y268S: 0 %). The second contact (Trp136-Met133) was reduced in mutant systems (Y268C: 59 %, Y268N: 0 % and Y268S: 16 %). Residue-residue contacts of active site residue Trp136 to residues Gln132 and Pro260 were lost in all three mutant systems. The active site residue Gly137 was further assessed in Y268C and Y268N mutant systems, and a reduction in the third contact (Gly137-Ile141) for mutant system Y268C occurring at 70% of simulation time was found. We also observed that the fourth contact (Gly137-Pro260) was reduced in Y268C and Y268N mutant systems, where the respective contacts occurred at 50 % and 1 % of MD simulation time. The fifth contact (Pro260-Phe264) was lost in two runs in the Y268N mutant system and reduced in the Y268S mutant system, occurring at 50 % of the simulation time. Lastly, active site residue Phe264 was assessed where the sixth contact (Phe264-Asn/Ser268) was observed to be reduced in mutant systems Y268N and Y268S, occurring at 94 % and 5 % of the simulation time, respectively. From the Phe264 residue assessment, the seventh contact (Phe264-Pro260) was reduced to 38 % and 50 % in Y268N and Y268S systems, respectively. Linking residue contact frequencies to motif analysis (Chapter 3), these residues formed part of conserved Motif 2 (260, 264) and Motif 3 (136, 137). Notable is residues Pro260 and Phe264, forming a part of the PEWY motif, were observed to

be compromised (reduced/lost) in the mutant systems. This possibly suggest the likelihood of the mutations affecting substrate binding function by decreasing residue contact frequencies.

However, ATQ-bound proteins observed both reduction/loss and gain/enhancement in a total of five residue-residue interactions. Reduced interactions included Trp136-Gln132 occurring at 18 % (Run_1) and 8 % (Run_2) of MD simulation time for the Y268C mutant system. Interaction Met133-Asp244 was reported to occur 15 % (Run_2) and 28 % (Run_3) of simulation time in Y268N mutant system. Lastly, interaction Thr142-Val152 was reduced in Run_1 and Run_2, occurring for 1 % of simulation time in each run in the Y268S mutant system. These residues formed a part of conserved motif 3 (136, 133, 142) and are reported to experience reduction/loss of residue interactions. This suggests that in the presence of the mentioned mutations, the protein function is likely to be affected. More reduced residue interactions were reported in ATQ-bound proteins than holo proteins, which showed that the proteins' substrate binding involving the active site residues is likely to be disrupted. Evidently, we can link the compromised interactions to the presence of point mutations where the ability to block the ISP domain movement is impaired as reported in the literature [55]. Thus far, Y268C and Y268N mutant systems are likely to share similar resistance mechanisms. There were two gained interactions that were reported in the Y268C and Y268S mutant systems. These were: Val259-Cys334 in Run_1 (100 %) and Run_3 (87 %) for Y268C as well as Ile258-Thr254 in Run_1 (93 %) and Run_2 (86 %) for Y268S mutant systems. Active site residues Ile258 and Val259 (forming a part of conserved motif 2) participate in the supporting the catalytic function at the Qo site. Thereby, these residue interactions could possibly compensate for the compromised interactions reported once ATQ- drug binds.

In residue position 268, forming a part of conserved Motif 2, reported four distinct residue-residue contacts in the WT holo protein (Tyr268-Gly333, Tyr268-Val259, Tyr268-Pro260 and

Tyr268-Phe264) where the interactions were maintained for 66 %, 90 %, 99 % and 99 % of the simulation time, respectively, while they were compromised in the presence of the mutation. Contact Tyr268-Gly333 was reduced in all three mutant systems (6 %, 6 % and 0 %) while Tyr268-Val259 was reported in Y268S and Y268C mutant systems (24 % and 3 %, respectively). Contacts Tyr268-Pro260 and Tyr268-Phe264 were reduced (0 % and 8 %) in mutant systems Y268N and Y268S, respectively. Also, residues Gly333 and Val259 were reduced/lost in all mutant systems, while the same was seen for Pro260 only in the Y268N mutant system. However, in ATQ-bound systems two interactions (Tyr268-Gly333 and Tyr268-Val259) were reduced in all mutant systems (Y268C, Y268N and Y268N) in at least two runs, which further supports the effect of the mutations on substrate binding.

Interatomic interactions predicted by DynaMut (Chapter 3: **Figure 3.7**) revealed three interaction changes: Tyr268-Pro260, Tyr268-Leu265 and Tyr268-Leu271. Out of the three interactions, the contact map findings of the reduction of hydrophobic interaction (Tyr268-Pro260) were in good agreement with the DynaMut predictions in **Figure 3.7**, while vdW interactions Tyr268-Leu265 and Tyr268-Leu271 were observed 100% of simulation time as seen in **Figure S5.7**. Additional interactions between residues Tyr268 and Pro260, and Leu265 and Leu271, were reported in the mutant proteins although not observed in **Figure 5.7**. This varied observation could be because prediction analysis was performed on a static structure, as opposed to weighted residue contact map which samples vast data over MD simulation. Overall, active site residues Val140, Val259, Pro260 and Phe264 crucial in substrate binding (forming a part of PEWY conserved motif) reported either reduction/loss of contacts in mutant systems (Y268C and Y268N), (Y268C, Y268S and Y268N), Y268N and Y268S, respectively. It was observed that structurally, active site residues Val259 and Pro260 are in a flexible loop; hence the interactions in the mutant systems are likely to be affected. Out of the three mutations, mutations that had the greatest effect on interactions were determined by ranking the residues

that experienced reduction/loss compared to total number of residues forming contacts in WT. From this, Y268S reported 20 % while mutant systems Y268C and Y268N reported 30 % each, showing the latter two systems had the greatest effect on the protein's residue interaction network. Structurally, four contacting residues (Leu265, Phe267, Ala269 and Leu271) located in the protein's alpha-helices formed sufficient stabilising interactions due to their high conservation [318]. These residues form a part of conserved motif 2 located in the Qo active site, hence important in preserving the catalytic function.

In summary, the molecular size and the physicochemical differences in residue-residue interactions were analysed in each of the mutations (Y268C/N/S). Residue Tyr268 supports substrate binding due to its aromatic side chain that stabilises the substrate [319]. A large hydrophobic residue or an aromatic side chain would maintain optimal enzyme activity at this residue position. In this study, all three mutations lacked the aromatic side chain, and in terms of size, both Ser and Cys residues are small compared to the Asn residue. Therefore, we would expect a severe mutation effect in Y268S and Y268C as opposed to the Y268N mutant system with regards to interatomic interactions. From previous findings (**Figures 5.5** and **S5.7**) Y268S and Y268C were likely to cause higher resistance levels than the Y268C mutation as they were the most affected. These findings corroborate the prevalence of cytochrome *b* mutations as previously reported in order: Y268S, Y268C, Y268N [161–163,320]. Overall, these findings are in good agreement with previous studies showing reduced hydrophobic interactions at the protein's active site due to mutations [203].

Overall, four residues out of a total of 20 contacting residues, representing ~20 % of the residues forming contacts in WT, were observed in all the mutant systems. Structurally, the four contacting residues (Leu265, Phe267, Ala269 and Leu271; residues that form motif 2

pattern) are in the alpha helices of the protein, which were reported to be conserved thus more stable and able to form sufficient stabilising interactions [318].

5.3.6 Effect of mutation on protein communication and residue accessibility

Using dynamic residue network (DRN) analysis, mutant effects on the protein's residue accessibility and communication were investigated focusing on the last 100 ns (250 - 350 ns time scale) of MD simulation. This was performed in holo and ATQ-bound systems to identify significant residues involved in intra-protein communication in the WT and mutant systems as employed in previous studies [295,321,322]. L highlighted the accessibility of a residue to all other residues, while BC was indicative of the importance of a residue to protein communication. The residues associated with lower L measurements were more accessible while those with higher BC values were highly used in accessing the rest of the residues. Both the low L and high BC measurements pointed to the importance of the residues in intra-protein communication and function. The results in **Figure 5.6** and **Table S5.1** highlighted key residues in holo protein based on their high BC values, with some identified to be active site residues important in enhancing substrate binding and stability [203,315]. The residues with high BC values included: 85, 120, 123, 124, 140, 141, 160, 260–265, 268, 271 and 272, out of which Val120, Phe123, Val124, Val140, Ile141, Pro260, Glu261, Phe264, Tyr268 and Leu271 were active site residues. Altogether, most of these residues form a part of motif 1 (85), motif 2 (260-265, 268, 271-272), motif 3 (120, 123-124, 140-141) and motif 4 (160) which are conserved motifs between *P. falciparum* cytochrome *b* and its human homolog. Residues 260-263 form PEWY conserved motif that is essential in Qo substrate binding [311]. In the absence of ligand binding, these findings suggest that the protein communication is maintained

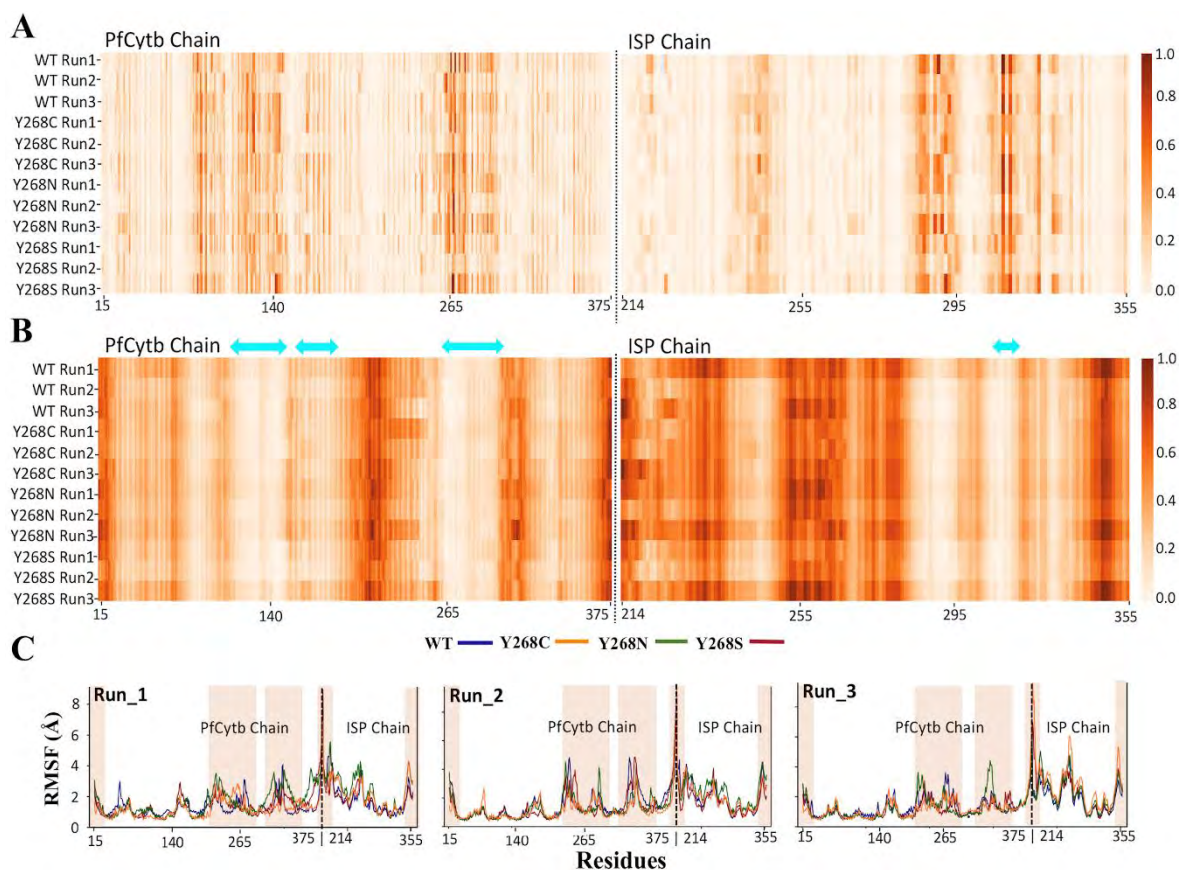


Figure 5.6: Dynamic residue network (DRN) analysis. Heat map representation of normalised values of (A) *Betweenness centrality* (*BC*) and (B) *Average shortest path* (*L*) for PfCytb and ISP chains. *L* values were correlated with root mean square fluctuation (RMSF) which is plotted in (C) where the WT (wild-type) and mutant systems Y268C, Y268N and Y268S are represented by the blue, orange, green and red lines, respectively. (C) Black dotted lines in RMSF plots show an intersection between the two chains. Two-sided cyan arrow on the *L* heat map represents active site residues. Each protein system had triplicate runs presented in (A–C). Reproduced from [141].

A heat map of *BC* values was plotted to compare both holo and ATQ-bound proteins in triplicate runs (**Figure S5.8**). High *BC* values (**Table S5.1**) were reported for all holo systems, with some residues being outstanding due to their multiple occurrences across protein systems.

Important to note is Trp262 residue which was observed to have high *BC* values across WT and mutant systems in all triplicate runs. Analysis of each of the systems reported high *BC* values as follows: residues Trp262 and His320 across the triplicate runs and residues Phe123, Val259, Phe264 and His301 in at least two runs of the WT system. In the mutant systems,

residues Leu271 and His320 are common across all runs, and Phe123, Ile141, Glu261 and Cys268 are common in at least two runs in Y268C. Residues Phe264, Asn268 and His320 were common across all runs and residues Pro260, Glu261 and His301 in at least two runs of the Y268N system. Lastly, the Y268S mutant system reported residue His320 to be common across triplicate runs and residues Phe123, Val140, Ile141 and Phe264 to be common in at least two runs. These high *BC* residues are part of conserved motifs 2 (259, 262, 264, 268, 271) and 3 (123, 140-141). This suggests that despite the presence of mutations, the proteins' overall function including substrate binding is likely to be supported due to the presence of the conserved motifs.

In ATQ-bound systems, the residues with high *BC* values (**Table S5.2**) were similarly observed in holo proteins. In the WT system, residues Phe169, Val259, Phe264, Cys299, His301 and His320 were reported in at least two runs. In mutant systems, residues (Phe123, Val124, Tyr263, Phe264, Leu271, His301 and His320), (Met133, Tyr263, Phe264, Asn268, His301 and His320) and (Phe123, Gly137, Val140, Tyr263, Phe264, Ser268, His301 and His320) were observed in Y268C, Y268N and Y268S systems, respectively. In summary, it was noted that active site residue Phe264 and metal coordinating residues (His301 and His320) were reported in at least two runs in WT and Y268C, Y268N and Y268S mutant systems. The high *BC* residues are part of conserved motifs 2 (259, 263, 264, 268), motif 3 (123, 124, 137, 140) and motif 4 (169). These residues are observed to gain more central role in protein communication upon ATQ binding even in the mutant systems. Overall, both holo and ATQ-bound systems reported a similar number of active site residues that had high *BC* values, indicating high connectivity in the proteins' catalytic region. Catalytic residues Trp262 and Tyr263 form the mitochondrial PEWY (Pro260-Glu261-Trp262-Tyr263) conserved motif [323] which is found on the protein's loop region. Based on previous findings, they have been reported to maintain the function of the Qo active site [323]. On the other hand, residues His301 and His320 which

coordinate the [2FE-2S] metal cluster of the ISP chain were reported in not only the WT but also mutant systems. These findings show that the protein function is likely to be maintained when the ATQ drug binds even in the presence of mutations. This is a positive step in drug discovery towards identifying novel inhibitors that can withstand ATQ resistance caused by mutations [315].

Further, the effect of Y268C/N/S mutations on protein communication was investigated by calculating changes to average BC (ΔBC : WT - Mutant). A threshold value of ± 2 standard deviations (SD) of the ΔBC values was used for each holo and ATQ-bound PfCytb-ISP protein system after the data was confirmed to assume a normal distribution (**Figure S5.9**). Significant residues with greater or less than 2SD were reported and discussed (**Tables S5.3 and S5.4**). In the mutant systems, negative and positive ΔBC signify an increase and a decrease in residue connectivity, respectively. The negative and positive values represent enhanced and diminished residue participation in protein communication of the ligand-bound systems. All significant residues in holo and ATQ-bound systems were mapped onto the protein structure (**Figure 5.7**), with most of the residues located either in the Qo active site or within its proximity. Across all the mutant systems, Trp262 residue was observed to lose centrality ($+\Delta BC$ values). Specific to each mutant system, the following residues were observed to lose importance: Ile119 in Y268C; Phe123, Val140 and Ile141 in Y268N; as well as Ile119 and Val259 in Y268S mutant systems (refer to **Table S5.3**). As for ATQ-bound proteins, both increased and decreased connection was reported in the mutant systems (**Table S5.4**). In terms of loss of connectivity, residues Val259 in Y268C and His301 in Y268S were noted in at least two runs. However, residues Phe264 in Y268C, His192 and His320 in Y268N, and Phe123 and Tyr263 in Y268S systems gained connectivity.

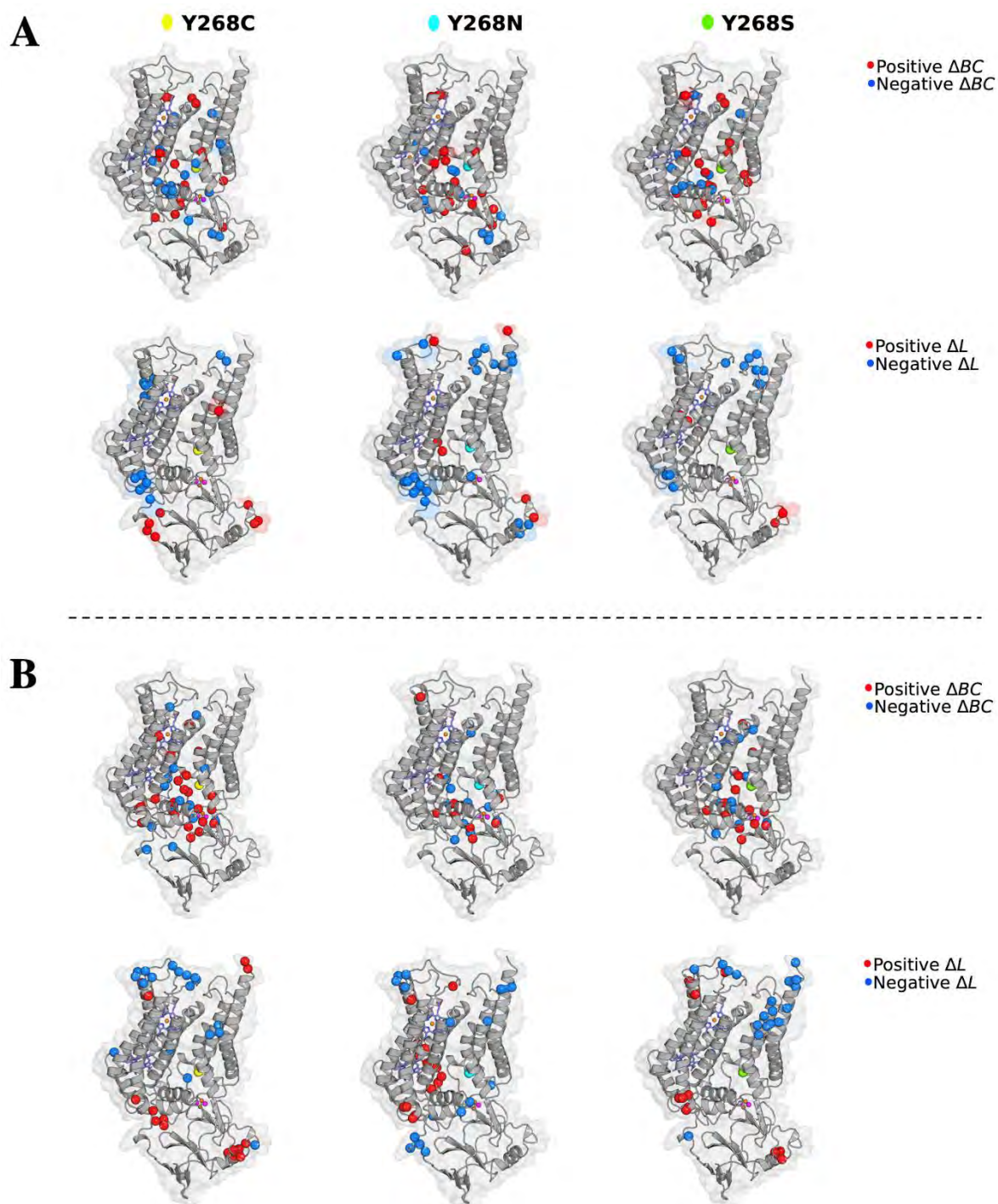


Figure 5.7: Effect of the three mutations Y268C, Y268N and Y268S on the protein intra-communication in the PfCytb-ISP protein. (A) PfCytb-ISP holo protein (B) PfCytb-ISP ATQ-bound protein. Residues exhibiting a significant change in average *betweenness centrality* (ΔBC) are mapped onto the structure. ΔBC was calculated as WT (wild-type) — mutant. Therefore, the red spheres represent residues with positive ΔBC values, i.e., residues that lost centrality in the mutant model relative to the WT. The blue spheres represent residues with negative ΔBC values, i.e., residues that gained more connectivity in the mutant models relative to the WT. (B) Residues exhibiting a significant change in average accessibility (ΔL) are mapped onto the structure. ΔL was calculated as WT — mutant. Therefore, the red spheres represent residues with positive ΔL , i.e., residues that increased accessibility in the mutant model relative

to the WT. The blue spheres represent residues with negative ΔL values, i.e., residues that decreased accessibility in the mutant model relative to the WT. Reproduced from [141].

Residues Trp262 and Tyr263, which form the PEWY motif, are located near the point mutations (Y268C/N/S). As such, residues with compromised connectivity in this region are likely to affect substrate binding. Literature confirms that the effect of mutations in close proximity to the PEWY motif would impair the redox process [324]. It is evident that a similar pattern was observed across the mutant systems focusing on residues with high BC values. This suggests the mutant systems are likely to disrupt protein communication as most of them have reported active site residues losing connectivity compared to their respective WT systems. The residues that gained connectivity in ATQ-bound systems included: Phe123, (Pro260 and Phe264), and Tyr263 in the Y268C, Y268N and Y268S mutant systems, respectively, which could indicate compensatory mutant mechanisms. Comparing with weighted residue contact map analysis (refer to **Section 5.3.5**), these findings are in good agreement, where residues Val140 and Val259 were observed to have reduction and loss of contacts across the mutant systems. This provides more insights into the effect of these mutations on protein function and ATQ activity.

Residue accessibility was investigated using average L , where those with lower L values indicated stability. The residues that were reported to have lower L values (< 0.25) included: 116, 119, 122–123, 126, 137, 140–144, 260–269 and 271–272, out of which $> 95\%$ were Qo active site residues and formed part of conserved motifs 2 and 3. This feature further supports the fact that these residues reported increased residue accessibility within the active site. Besides these residues, metal coordinating residues His78, His92, His173, His187, Cys299, His301, Cys317 and His320 were additionally reported to have lower L values. These were observed to correspond to the Qo substrate binding site as well as regions of catalytic importance (**Figure 5.6B** and **Figure 5.7B**), which is in good agreement with previous studies

[321]. It was noted that both WT and mutant systems displayed a similar pattern (high (+) L values), suggesting that residue accessibility around catalytic sites is maintained even in the presence of mutations. These findings were observed to directly correlate with residues having high residue fluctuations displaying the highest average L values (**Figure 5.6B,C**) which correspond to the protein's loop regions, a previously reported correlation [325]. The residues showing major changes in L for both holo and ATQ-bound systems were reported (**Tables S5.5 and S5.6**). Interestingly, residues Cys214, Trp221 and Ile222 were observed to have low (−) ΔL values in the holo protein of Y268C and Y268N mutant systems. These residues are located at the heterodimer interface, specifically at the C-terminal regions of the ISP chain. This evidently suggests the possibility of the residues being less accessible, showing that the mutations have a potential of destabilising the protein specifically at the heterodimer interface. However, in ATQ-bound systems, some interface residues were reported to have high (+) ΔL values in all mutant systems across at least two runs, showing their availability for interaction.

PART II

TARGET 1: Qo ACTIVE SITE

To investigate the effect of ligand (hit compounds) binding on the protein function and stability, the residue fluctuations, the RMSD of the protein backbone, center of mass and hydrogen bond distribution of the MD trajectories for the protein-ligand complexes in PART II and III were analysed.

5.3.7 Docking analysis for Qo active site; Potential hits identified through virtual screening

The docking experiment was validated by redocking co-crystallised ligand stigmatellin (SMA) of *Bos taurus* (PDB ID: IPP9) located in the Qo active site to reproduce its original docked pose. Additionally, the SMA ligand was cross-docked into the PfCytb-ISP protein model for validation purposes. From the control docking experiments, SMA had a good binding energy of -8.7 kcal/mol, and the docking poses are shown in **Figure S5.10**, where the molecular structural overlay of the two docking experiments reported a similarity score of 0.83. After molecular overlay, the head groups of the SMA ligand in the redocked and cross-docked experiments were identical, with few differences noted in the tail group due to the different positions assumed by the rotatable bonds. A total of 1,020 compounds were docked, and the compounds with the best docked poses as well as lowest energy conformation were selected for further selective inhibition analysis.

To identify selective inhibitors, the compounds having a binding energy of < -8.7 kcal/mol were selected, and their ligand interactions evaluated. This binding energy cut-off was applied to identify compounds with better binding affinity compared to SMA (co-crystallised drug: -8.3 kcal/mol) and ATQ drug which was previously reported to bind at -8.7 kcal/mol [141]. A total of 38 inhibitors (19 SANCDB and 19 MMV compounds displayed in **Table S5.5**) were

bound in the PfCytb-ISP Qo active site, while binding elsewhere (outside the Qo active site) in the HsCytb-ISP protein. From these, heterodimer interface analysis was performed on the HsCytb-ISP protein to evaluate whether the 38 drug compounds were binding in its interface. Compounds that bind in the interface are likely to interfere with important protein-protein interactions in the human protein. As shown in **Figure 5.8**, 9 out of the 38 selective inhibitors were not only found to bind in the vicinity of the interface but also formed inter-molecular interactions, including hydrogen bonds, with the protein. Thus, a total of 29 compounds remained for further analysis as the 9 were excluded.

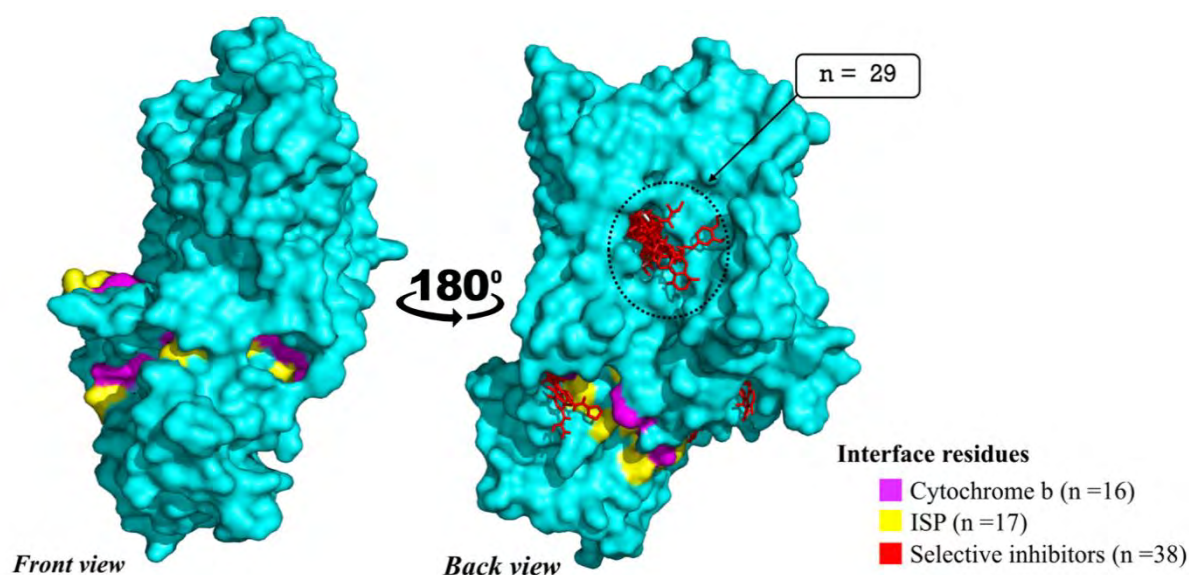


Figure 5.8: A surface representation of the interface residues in the HsCytb-ISP protein model and the 29 selective inhibitors included in MD simulation runs for further analysis. The protein shown as surface presentation is coloured in cyan, where heterodimer interface from Cytb is coloured in purple while ISP is coloured in yellow.

5.3.8 Molecular dynamics trajectory analysis: Eight potential hits identified out of 29 selective compounds via MD simulations

Due to the importance of the Fe^{2+} metal in the protein's function, previously derived force field parameters [141] were applied in the protein models before the MD simulation experiments. The charge parameters were applied to ensure that the metal coordinating sphere and

geometries were preserved during simulations. 350 ns MD simulations were performed where the PfCytb-ISP integral membrane protein was embedded into a POPE: POPC phospholipid bilayer. MD simulations were performed on a total of 29 potential hit compounds on the WT model after which 15 were unstable hence eliminated from further analysis. Thereafter, MD simulations were performed on the remaining 14 compounds assessing their stability not only in WT but also on the frequently reported PfCytb point mutations. Therefore, a total of 14 potential compounds in the WT as well as mutant systems (Y268C, Y268N and Y268S) equivalent to $14 * 4 = 56$ systems were analysed for stability. After 350 ns MD simulations, a total of eight compounds (1-4 (MMV007638, MMV019551, MMV659004, MMV676444), 6-8 (SANC00373, SANC00459, SANC00548) and 14 (SANC00725) were stable in WT and mutant systems (Y268C, Y268N and Y268S) remained stable in the PfCytb-ISP Qo active site in WT (**Figure S5.11A**) and mutant proteins (**Figure S5.11B**). Thereafter, to evaluate the allosteric property of these compounds, MD simulations were performed on the eight compounds bound outside the Qo active site in the HsCytb-ISP protein. None of them was stable in the human protein, thus the eight compounds were regarded as potential selective hit compounds.

Ligand interactions for the eight compounds that remained stable in WT and all the mutant proteins were assessed as shown in **Figure 5.9**.

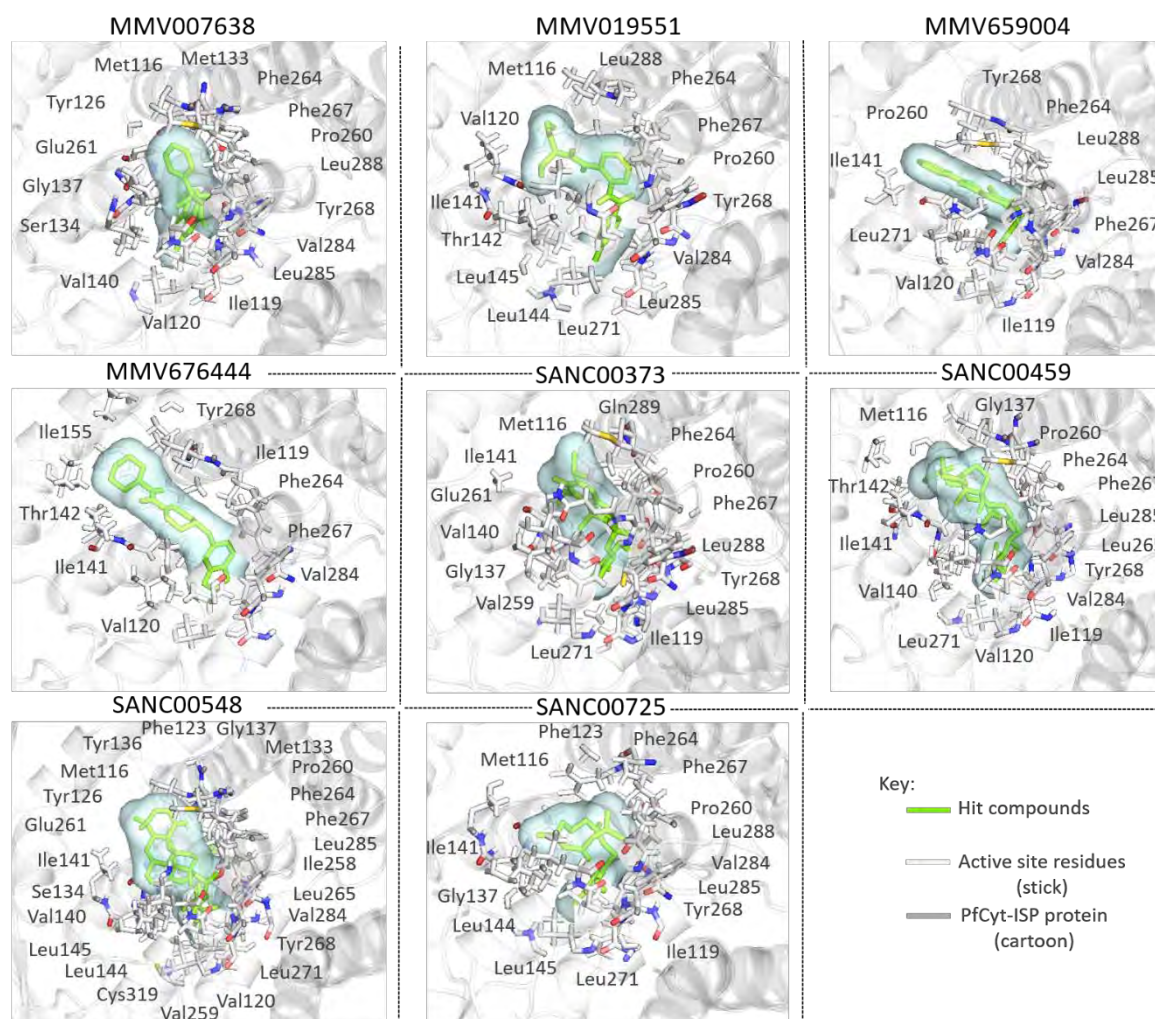


Figure 5.9: 3D ligand interaction network for eight PfCytb-ISP selective inhibitors in the Qo active site pocket: (A) MMV007638, (B) MMV019551, (C) MMV659004, (D) MMV676444, (E) SANC00373, (F) SANC00459, (G) SANC00548, (H) SANC00725. The protein in cartoon representation is coloured dark grey while the ligands and the interacting active site residues are displayed in stick representations coloured in green and grey, respectively.

According to previous studies, the active site residues found in the PfCytb domain included: Met116, Ile119, Val120, Phe123, Val124, Met133, Trp136, Gly137, Val140, Ile141, Thr142, Leu144, Leu145, Ile155, Phe169, Leu172, Ile258, Val259, Pro260, Glu261, Trp262, Tyr263, Phe264, Phe267, Tyr268, Leu271, Val 284 and Leu285, while only His320 was located on the ISP domain [107]. From the docking results, it was further noted that the eight compounds formed interactions with most of the active site residues as reported previously [107]. Exceptions to note included MMV007638, SANC00373 and SANC00548, which formed

interactions with residue Glu261. Previous studies reported Glu261 residue to be essential in the formation of a hydrogen bond that supports substrate binding [220,312]. All these ligand systems were monitored and assessed after MD to determine whether they formed sufficient interactions for their stability (further detailed in Section 5.3.14).

Even though van der Waals (vdW) forces are the weakest of all inter-molecular interactions, this study reports that they account for most interactions observed across the eight ligand systems. Despite being weak forces, it has been reported that vdW forces accumulate together between two objects to form a stronger energy force that supports ligand binding [326,327]. In this case, these forces are responsible for stabilising the ligand in its active site pocket. Evidently, this is well supported by the strong binding energies seen after docking experiments.

Table 5.1: Chemical information and pharmacokinetic profile of the selective inhibitors derived from MMV and SANCDB compounds.

Compound ID	Chemical Formula	Pharmacokinetic properties					Binding Energy (kcal/mol)	PAINS # alerts
		MW	nRB	HBA	HBD	iLogP		
MMV007638	C20H20N2O4S2	416.5	8	5	1	2.4	-9.5	0
MMV019551	C20H22N2O4	354.4	9	4	2	2.8	-8.9	0
MMV659004	C21H21ClN4	364.9	4	3	0	3.7	-9.9	0
MMV676444	C19H27N3O3	345.4	5	4	1	3.7	-8.8	0
SANC00373	C27H30O6	450.5	5	6	3	3.6	-9.7	1
SANC00459	C33H50O8	574.8	7	8	2	4.1	-9.3	0
SANC00548	C36H58O9	634.8	8	9	6	3.8	-8.9	0
SANC00725	C22H32O3	344.5	9	3	2	4.1	-8.3	0

The binding energy scores and the chemical structures of the eight selective compounds are shown in **Table 5.1** and **Figure 5.10**, respectively.

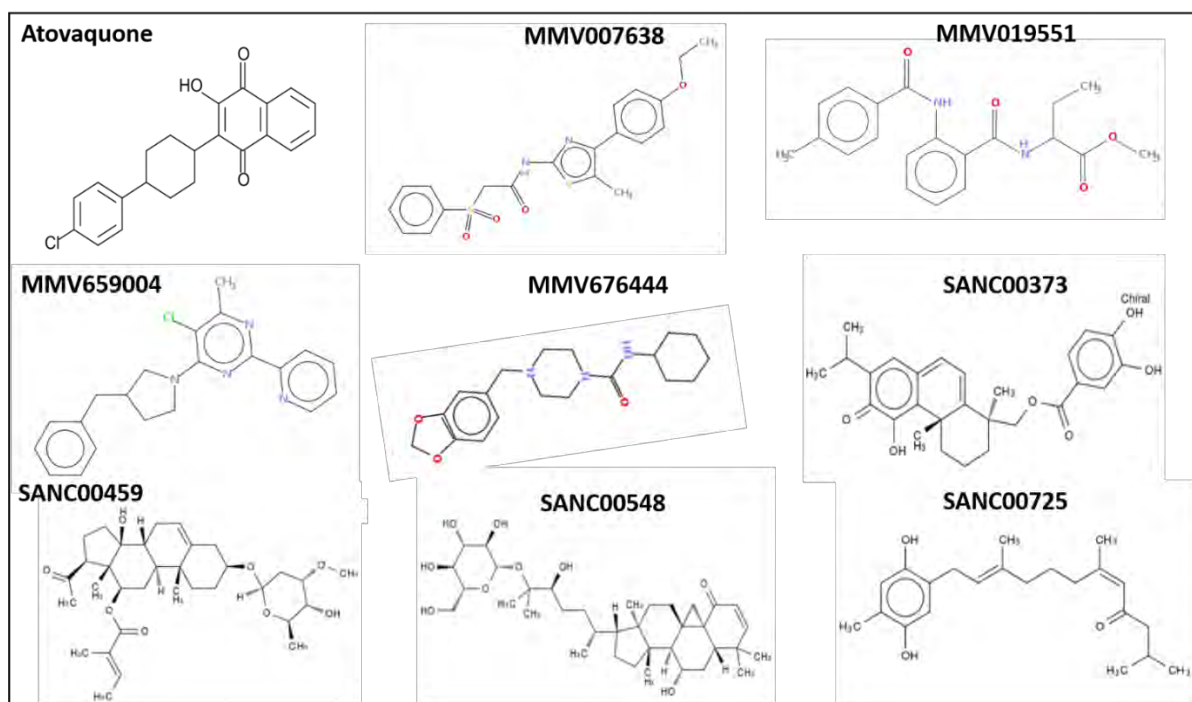


Figure 5.10: 2D structures of the ATQ (reference) MMV and SANCDB compounds. MMV and SANCDB compounds were docked onto both PfCytb-ISP and HsCytb-ISP enzyme and were identified as selective hit compounds.

5.3.9 General information on the potential hit compounds

A summary of the general chemical information of the hit compounds and their chemical structures is shown in **Table 5.1** and **Figure 5.10**, respectively. From this, MMV compounds (MMV007638, MMV019551, MMV659004 and MMV676444) were retrieved from the Malaria/Pathogen box where they have been tested and they showed antimalarial activity against *P. falciparum* 3D7 [328]. For SANCDB compounds, SANC00373 is a terpenoid extracted from the *Plectranthus purpuratus* plant (spp. tongaensis) and its medicinal usefulness is currently unknown [329]. SANC00459 is a steroid extracted from *Hoordia gordonii* plant that has been previously used as an appetite suppressant [330]. SANC00548, a sutherlandioside D, is a cyclooctane extracted from *Sutherlandia frutescens* plant and has multipurpose medicinal uses including treatment of cancers, eye problems and many other diseases [331].

SANC00725 is hydroquinone extracted from *Leminda millecra* Griffiths (spp. Arminacea) plant [332].

The Pan-assay interference compounds (PAINS) filters were checked to assess whether the potential hit compounds inhibit multiple drug targets. The SWISSADME output results in **Table 5.1** revealed that the eight potential hits were suitable drug candidates except for PAINS alert on compound SANC00373. This compound has been previously reported to bind (potential hit compound) cysteine proteases of the *Plasmodium* parasite, also known as falcipains [333]. The Pan-assay interference compounds (PAINS) filters were checked to assess whether the potential hit compounds inhibit multiple drug targets. The SWISSADME output results in **Table 5.1** revealed that the eight potential hits were suitable drug candidates except for PAINS alert on compound SANC00373. This compound has been previously reported to inhibit cysteine proteases of the *Plasmodium* parasite, also known as falcipains [333].

5.3.10 Global protein dynamics analysis: Root mean square deviations (RMSD) and radius of gyration (Rg)

A comparative analysis was performed on the eight ligand systems (WT and mutant states) in the presence and absence of POPE: POPC bilayer to evaluate the importance of performing MD simulations in a phospholipid membrane. Here, 350 ns of MD simulations was performed with and without phospholipid bilayer as a proof of concept to analyse protein stability. Interestingly, lower RMSD values were observed in the ligand systems setup embedded in a membrane as shown in **Figure 5.11**. This evidently showed that the experiments performed with protein embedded in a membrane portray better system stability than those without the membrane.

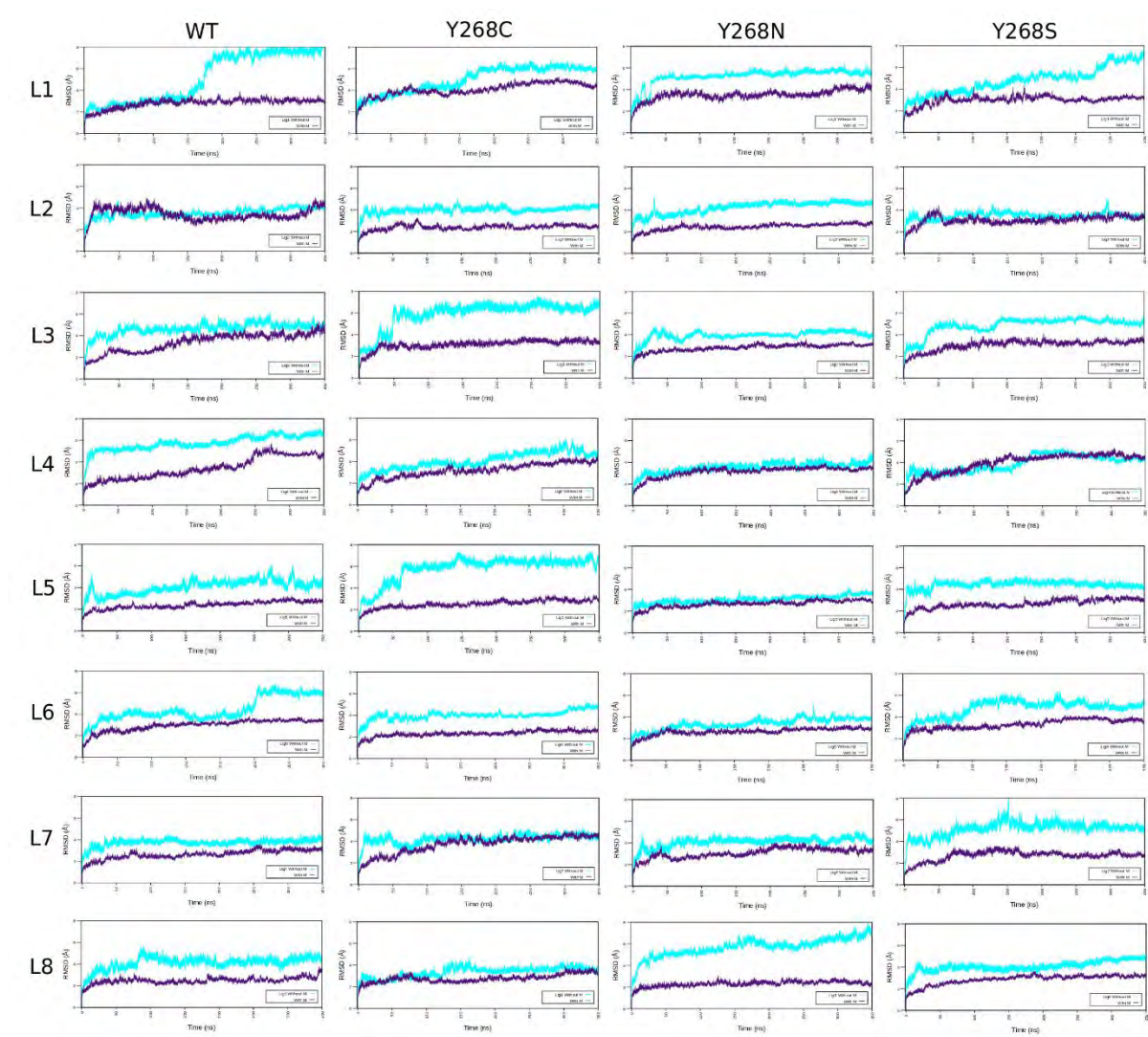


Figure 5.11: RMSD Line graph presentation of eight ligand systems (Lig1: MMV007638, Lig2: MMV019551, Lig3: MMV659004, Lig4: MMV676444, Lig5: SANC00373, Lig6: SANC00459, Lig7: SANC00548, Lig8: SANC00725) comparing MD simulations performed with (coloured in navy blue) and without (coloured in cyan) phospholipid membrane.

Comparative RMSD analysis was performed after the 350 ns MD simulations where **Table 5.2** showed mean RMSD values of 2.65, 2.73, 2.63 and 2.86 Å in all the eight ligand systems across the WT and Y268S, Y268N and Y268C systems, respectively. In comparison with the WT, the mean values showed that Y268C and Y268S mutant systems converged at a higher RMSD.

Table 5.2: Mean protein backbone RMSD values for WT and mutant (Y268C, Y268N and Y268S) ligand-bound systems. Values in Å.

	LIGAND BOUND SYSTEMS			
Ligand systems	WT	Y268C	Y268N	Y268S
MMV007638	2.80	4.09	3.51	2.99
MMV019551	3.44	2.38	2.44	3.07
MMV659004	3.38	3.10	2.81	3.07
MMV676444	3.42	3.18	3.14	3.99
SANC00373	2.39	2.52	2.66	2.66
SANC00459	2.97	2.30	2.65	3.32
SANC00548	2.63	3.71	3.06	2.68
SANC00725	2.48	2.76	2.30	2.85
MEAN	2.93	3.00	2.82	3.08

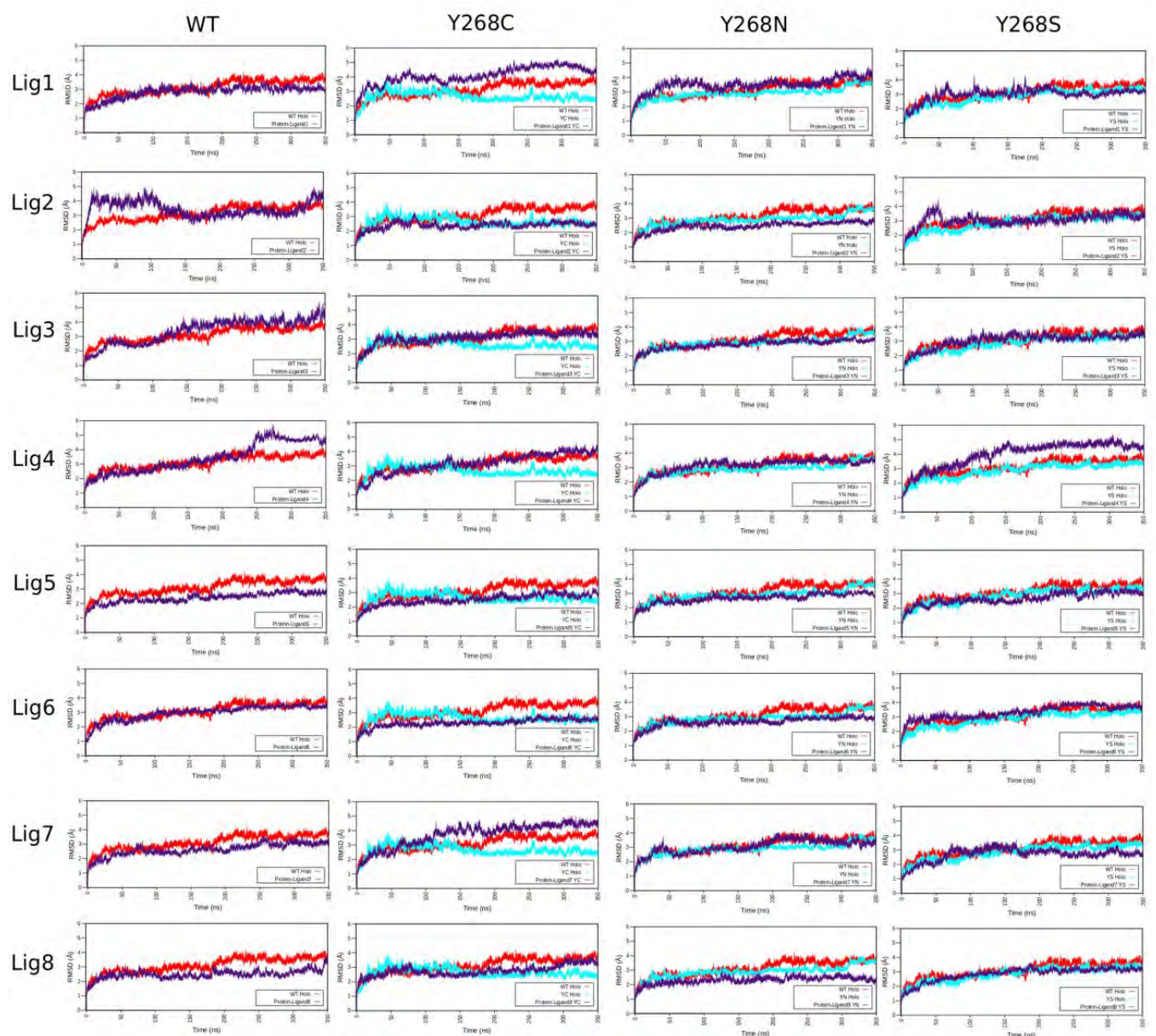


Figure 5.12: RMSD density distribution for PfCytb-ISP holo and 8 ligand systems in WT, Y268C, Y268N and Y268S states across 350 ns MD simulations. The ligand systems are as follows: Lig1 (MMV007638), Lig2 (MMV019551), Lig3 (MMV659004), Lig4 (MMV676444), Lig5 (SANC00373), Lig6 (SANC00459), Lig7 (SANC00548) and Lig8 (SANC00725). A colour scale is provided.

To further understand the protein conformations of all holo and ligand systems presented in line graphs (**Figure 5.12**), present in the eight ligand systems, RMSD distribution in the form of histogram plots was compared across the WT and mutant (Y268C, Y268N and Y268S) systems. The changes in conformations across 350 ns MD simulations were assessed comparing PfCytb-ISP holo protein (protein with cofactors only; ligand-free) and ligand-bound

systems; the holo protein was used as the reference (**Figure 5.13**). In comparison to PfCytb-ISP holo protein, all the eight ligand systems in the WT and mutant systems sampled both unimodal and bimodal distributions. Focusing on major conformations and the mean indicated by the dotted lines in **Figure 5.13A**, there was no significant difference in mean RMSD values of the ligand systems compared to their holo proteins. Also, the histogram plots showed that most ligand systems assumed a normal distribution in terms of conformations sampled. This showed that the protein backbone conformation was not affected by ligand binding.

However, several ligand systems displayed bimodal distributions, including but not limited to Lig7 (SANC00548) of the Y268C system, which sampled two distinct conformations as shown in **Figure 5.13A**. Here, the pairwise correlation between the holo and ligand-bound systems of WT, Y268C, Y268N and Y268S states was calculated to evaluate the differences in the systems. In the WT, most of the systems were positively correlated ($r = 0.64 - 0.82$) except Lig2 (MMV019551) and Lig8 (SANC00725) systems that displayed low correlation values of 0.30 and 0.49, respectively. In the Y268C system, most of the systems had correlation values > 0.50 except Lig4 (MMV676444), Lig5 (SANC00373), Lig7 (SANC00548) and Lig8 (SANC00725) systems that displayed low correlation values of 0.35, 0.46, 0.27 and 0.47, respectively. In the Y268N system, most of the ligand systems positively correlated ($r = 0.73 - 0.87$). Lastly, the Y268S system showed that most ligand systems positively correlated ($r = 0.68 - 0.91$) except Lig2 (MMV019551) that displayed a correlation value of 0.37. Those that displayed positive correlation denote that the systems were positively correlated; hence there were no distinct differences in the major conformations sampled during simulation. The unimodal conformations reported in most ligand systems confirm that they all achieved a stable conformation across MD simulation. It also suggests the high likelihood of the ligands remaining bound in the active site pocket (further detailed in the ligand RMSD analysis). These results showed the preliminary effect of ligand binding as well as mutation on the protein

conformations. This will be further elucidated by both global and local residue network analysis.

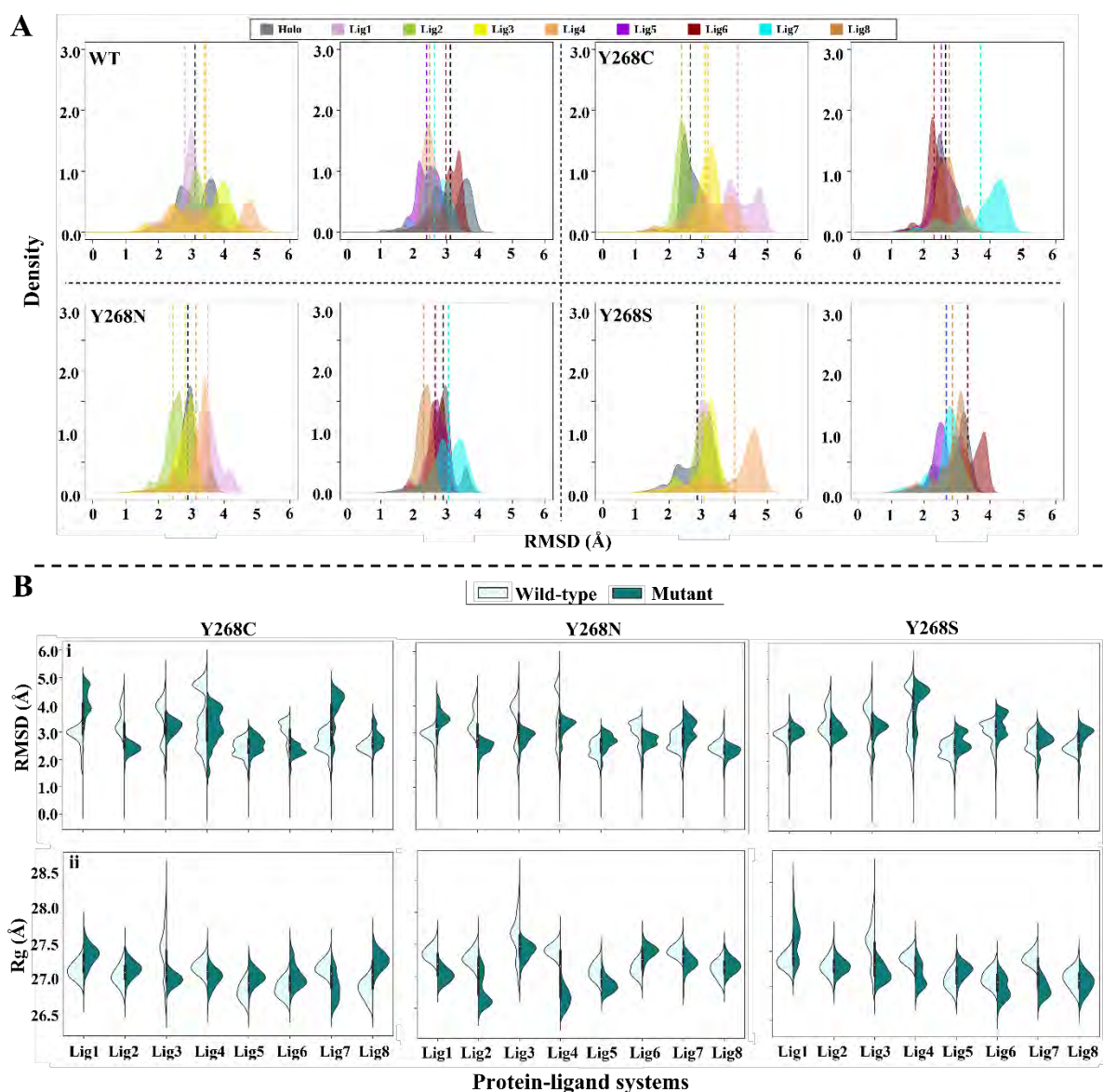


Figure 5.13: (A) Histogram plot showing backbone RMSD density distribution for PfCytb-ISP holo and 8 ligand systems in WT, Y268C, Y268N and Y268S states across 350 ns MD simulations. The y and x-axis represent the number of conformations versus the number of times specific conformation is sampled, respectively. The ligand systems are as follows: Lig1 (MMV007638), Lig2 (MMV019551), Lig3 (MMV659004), Lig4 (MMV676444), Lig5 (SANC00373), Lig6 (SANC00459), Lig7 (SANC00548) and Lig8 (SANC00725). The respective coloured dotted lines represent mean RMSD values for the ligand-free and ligand-bound systems. (B) Violin plot presentation of i) RMSD and ii) Rg for WT and mutant ligand-bound systems during 350 ns MD simulations; WT in light cyan and MT in dark cyan colour.

Further, the effect of mutations on the protein backbone was analysed across all eight ligand systems comparing the WT and mutant ligand-bound systems as displayed in **Figure 5.13B i)**. Here, the comparative analysis of the WT versus mutant models showed unimodal distribution in all mutant complexes except in ligand system 7 (SANC00548) in Y268C, ligand system 4 (MMV676444) in Y268N, and ligand systems 3 (MMV6590004), 4 (MMV676444), and 7 (SANC00548) in Y268S mutant systems. Backbone distribution across the three mutant systems showed similar major conformations distributed at approximately 3 Å. However, in terms of the RMSD distributions across the eight ligand systems, we observed that Y268C and Y268N systems displayed similar patterns, unlike the Y268S systems that sampled a unique profile in some ligand systems. This suggests that the three mutations could possibly have varied effects on the protein backbone.

Figure 5.13B ii) shows the assessment of the protein compactness for the eight ligand systems in both WT and mutant states. This indicator allows assessment of the distribution of atoms around the centre of mass (COM) of the protein to both of its termini. WT ligand systems compacted between mean Rg range of 27.0 – 27.5 Å, while their respective mutant states converged within the range of 26.7 – 27.8 Å. Compared to corresponding WT systems, all mutant systems compacted at a lower Rg value except in ligand systems 5 (SANC00373) and 8 (SANC00725) of Y268C, ligand systems 6 (SANC00459) and 8 (SANC00725) in Y268N, and ligand systems 1 (MMV007638) and 8 (SANC00725) in Y268S systems as shown in **Figure 5.14**.

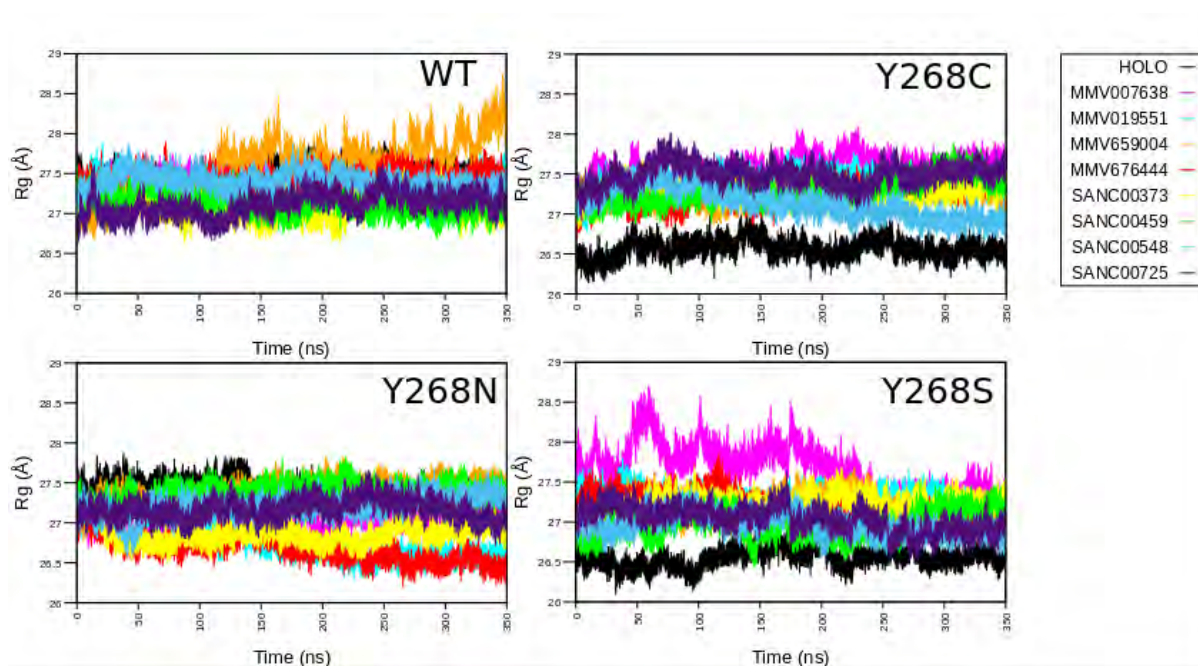


Figure 5.14: Radius of gyration line plots for PfCytb-ISP holo and eight ligand systems in WT, Y268C, Y268N and Y268S systems. A colour scale has been provided.

A comparison with the holo protein showed that Y268C and Y268S mutant proteins experienced changes, unlike the Y268N mutant protein. Despite this observation, the eight ligand systems are reported to be compact, with the holo protein being more compact.

The results depicted compactness in the mutant systems and suggested that the binding of the ligands stabilised the systems as observed by the reduced Rg values compared to the WT proteins.

5.3.11 Principal component and free energy analysis

Comparison of the structural conformational changes of the holo and protein-ligand systems (MMV and SANCDB hit compounds) in their WT and mutant states (Y268C, Y268N and Y268S) is shown in **Figure 5.15**. Here, the figures showed the plots of PC1 and PC2 alongside a mapping of the free energies of the sampled conformations. As displayed in **Table 5.3**, PC1 represents dominant motions that occupied the largest variation (approximately > 50 %) followed by PC2 (approximately > 20 %). With respect to the free energies, the conformations

that displayed low energy wells were likely to contribute favourably to the overall stability of the protein systems.

Table 5.3: Eigenvalues in fractions representing PC1, PC2 and PC3 for each of the eight ligand systems in both WT and mutant systems; Lig1 (MMV007638), Lig2 (MMV019551), Lig3 (MMV659004), Lig4 (MMV676444), Lig5 (SANC00373), Lig6 (SANC00459), Lig7 (SANC00548) and Lig8 (SANC00725).

Systems		Lig1	Lig2	Lig3	Lig4	Lig5	Lig6	Lig7	Lig8
WT	PC1	0.64	0.67	0.65	0.60	0.56	0.69	0.51	0.46
	PC2	0.23	0.17	0.24	0.31	0.27	0.17	0.28	0.29
	PC3	0.13	0.15	0.11	0.10	0.17	0.14	0.21	0.26
Y268C	PC1	0.59	0.50	0.44	0.59	0.56	0.63	0.60	0.44
	PC2	0.24	0.28	0.36	0.26	0.25	0.24	0.29	0.38
	PC3	0.17	0.22	0.20	0.15	0.18	0.13	0.11	0.18
Y268N	PC1	0.56	0.63	0.54	0.73	0.74	0.66	0.58	0.53
	PC2	0.28	0.24	0.30	0.18	0.14	0.21	0.27	0.30
	PC3	0.15	0.13	0.17	0.09	0.13	0.13	0.15	0.17
Y268S	PC1	0.50	0.47	0.44	0.82	0.43	0.63	0.69	0.59
	PC2	0.28	0.33	0.35	0.11	0.39	0.21	0.19	0.29
	PC3	0.21	0.20	0.21	0.07	0.18	0.16	0.13	0.12

In the holo form, one major conformation of low energy was sampled in the WT as well as in all mutant systems. These conformations were mainly sampled between -20 and -40 along PC1 in the WT, Y268C and Y268S systems. However, the Y268N system was observed to sample a different conformation along PC1 from 20 to 40. Interestingly, along PC2, both WT and all the mutant systems displayed a similar profile where the major conformation was observed between -10 and 10. This similar profile in the WT and mutant systems suggests that in the presence of these point mutants, the protein's stability is likely to be maintained, as evidenced by the low free energy conformations.

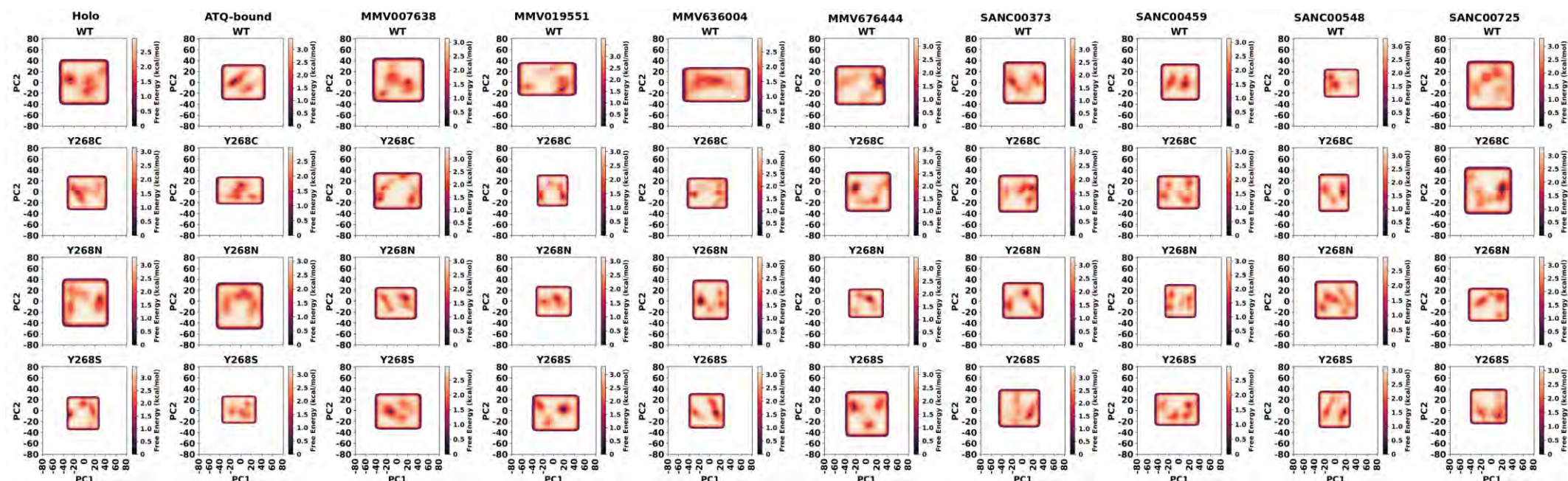


Figure 5.15: 3D Principal component analysis (PCA) plot for holo and eight protein-ligand systems (MMV and SANCDB potential hits). PC1 versus PC2 are plotted as a function of free energy represented as kcal/mol.

Upon binding of each of the four MMV compounds, the complexes displayed a major conformation of low energy in their WT states. However, we noted that in the mutant states, all the systems sampled the lowest free energy conformation (indicated in black colour) except for the MMV007638 ligand in the Y268N mutant system.

Focusing on the dominant motion (PC1) in each ligand system, the major conformations were sampled from -20 to 0 (MMV019551 and MMMV656004) and from -40 to -20 (MMV007638 and MMV676444) in their WT state. In the mutant state, all the four ligand systems sampled the conformations between 0 to -40. This is consistent with previous findings where ATQ-bound systems displayed similar conformational sampling patterns [141].

Comparative analysis of the holo and SANCDB protein-ligand complexes demonstrated unique profiles in each of the systems. Of interest were the low energy conformations sampled in all the four ligand systems across WT and mutant states, except in several systems. These include WT (SANC00548 and SANC00725), Y268N (SANC00548 and SANC00725), and Y268S (SANC00373) ligand systems. In the WT state, ligand systems SANC00373 and SANC00459 sampled one distinct protein conformation along PC1 from 0 to -40. Even though systems SANC00548 and SANC00725 failed to sample native state conformations (no lowest energy wells), their conformational sampling profile was similar to the holo protein. In the mutant state, ligand systems SANC00459 and SANC00725 showed similar conformation profiles in Y268C and Y268S, where the lowest energy wells were sampled equally along PC1 from (0 to 20) and (-20 to 0), respectively. This observation could suggest that Y268C and Y268S share similar resistance mechanisms as previously reported [141].

On the other hand, ligand systems SANC00373 and SANC00548 displayed different conformational free energies in their mutant states. Ligand system SANC00373 displayed low energy conformations from -20 to 20 along PC1, which occupied similar conformation from 0 to -20 along PC2 in Y268C and Y268N states while Y268S lacked low energy conformation. Ligand system SANC00548 sampled two conformations in Y268C, where the major conformation was sampled from 0 to -40 along PC1 in both Y268C and Y268S. The two conformational clusters were also evidently observed in the RMSD results (**Figure 5.13**). In

the Y268N mutant state, the same system also failed to sample the native state conformation although similar conformational patterns were observed in Y268N states. As such, these findings suggest that even though these three mutations occur in the same position, their mechanisms of action could be different.

Overall, the results agree with RMSD results reported in **Figure 5.13**, where dominant low energy conformations were reported in the eight protein-ligand complexes. These results support the likelihood that these drug compounds may stabilise the protein not only in WT but also in the presence of mutations.

5.3.12 Ligand stability during MD simulations

To ascertain the selectivity of the potential hits on HsCytb-ISP, all the eight ligands were unstable in the HsCytb-ISP Qo active site during 350 ns MD simulation, thus indicating their selectivity against the *Plasmodium* parasite. These ligands were confirmed to remain in the Qo active site throughout 350 ns MD simulation as shown in **Figure 5.16**. Ligand RMSD distribution (**Figure 5.17**) was calculated showing all the eight hit compounds (MMV007638, MMV019551, MMV659004, MMV676444, SANC00373, SANC00459, SANC00548 and SANC00725), and reported a mean range of 0.75 - 3.2 Å across both WT and mutant systems.

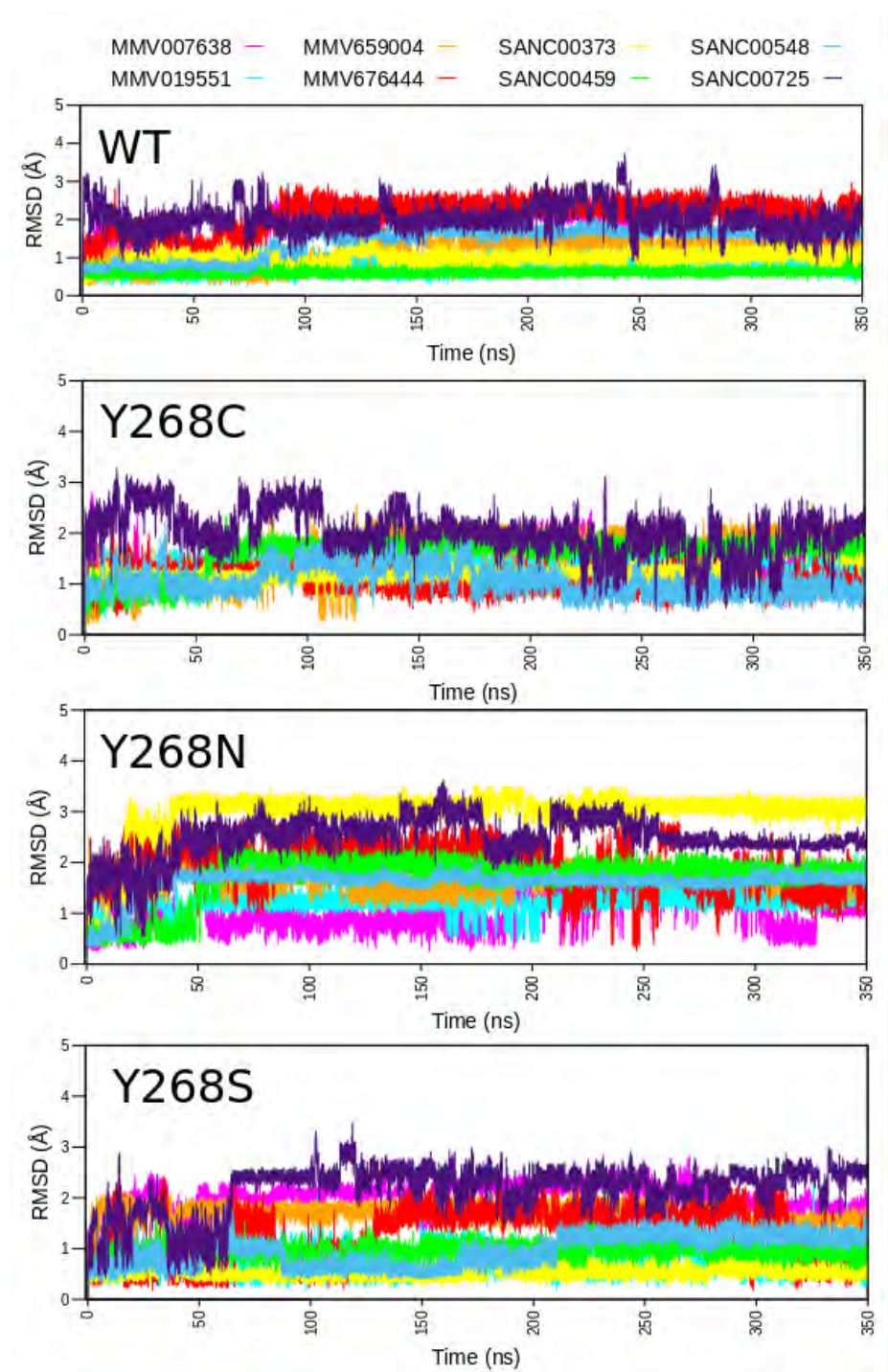


Figure 5.16: Ligand root mean square deviation (RMSD) for eight ligand systems (MMV007638, MMV019551, MMV659004, MMV676444, SANC00373, SANC00459, SANC00548, SANC00725) in WT and mutant states (Y268C, Y268N and Y268S) during 350 ns MD simulation.

However, violin representation of the ligand RMSD in **Figure 5.17** showed that the eight hit compounds behaved differently in the Qo active site based on the conformations sampled

during simulation, thus they were grouped into two classes. The two classes represent the ligands' degree of stability with reference to the WT system and their stability in their respective mutant systems. Therefore, class I consisted of MMV019551, SANC00373 and SANC00459, while class II consisted of MMV007638, MMV659004, MMV676444, SANC00548 and SANC00725. Compounds in class I were reported to be stable in the WT PfCytb-ISP protein as well as in Y268C, Y268N and Y268S mutant states during the entire simulation. These three compounds sampled one ligand conformation, portraying their stability in the active site.

On the other hand, compounds in class II sampled two ligand conformations due to the rotatable bonds that kept flipping and moving in the active site during the simulation. As shown in **Figure 5.16**, the jumps observed in some systems were due to ligand movements within the active site pocket. Also, it was noted that the rotatable bonds in the ligands allowed certain ligand structures to keep flipping/rotating during MD simulation while remaining bound in the active site. This suggests that compounds in class II are less stable than those in class I even though they are reported to remain bound in the active site.

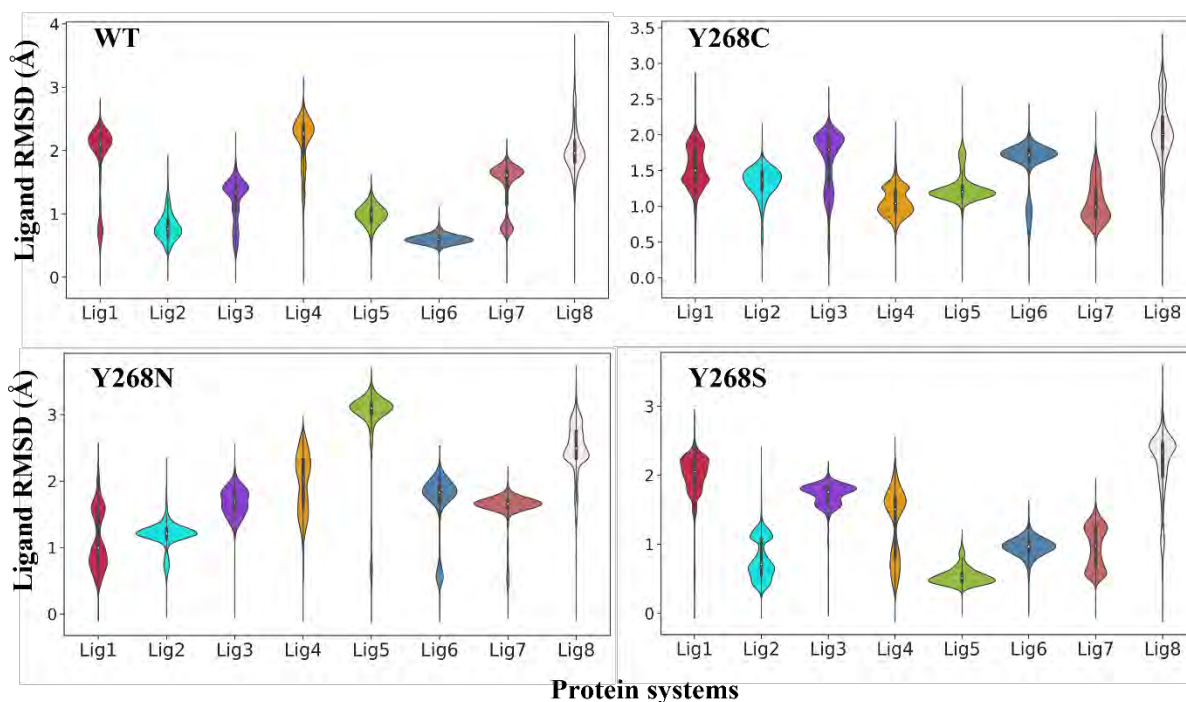


Figure 5.17: Ligand RMSD for eight ligands during 350 ns MD simulation performed in WT, Y268C, Y268N and Y268S systems as indicated in the figure. The ligands include (Lig1: MMV007638, Lig2: MMV019551, Lig3: MMV659004, Lig4: MMV676444, Lig5: SANC00373, Lig6: SANC00459, Lig7: SANC00548, Lig8: SANC00725).

Additionally, compared to the WT system, the ligands had varied RMSD deviations across the mutant systems where similar patterns were observed in the Y268C and Y268S systems. In the Y268N system, ligands MMV676444 and SANC00373 were observed to have increased ligand RMSD. Even though the Y268 mutations occur in the same residue position, their implication on ligand stability is shown to be different. This varied observation could possibly be because of the nature of the active site pocket in mutant systems, whose volume has previously been reported to be altered (increased) in the presence of mutations [62].

Similarly, the COM between each ligand and the active site was performed in the WT and mutant (Y268C, Y268N and Y268S) states. **Figure 5.18** further supports each ligand's stability, where the distance between ligand and active site was measured. Line (**Figure 5.18A**) and violin (**Figure 5.18B**) plot representation of the COM showed that the distance of the

ligand binding site did not vary during the 350 ns MD simulation, though some systems moved slightly away from the active site (**Figure 5.18A**). This confirms the ligands' similar degree of stability in the active site as reported in **Figure 5.17**. However, in the three mutant systems, there was an increased COM of up to 8 Å compared to the WT systems, for which most systems except that of MMV007638 reported COM of 6 Å. This further suggests the possibility of the mutations altering the active site pocket.

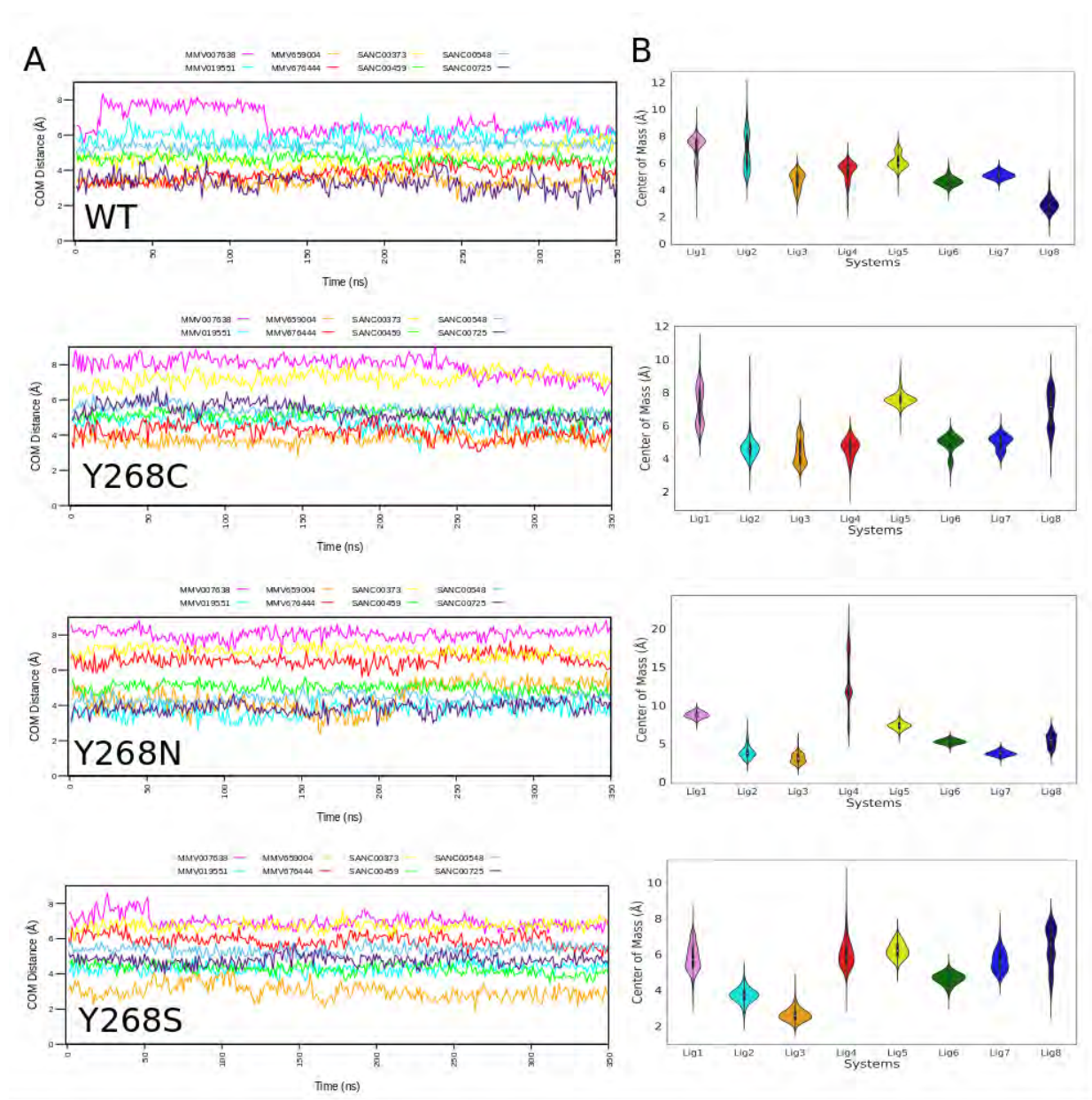


Figure 5.18: Centre of mass (COM) distance between ligand and the Qo active site. (A) Line graph showing COM across 350 ns MD simulation. (B) Violin plot representation of COM showing conformation sampling during MD simulation.

Overall, this trajectory analysis supports the fact that these eight compounds have the potential to inhibit the PfCytb-ISP protein. Based on the degree of stability, the findings also showed that class I potential hit compounds not only bound significantly strongly in WT but also in the presence of the three point mutations (Y268C, Y268N and Y268S).

5.3.13 Effect of the eight potential hit compounds on protein residue flexibility

Besides the global analysis of the changes occurring in the protein backbone, residue fluctuations on the C α atoms occurring during MD were investigated as shown in **Figure 5.19**. This was assessed in the ligand-free and ligand-bound states of the WT, Y268C, Y268N and Y268S systems. The results showed that the greatest RMSF changes were located towards the C-terminal of the PfCytb subunit and the intersect between the two chains (PfCytb and ISP), corresponding to loop 359-366 at the N terminal, which showed high fluctuations. This is because the connecting residues of the two heterodimers are far apart, distance-wise. Similarly, high fluctuations were observed in the loop regions (PfCytb chain: 21-27, 187-180, 207-213, 218-232, 265-285, 294-315, 334-339; ISP chain: 218-235, 238-243, 270-273, 277-282, 297-306, 317-343; *P. falciparum* numbering) of the PfCytb-ISP protein. Although distant from the substrate binding site, these loop regions are known to be highly flexible, accounting for the most fluctuations observed in **Figure 5.19**. The loop residues listed above account for 33 % of the total 503 protein residues that make up both chains A and B of the PfCytb-ISP protein.

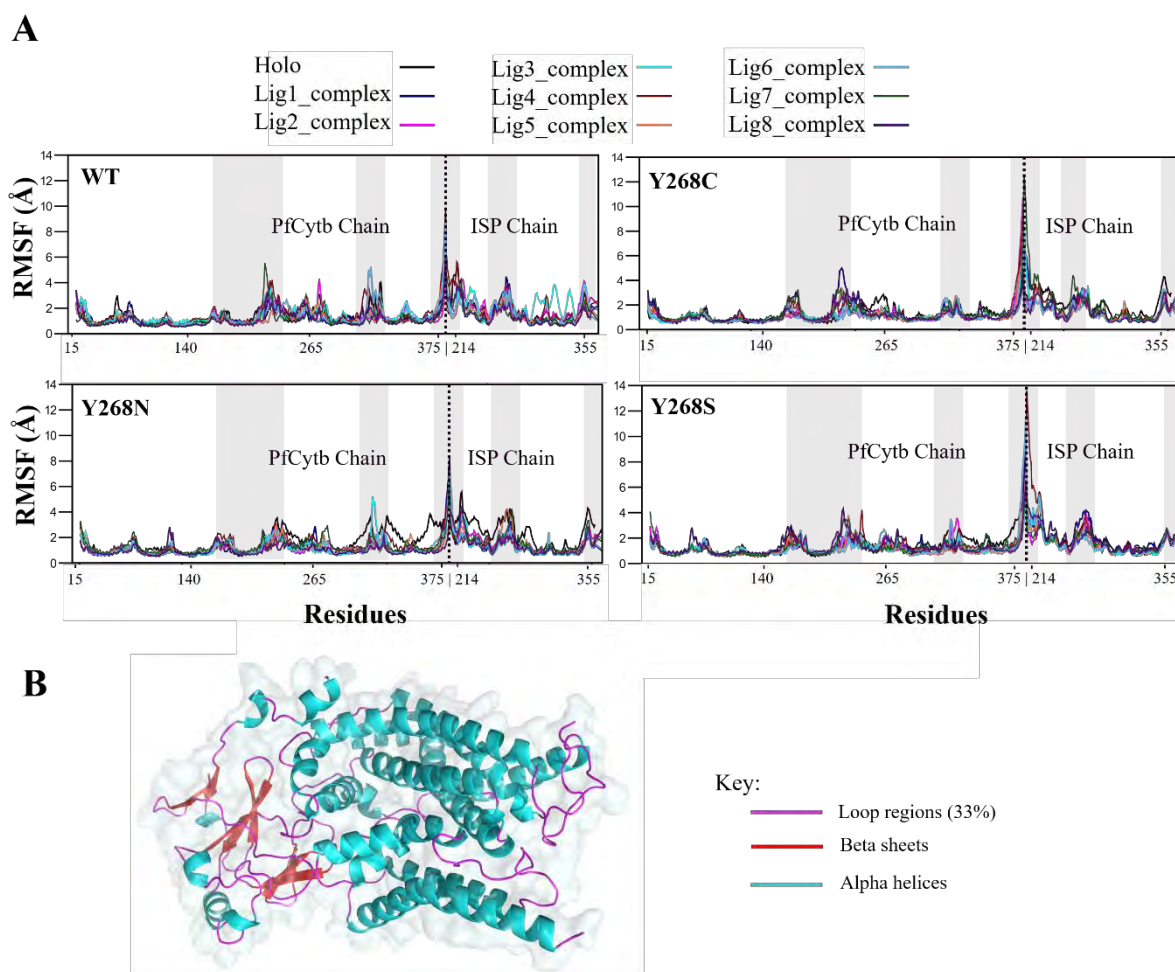


Figure 5.19: (A) Root mean square fluctuations comparing ligand-free and ligand-bound protein systems for 8 ligands performed in WT and mutant states (Y268C, Y268N and Y268S) during 350 ns MD simulation. (B) Cartoon representation of PfCytb-ISP protein showing secondary structures including flexible loops highlighted in purple colour.

A comparison was performed between the WT, Y268C, Y268N and Y268S holo structures and their corresponding ligand-bound systems. From this analysis, a similar pattern was observed across all ligand-free versus ligand-bound systems, specifically in the loop regions. However, upon ligand binding, there were some differences noted. Fluctuations in the ligand systems were observed where the eight ligand systems behaved differently in the WT and mutant states. For instance, Y268N ligand system eight portrayed increased fluctuations in regions 90-110, while in the WT, Y268C and Y268S, the same regions displayed decreased fluctuations.

Another interesting observation is the decreased fluctuations observed in the ISP chain of all the Y268C ligand systems as compared to the WT, Y268N and Y268S systems. This observation highlights that the mutations have a varied impact on per-residue flexibility. Overall, the eight ligand systems portrayed similar fluctuation patterns in both WT and mutant states. Worth noting is the observation that the active site residues in PfCytb chain (Met116, Ile119, Val120, Phe123, Val124, Met133*, Trp136*, Gly137, Val140, Ile141, Thr142, Leu144*, Leu145, Ile155*, Phe169, Leu172, Ile258, Val259, Pro260, Glu261, Phe264, Phe267, Tyr268, Leu271*, Val284, Leu285, Leu288*) in the unshaded regions, displayed decreased fluctuations. Thus, the overall function of PfCytb-ISP protein was likely maintained across the holo and ligand-bound protein systems.

5.3.14 Hydrogen bond, vdW, and other interactions maintaining the stability of protein-ligand complexes

To investigate protein-ligand structural stability for all eight ligand complexes in four states (WT, Y268C, Y268N and Y268S), molecular interactions such as vdW and hydrogen bonds were investigated during 350 ns MD simulation. The protein-ligand interactions in a total of 32 protein systems were assessed. The hydrogen bond, which is considered a weak inter-molecular interaction, has been well described in increasing ligand binding affinities [334,335]. The hydrogen bond distribution in all ligand systems varied considerably as shown in **Figures S5.12** and **S5.13**, representing MMV and SANCDB compounds, respectively.

MMV drug compounds in **Figure S5.12** were observed to form at least one hydrogen bond in the four protein states (WT, Y268C, Y268N and Y268S) except the following drug compounds; MMV007638 in Y268C mutant system, MMV019551 in two systems (Y268C and Y268S), and MMV659004 in three systems (WT, Y268C and Y268N). Here, hydrogen bonds observed

in < 10 % of MD simulation time were considered short-lived/weak, thus not significant in providing ligand stability in the active site. **Figure S5.12** showed (hydrogen bond fractions column) that for 79 % and 28 % of the simulation time, there was a hydrogen bond formed between the oxygen atom of MMV007638 and the nitrogen atom of residue Val284 in Y268N and Y268S systems, respectively. Hydrogen bond formation was also reported between the oxygen atom of MMV019551 and the nitrogen atom of residue Gln289 in the WT and Y268N systems for 89 % and 11 % of the simulation time, respectively. Similarly, hydrogen bonds were observed for 10 %, 12 % and 42 % of the simulation time between the oxygen atom of MMV676444 and the Met116 residue (WT) and the nitrogen atoms of Trp136 (Y268C) and Trp140 (Y268N) residues, respectively.

On the other hand, SANCDB compounds displayed in **Figure S5.13** portrayed the highest number of residues forming significant hydrogen bonds for the stability of the compounds in the protein's active site. The highest hydrogen bond fraction, representing a high percentage of bond occupancy formed across MD simulation time, for each of the systems was further discussed. For SANCDB, hydrogen bonds were observed for 40 %, 80 %, 60 % and 25 % of the simulation time between the oxygen atom of SANC00373 and the Met133, Tyr126, Glu261 and Trp136 residues in the WT, Y268C, Y268N and Y268S systems, respectively. In the SANC00459 drug, hydrogen bonds were observed for 60 %, 60 %, 80 % and 60 % of the simulation time, between the oxygen atom of the drug and the Met133, Ile258, Asn268 and Ser268 residues in the WT, Y268C, Y268N and Y268S systems, respectively. In the SANC00548 drug, hydrogen bonds were observed for 64 %, 48 %, 72 % and 75 % of the simulation time, between the oxygen atom of the drug and the Tyr249, Ile244, Glu261 and Met133 in the WT, Y268C, Y268N and Y268S systems, respectively. In the SANC00725 drug, hydrogen bonds were observed for 27 %, 30 %, 60 % and 50 % of the simulation time, between the oxygen atom of the drug and the Pro260, Gln289, Ile258 and His320 residues in the WT,

Y268C, Y268N and Y268S systems, respectively. Overall, hydrogen bonds were formed with active site residues Met133, Trp136, Ile258, Pro260, Glu261 and Val284 that stabilised the drug compounds in the Qo active site.

Further ligand stability was assessed by calculating the number of vdW interactions compared to hydrogen bonds formed at 25 ns time step ($n = 15$ timepoints) across the 350 ns MD simulation, as highlighted in **Table S5.6**. From these results, we reported a consistently high number of vdW interactions (average range of 6 - 17) in all ligand systems across all 15 timepoints. Also, the ligand systems accumulated at least two hydrogen bonds across MD simulation in both MMV and SANCDB compounds. However, it was noted that SANCDB compounds consistently formed above two hydrogen bonds each across the 15 timepoints in WT, Y268C, Y268N and Y268S protein systems.

Table S5.6 highlights the specific active site residues forming hydrogen bonds. Among the active site residues forming hydrogen bonds is Glu261, which has been reported to contribute to ligand stabilisation in the Qo substrate binding pocket [107,201]. Out of the eight ligand systems, the findings showed that two ligand systems, SANC00373 and SANC00548, in both WT and mutant states, formed a hydrogen bond with residue Glu261. This further supports ligand RMSD results where these potential hits were observed to be stable across the entire 350 ns MD simulation.

5.3.15 Dynamic Residue Network (DRN); Global Residue Network analysis

The network analysis for this study was performed adapting two approaches namely, global, and local residue network analysis. Global residue network analysis focus on calculating residue network changes in the entire protein while local residue network focuses on a subset of residues on interest. In this case, the whole protein as assessed globally while a subset

composed of active site residues was assessed. The findings here will be correlated to motif analysis in chapter 3.

5.3.15.1 Five centrality measures identified key residues ‘hubs’ important in protein function

To determine the PfCytb-ISP protein residue interaction network and explore the residues central in protein communication and function, a total of five centrality measures were calculated from the dynamic residue network (DRN) as detailed in the methods section. Briefly, the input data captured the last 100 ns (250 - 350 ns times scale), representing the equilibrated phase of the MD simulation, where a 6.7 Å cut-off distance between residues in a network and a step size of 100 frames was applied. The five centrality metrics include; *betweenness centrality (BC)*, *closeness centrality (CC)*, *eigen centrality (EC)*, *degree centrality (DC)* and *katz centrality (KC)* [308]. The DRN calculations were performed on a total of 36 (four holo and eight ligand systems across WT, Y268C, Y268N and Y268S) protein systems. High centrality residues (having significantly high/positive centrality values) for each of the systems were obtained. Thus, top 4 % cut-off was used to calculate the highly significant residues in each of the ligand systems ($4 * 8 = 36$). The residues with high values for the *BC*, *CC*, *DC*, *EC*, and *KC* metrics are shown in **Figures 5.20 – 4.24**, respectively. These high centrality values correspond to residues with possible implications on the protein function. From these significant residues (comparing holo and ligand systems in WT and mutant), hubs with > 90 % occurrence in each of these systems was identified. In this study, ‘all systems’ is inclusive of both holo and ligand systems.

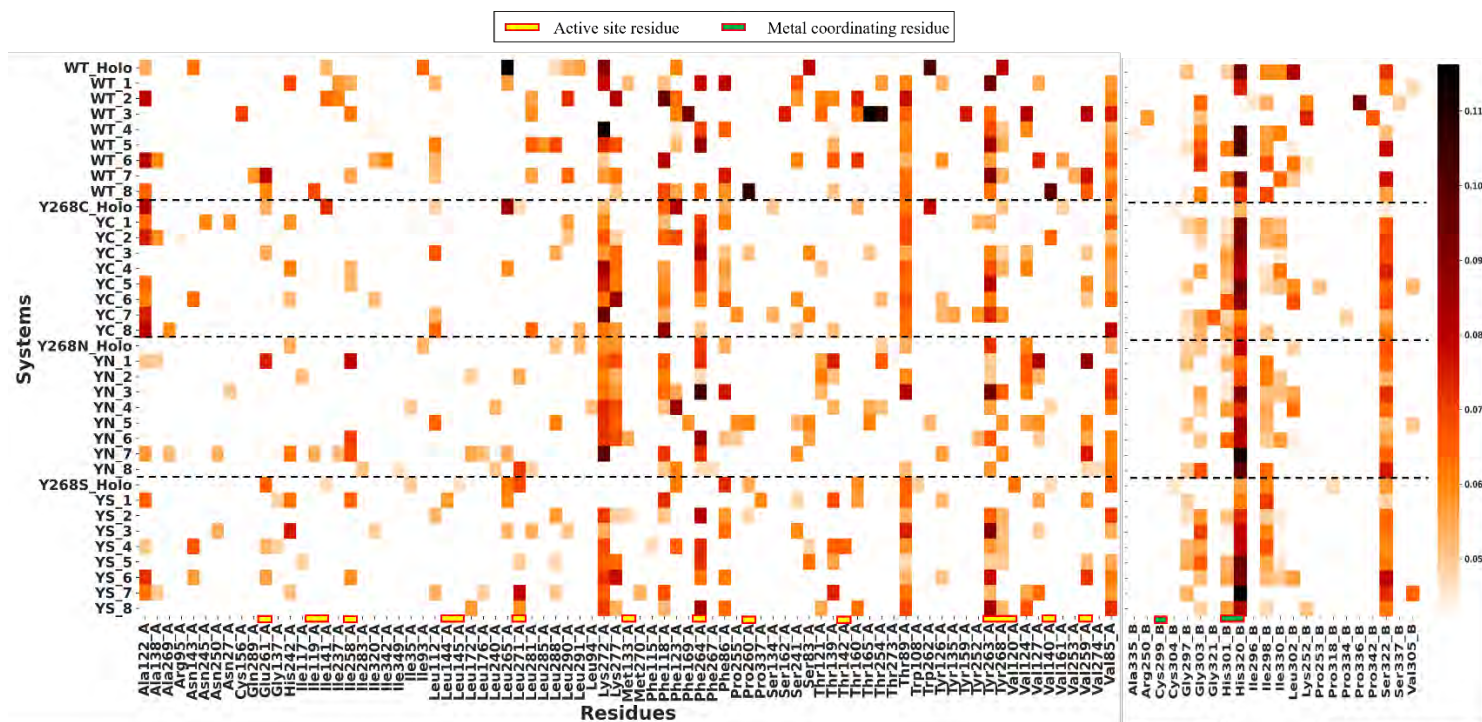


Figure 5.20: A heat map representation of top 4 % *Averaged BC* representing significant residues affecting protein communication across eight ligand systems in their WT and mutant states. Active site residues and metal interacting residues are displayed in purple and green squares, respectively.

The function of the protein-ligand systems was assessed with respect to the compounds binding at the Qo active site. Generally, it was noted that there was a significant loss of functionality in the active site residues in not only WT but also all the three mutant states across all the eight ligand systems. Looking at each of the centrality metrics, each ligand system reported significant residues that function in substrate binding. We also noted that the drug compounds behaved uniquely from one state to another (WT, Y268C, Y268N and Y268S).

In *BC* metric (**Figure 5.20**), the residues frequently observed across the systems included; Tyr263, Phe264, Tyr268, His301 and His320, out of which Phe264 and Tyr268 are active site residues important in substrate binding and stability [107,203]. The metal coordinating residues (His301 and His320) of the [2FE-2S] cluster play a major role in stabilising cofactors which enhance the electron transfer (ET) process [336,337]. Besides active site and metal

coordinating residues, residues Val85, Thr89 and Lys272 from chain A and Ser322 from chain B were found to be *hubs* across > 90% of all systems. Out of these residues, Lys272 and Ser322 were found to be near the Qo active site. These residues are conserved including Motif 1 (85, 89), Motif 2 (263, 264, 268). The latter formed the PEWY motif essential in substrate binding hence protein communication is increased among the ligand systems.

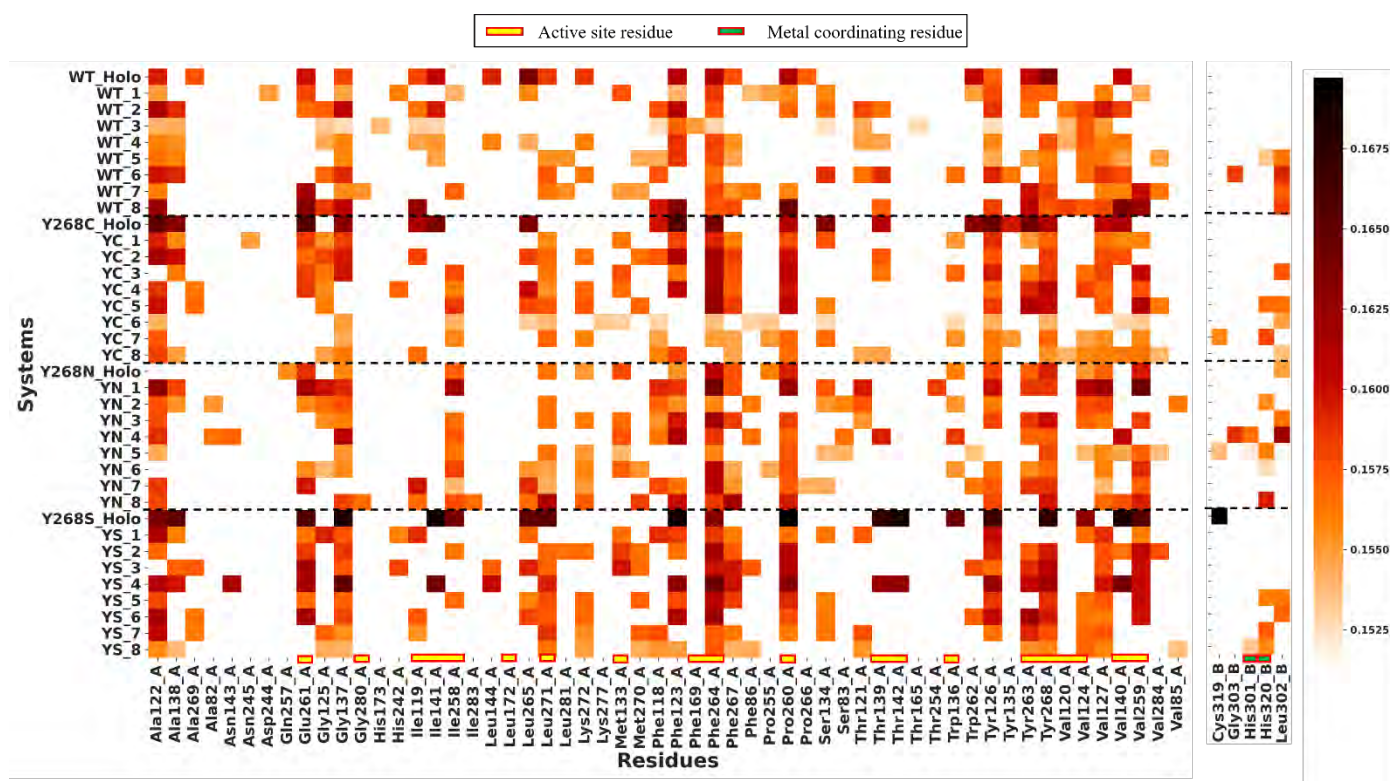


Figure 5.21: A heat map representation of top 4 % *Averaged CC* representing significant residues affecting protein communication across eight ligand systems in their WT and mutant states. Active site residues and metal interacting residues are displayed in purple and green boxes, respectively.

In *CC* metric (**Figure 5.21**), Val124, Gly137, Val140, Pro260, Glu261, Tyr263, Phe264, Tyr268, Leu271 and His320 were frequently observed across the systems. Using this metric, *hubs* including Ala122, Tyr126, Gly137, Pro260, Phe264, Tyr263, Tyr268 and Leu271 were reported from chain A only. Notably, the residues formed a part of conserved motifs 2 (260,264, 268, 271) and motif 3 (122, 126, 137) in support of the overall function of the protein.

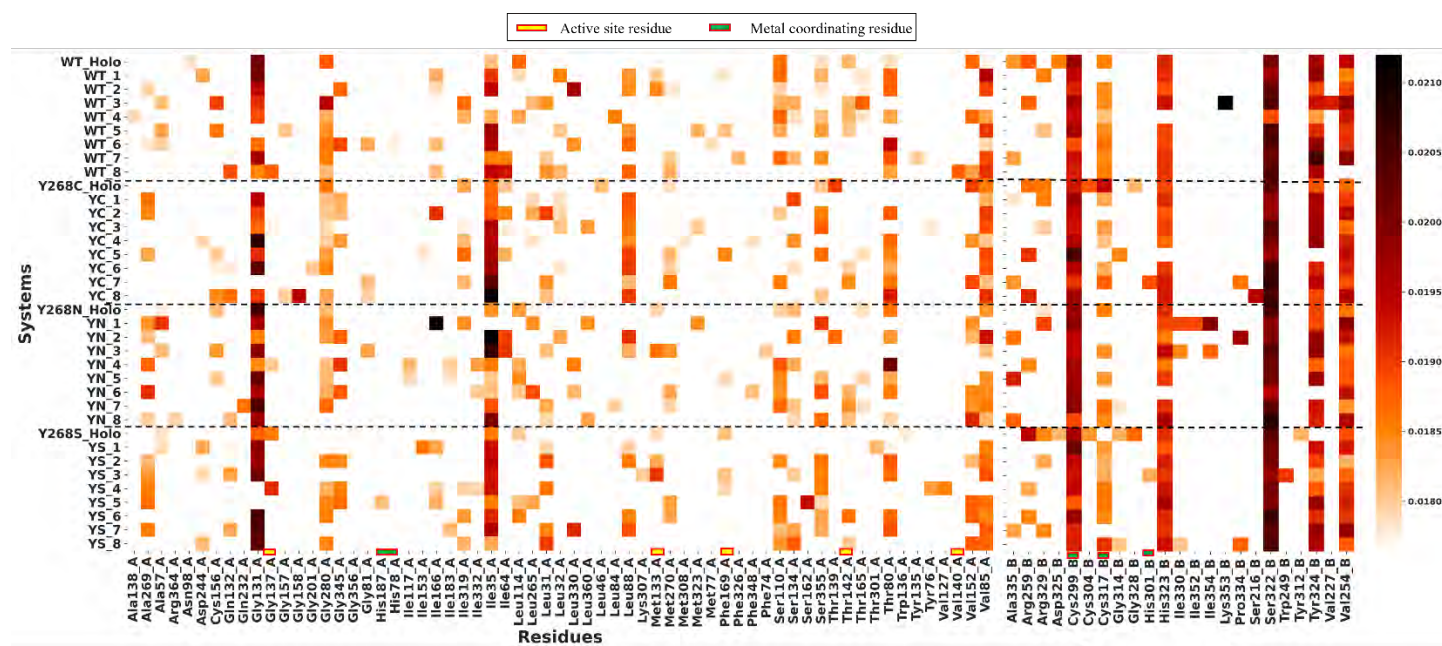


Figure 5.22: A heat map representation of top 4 % *Averaged DC* representing significant residues affecting protein communication across eight ligand systems in their WT and mutant states. Active site residues and metal interacting residues are displayed in purple and green boxes, respectively.

In the *DC* metric (**Figure 5.22**), residues Thr142, Cys299 and Cys317 were reported to have high *DC* values. On the other hand, *hubs* included Ile35, Thr80, Val85 and Gly131 from chain A, and Cys299, Ser322, His323 and Tyr324 from chain B. These residues were correlated to motif analysis from chapter 3 belonging to conserved motif 1 (80, 85), motif 3 (122, 131, 142) and motif 5 (35).

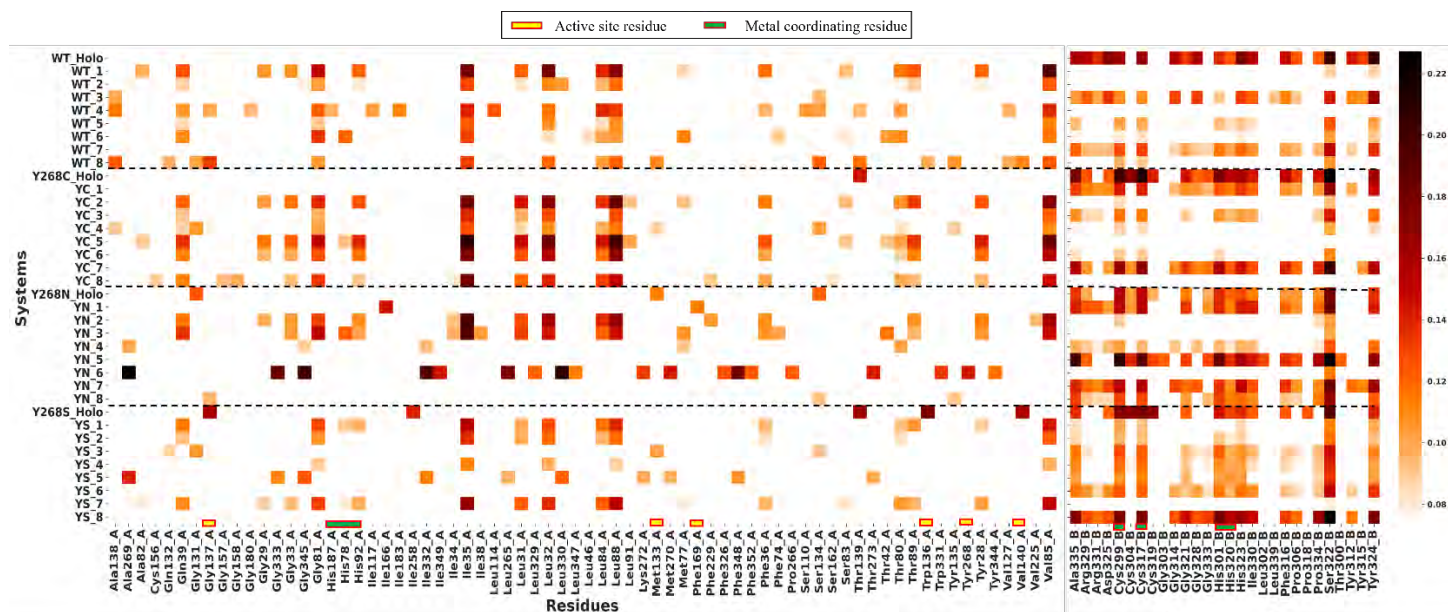


Figure 5.23: A heat map representation of top 4 % *Averaged EC* representing significant residues affecting protein communication across eight ligand systems in their WT and mutant states. Active site residues and metal interacting residues are displayed in purple and green boxes, respectively.

In the *EC* metric (**Figure 5.23**) and *KC* metric (**Figure 5.24**), a similar profile was observed where metal interacting residues Cys299, His301, Cys317 and His320 were reported as *hubs* and were only reported in chain B. These residues coordinate the [2FE-2S] cluster that participates in the ET process in the protein.



Figure 5.24: A heat map representation of top 4 % *Averaged KC* representing significant residues affecting protein communication across eight ligand systems in their WT and mutant states. Active site residues and metal interacting residues are displayed in purple and green boxes, respectively.

Overall, the observed active site residues were central in the Qo substrate binding site, while metal interacting residues were important in coordinating the metal centre, altogether promoting the ET process. *BC* metric reported a high number of significant residues that were reported to form a part of conserved motifs important in substrate binding. Thus, the *BC* metric was used to elucidate the effect of mutation and hit compounds binding on the active site.

Furthermore, comparative residue analysis across the eight ligand systems for the WT and mutant systems was performed by mapping these significant residues onto the protein structure for further analysis as shown in **Figures S5.14 – S5.18**. From the mapping, **Figure 5.25** shows the communication pattern as outlined by the 5 DRN metrics using the WT protein for discussion. The ET pathway in PfCytb-ISP outlined previously shows that electrons are transported from the Qo active site to [2FE-2S] cluster area as well as from Qo to Qi active site, altogether pumping a total of 4 electrons from the matrix to the inner mitochondrial

membrane. As such, we noted an interesting communication pattern observed by each of the metrics. *BC* metric showed communication hubs around Qo and Qi active site, while *CC* explicitly formed communication hubs just within the Qo active site. *DC* metric, on the other hand, showed communication hubs spread across Qo to Qi then to [2FE-2S] cluster region in ISP chain, while *EC* metric focused solely on ISP chain where communication hubs surrounded electron transfer process going on in the cluster region only. Lastly, *KC* highlighted communication hubs in the Qi active site of PfCytb and around the [2FE-2S] cluster region in ISP chain. In the presence of mutations (**Figures S5.14 – S5.18**), various ligand systems observed loss of communication patterns due to the mutations. For instance, the *CC* metric identified communication pattern changes in the MMV019551 ligand system (**Figure 5.21**) in the presence of Y268C, suggesting that the mutation has a probability of affecting protein function.

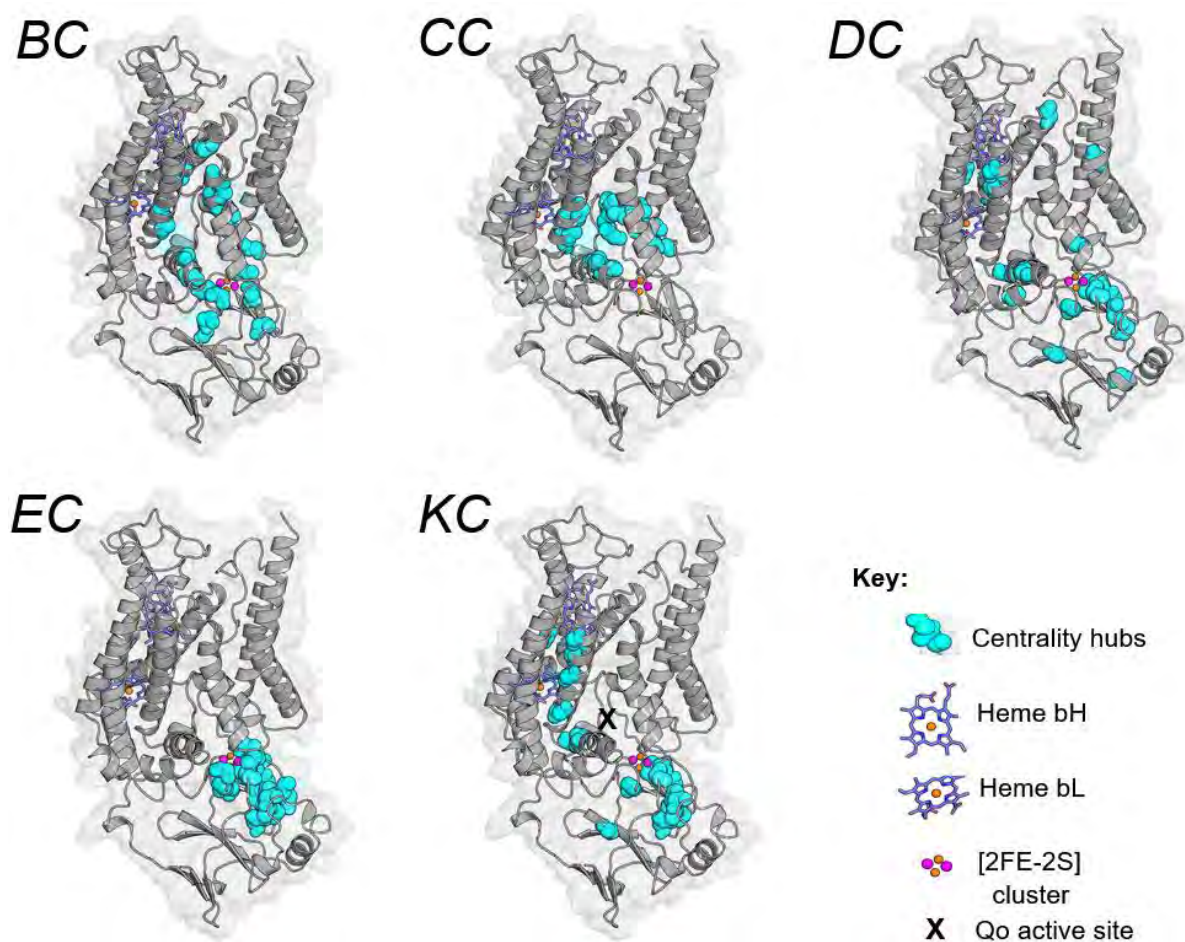


Figure 5.25: Structural mapping of the top 4 % residues as depicted by five DRN centrality metrics (*Averaged BC, CC, DC, EC, and KC*). This is a representation of the communication hubs in the WT model to assess the communication patterns. A key is provided highlighting the colour and representations of different features in the protein.

Cumulatively, common residues with high values across eight ligand systems using all the five centrality measures were calculated (**Table 5.4**).

Table 5.4: Significant residues common across eight ligand systems in each protein system. The assessment was performed using five centrality metrics; $\pm 2SD$ was used to obtain high values that correspond to functional residues. Residues forming the active site residues important in substrate binding are indicated in green colour, while those located in the metal coordination sphere are shown in blue colour.

		DRN Centrality metrics			
		WT	Y268C	Y268N	Y268S
Averaged Betweenness Centrality (BC)	+	<i>PfCytb chain:</i> VAL_85, THR_89	<i>PfCytb chain:</i> VAL_85, THR_89, PHE_264 , LYS_272, HIS_301 , HIS_320 , SER_322	<i>PfCytb chain:</i> VAL_85, LYS_272, HIS_320 , SER_322	<i>PfCytb chain:</i> THR_89, GLY_297, ILE_298, HIS_320 , SER_322
Averaged Closeness Centrality (CC)	+	None	None	None	None
Averaged Degree Centrality (DC)	+	None	None	<i>ISP chain:</i> CYS_299 , CYS_317 , SER_322	<i>ISP chain:</i> SER_322
Averaged Eigen Centrality (EC)	+	<i>ISP chain:</i> CYS_299 , SER_322	<i>PfCytb chain:</i> ILE_35 <i>ISP chain:</i> CYS_299 , SER_322	<i>ISP chain:</i> SER_322	<i>ISP chain:</i> SER_322, HIS_323
Averaged Katz Centrality (KC)	+	<i>ISP chain:</i> SER_322	<i>PfCytb chain:</i> LEU_32, ILE_35, VAL_85, LEU_88 <i>ISP chain:</i> CYS_299 , CYS_317 , SER_322, TYR_324	<i>ISP chain:</i> CYS_299 , CYS_317 , SER_322	<i>PfCytb chain:</i> ILE_35, VAL_85 <i>ISP chain:</i> CYS_299 , SER_322, TYR_324

Specifically, the reported residue Tyr263 is in the mitochondrial PEWY (Pro260-Glu261-Trp262-Tyr263) conserved motif, also known as the Qo motif [34]. The PEWY motif, found on a loop region of this protein, has been reported to participate in substrate binding. From the results, residue Tyr263 was noted to have a high *BC* value in WT, Y268N and Y268S systems. Altogether, metal coordinating residues Cys299, His301 and His320 were observed in Y268C, Y268N and Y268S mutant states of all ligand systems across the five centrality metrics. These residues coordinate the [2FE-2S] cluster of the ISP chain, which then participates in the electron transfer process in the Qo active site [4,34]. These findings support previous findings on the importance of the conserved motif in protein function [34]. This highlights that protein function is likely to be maintained in the presence of mutations when the eight potential hits

bind to the Qo active site. These findings further suggest that the identified compounds are likely to withstand ATQ resistance caused by the point mutations.

5.3.16 Effect of ligand binding and mutations on intra-protein communication in the eight protein-ligand systems

To determine the effect of the potential hit compounds binding on the PfCytb-ISP active site as well as the effect of mutations on residue communication, the protein systems were compared using the DRN *BC* metric. As reported previously, the input data was sampled over an equilibrated phase of the MD trajectory [49]. For insights on how protein communication was affected in these ligand systems and in the WT and mutant states, changes in *BC* were calculated (ΔBC : Ligand-free minus Ligand-bound systems; WT Ligand-bound minus Mutant Ligand-bound). A threshold value of ± 2 SD of the *BC* values was used for all ligand systems as the values were centred around the mean.

As a metric to identify crucial residues in communication within the protein, it was noted that once each ligand binds, key changes in the active site residues were observed (**Table S5.7**), as well as in the presence of mutations (**Table S5.8**). In **Tables S5.7 and S5.8**, significant residues with greater or less than 2 SD ΔBC values were reported. Herein, positive ΔBC values represent decreased residue connectivity, whereas negative ΔBC signifies increased residue participation in communication in the protein systems. Altogether, the significant residues, denoting the effect of ligand binding, were mapped onto PfCytb-ISP protein structure where most of them were observed to be in the active site and a few of them outside the active site, as shown in **Figure 5.26**. Those that were found outside the active site were found not only in the WT but also in all mutant systems. This suggests the likelihood of the potential hit compounds altering distant residue communication patterns.

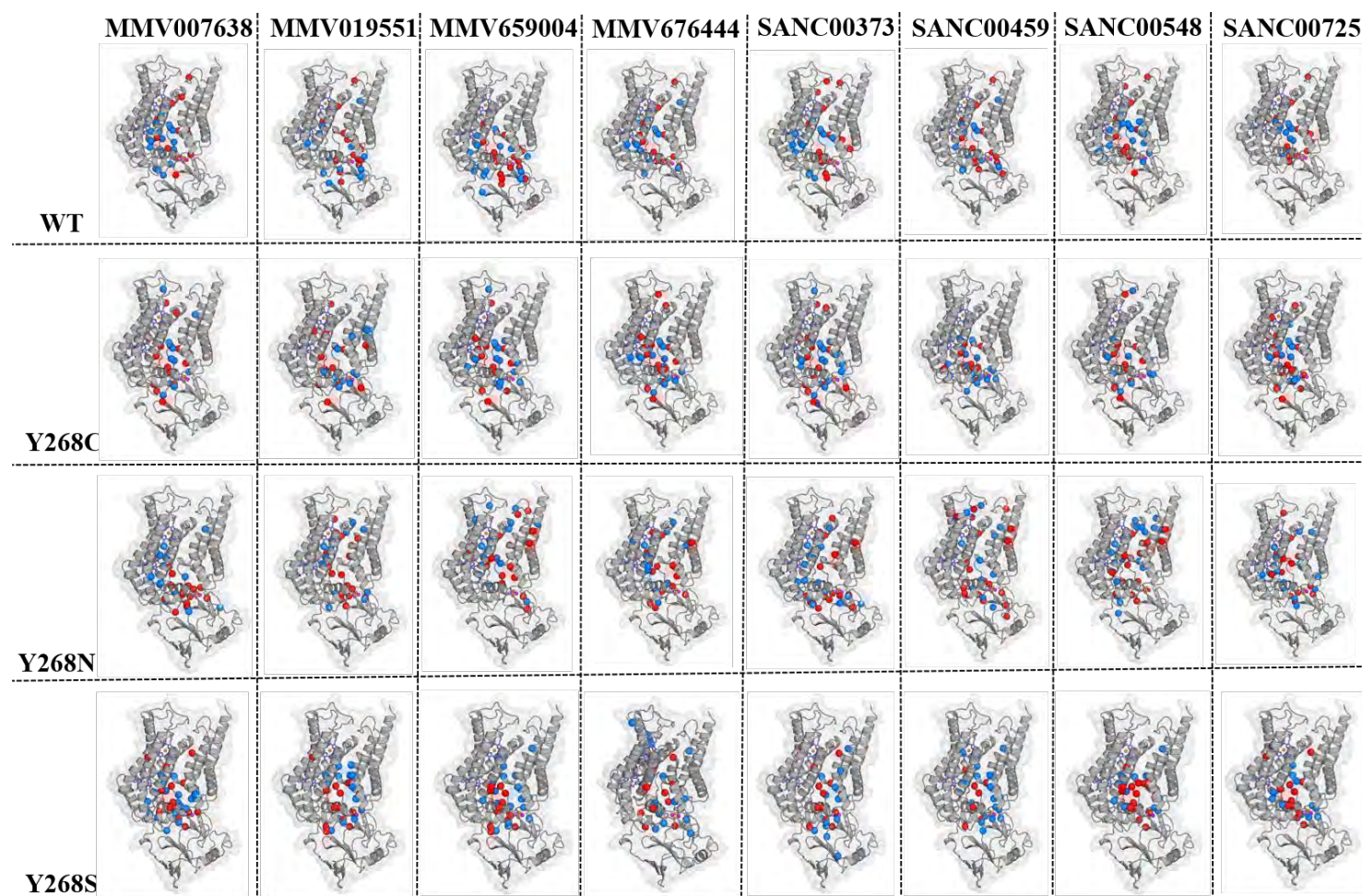


Figure 5.26: Structural mapping of significant residues onto PfCytb-ISP protein in WT and mutant models (Y268C, Y268N and Y268S). Protein communication was assessed in a total of eight ligand-bound protein systems. Positive and negative ΔBC values are shown as red and blue spheres, respectively.

A heat map representation of the ΔBC values showing the effect of ligand binding in all systems was presented in **Figure 5.27A**. Overall, the figure showed the regions having both increased and decreased protein communication due to ligand binding. The active site residues reported in each ligand system were presented in **Table S5.7** in bold. The residues frequently observed to have decreased protein communication across all ligand systems include: Phe123, Val140, Ile141, Val259, Pro260, Glu261, Trp262 and Phe264. Besides the fact that most of these residues participate in the formation of hydrophobic interactions that stabilise the substrate in the active site [21], other residues were noted to participate in protein function. These include

residues Glu261 and Trp26, which form a part of the conserved PEWY (Pro260-Glu261-Trp262-Tyr263) motif, essential in substrate binding [50,51]. Residue Trp262 decreased communication in all ligand systems except the SANC00548 system, while residue Glu261, important in forming a hydrogen bond with the substrate leading to its oxidation [35,52], showed decreased residue connectivity in all ligand systems.

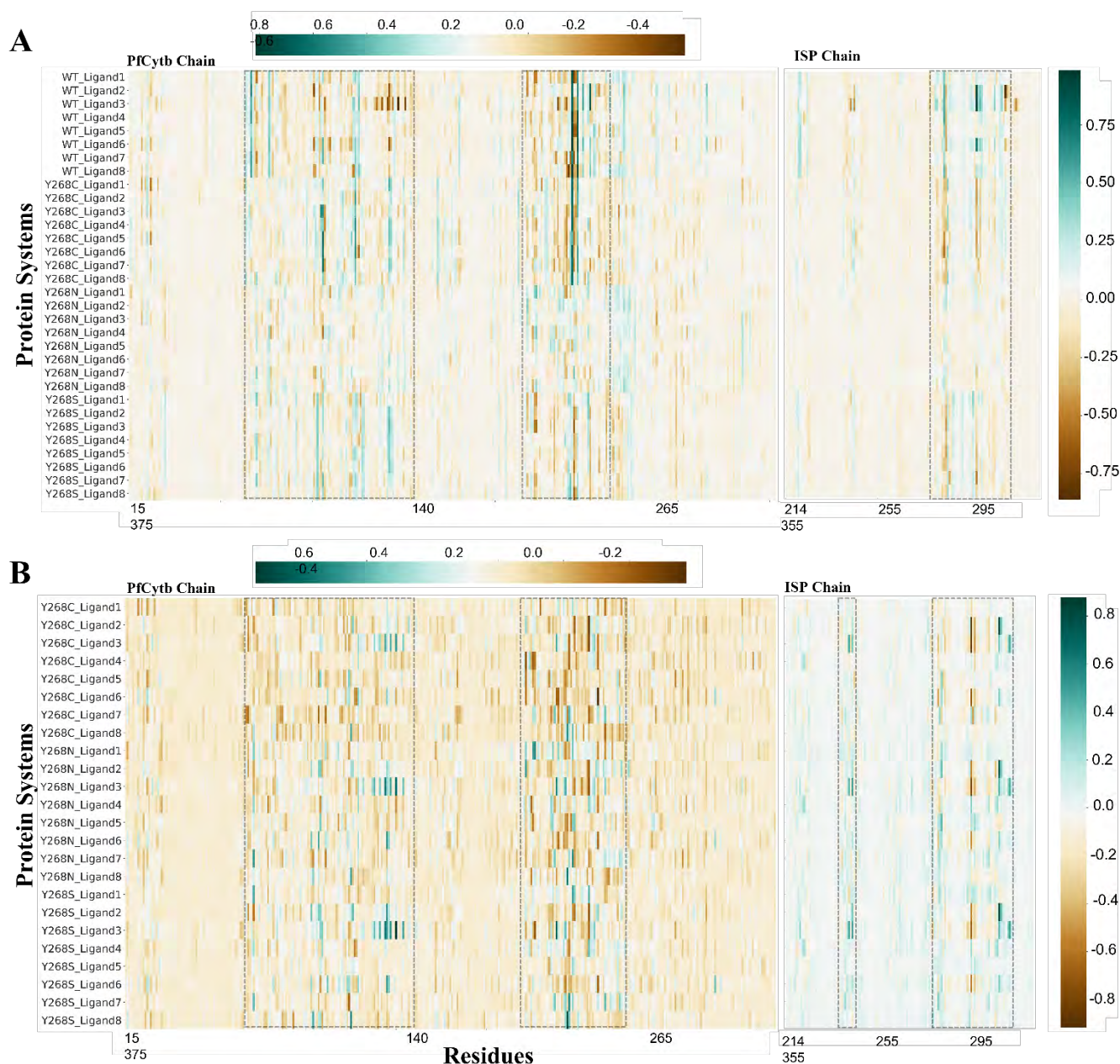


Figure 5.27: A heat map representation of significant residues affecting protein communication; change in BC . (A) Effect of ligand binding. (B) Effect of mutation in the presence of ligand. Regions where protein function is affected are highlighted in dotted lines.

Further analysis revealed Val259, Pro260, Glu261, Trp262, Tyr263, Phe264, His301 and His320 as residues with increased connectivity across all the ligand systems. Residues Pro260, Trp262 and Tyr263 also forms a part of the PEWY motif 2, which is essential in substrate binding, while His301 and His320 are metal coordinating residues essential in the protein's electron transfer process [4,53]. Notably, residue Tyr268, the point mutation residue position,

was reported in the SANC00548 ligand system of Y268N mutation. These residues showed negative ΔBC values, indicating that they remained central to communication in the WT and the mutant systems. However, some of these residues in some of the ligand systems contributed to decreased protein communication. This pattern is similar to that observed in the WT and mutant states of the ligand systems. Overall, as indicated by the dotted regions of **Figure 5.27A**, there is a balance in residues showing a decrease and an increase in protein communication, suggesting that the protein function is likely to be affected with the possibility of the function being maintained.

To determine the effect of the mutation in the presence of the eight ligand systems, **Figure 5.27B** showed ΔBC values for all systems where the residues that reported decreased or increased residue connectivity were highlighted (also compiled in **Table S5.8**). Across all the three mutant systems, a pattern was observed showing common regions across all ligand systems. The pattern highlighted regions/residues having decreased or increased protein communication in each ligand system. This is expected as all the eight ligands adopted a unique orientation in the active site pocket, as evidenced by the unique docked poses whose behaviour in the active site pocket is bound to be varied. Frequently observed active site residues that decreased in protein communication in each of the mutant systems across all ligand systems include: Y268C (Phe123, Leu144), Y268N (Phe123, Leu144, Phe169, Phe264) and Y268S (Val140, Glu261, Phe264). Those that gained importance in protein communication include Y268C (Leu271, Val259, His320), Y268N (Met133, Leu271, His320) and Y268S (Val259, Pro260, Glu261, His320). Some of these residues formed a part of the following conserved motifs; motif 2 (259, 260, 261, 264, 271) and motif 3 (123, 133, 140, 144, 169). Protein communication in all the ligand systems were likely maintained. All three mutations reflected that the active site pocket was affected in its residue communication, reporting different active site residues in each mutation. This likely suggests that all the mutations have different

resistance mechanisms despite being in the same residue position. However, the SANC0725 ligand system reported residues Ile119, Val140, Val259, Pro260 and Glu261 as decreasing protein communication in all three mutant systems. Specific to this system, we see that the hydrogen bond facilitated by residue Glu261 will likely be lost.

Additionally, we noted residue His320 which was reported to have gained communication in all three mutant systems. The residue is important in the metal coordination of the [2FE-2S] cluster which promotes the electron transfer process. Thus, this possibly suggests that when the eight hit compounds bind, the protein function is likely to be negatively affected in the presence of mutations. These findings agree with previous studies that report the cause of ATQ resistance to be the point mutations that alter the substrate binding site.

The effect of the three mutations in the absence of a ligand has previously been reported [37], where active site residue Trp262 was reported to decrease in protein communication across all mutant systems. In the eight ligand systems, other active site residues have been reported to decrease in communication, different from those reported in the absence of the ligand. This observation at the active site can be accounted for by the binding of the eight ligands.

5.3.17 DRN; Local Residue Network Analysis

5.3.17.1 Residue-residue interactions within the active site residues

A total of 28 active site residue-residue interactions were analysed in the eight ligand systems. Residue interactions in each WT ligand system were compared to their respective interactions in the mutant systems, where those that reduced, lost or gained/enhanced interactions were shown in **Table 5.5**. The table reports compromised residue-residue interactions in at least two mutant systems. This was performed to investigate whether the mutations affected the residue

interactions within the active site once the eight potential hit compounds were bound. The residue contacts were determined by applying a cut-off distance of 6.7 Å to all the 28 residues positions and an iteration step size of 100 frames. The active site residues included: Met116, Ile119, Val120, Phe123, Val124, Met133, Trp136, Gly137, Val140, Ile141, Thr142, Leu144, Leu145, Ile155, Phe169, Leu172, Ile258, Val259, Pro260, Glu261, Trp262, Tyr263, Phe264, Phe267, Tyr268, Leu271, Val 284, Leu285 [54]. All these amino acid residues are hydrophobic except Glu and Thr, which are charged polar and uncharged polar residues, respectively. In assessing reduction in residue interactions, a 9 to 10-fold reduction was considered in comparing WT and mutant systems to obtain top significant residue interactions that were notable affected.

Residue-residue interactions common across the three mutations were analysed and discussed. Overall, out of all the active site residues, a total of 20 and 15 unique active site residues were observed, across the eight ligand systems, to be reduced/lost and gained/enhanced, respectively. Due to the chemical structure of each hit compound and the unique interactions formed by each compound, the systems were discussed individually. This was to investigate the effect of these compounds' binding in the Qo active site as well as ascertain the effect of the mutations on the residue interactions in all respective ligand systems.

Since the Qo substrate binding site is hydrophobic in nature (based on the physico-chemical properties of the active site residues and the entire protein) most of the interactions were hydrophobic [187,338], while other interactions such as hydrogen bonds were reported. In MMV007638, hydrophobic interactions Ile258-Pro480, Tyr263-Ala122, Met133-His242 and Ile155-Thr165 were lost, while Thr142-Ser146 and Thr142-Leu145 contacts were gained in at least two mutant systems. In MMV019551, hydrophobic interactions Leu271-Val284, Glu261-Tyr126 and Tyr263-Ala122 were reduced/lost, while Leu145-Ile155 hydrophobic contact was

gained in at least two mutant systems. In MMV659004, residue-residue interactions Pro260-Leu265, Leu271-Val284 and Pro260-Tyr268 were reduced/lost, while Tyr268-Val259 contact was gained in at least two mutant systems. In MMV676444, hydrophobic interactions Pro260-Leu265 and Phe264-Phe123 were reduced in at least two mutant systems. Gained Interactions that are unique to each mutant system were as shown in **Table 5.5**.

In SANC00373, unique residue-residue interactions respective to each mutant system were reported to be reduced/lost, while Val259-Leu265, Tyr268-Val259 and Leu145-Ile148 hydrophobic contacts were gained in at least two mutant systems. In SANC00459, residue-residue interactions Trp136-Gln132, Leu144-Leu281 and Tyr268-Cys334 were reduced/lost in at least two mutant systems, while contacts Met133-His242 and Val259-Leu265 were gained in the Y268C mutant system only. In SANC00548, residue-residue interactions Ile141-Ile155, Leu144-Leu281, Glu261-Leu265, Thr142-Ile155 and Ile155-Thr165 were reduced/lost, while Thr142-Leu145, Thr142-Ser146 and Leu145-Ile155 contacts were gained in at least two mutant systems. In SANC00725, unique residue-residue interactions were reduced/lost in respective mutant systems as outlined in **Table 5.5** while Ile258-Trp136 and Tyr263-Ser83 contacts were gained in the Y268N mutant system only.

Overall, all the eight ligand systems reported compromised interactions (reduced/lost) as well as gained interactions. In all the ligand systems, the residues formed a part of either conserved motif 2 and 3 (refer to chapter 3), where both decreased and increased residue contact frequencies was reported. Worthy to note is that all ligand systems except MMV659004 and MMV676444 reported reduction or loss in Trp136-Gln132 contact. This contact is similarly reported when the ATQ drug binds in the Qo active site, as reported previously [37]. In comparison with the ATQ drug which has faced resistance, the eight ligand systems seem to display unique interactions that were also compromised. However, there was a balance noted

between reduced/lost and gained/enhanced interactions, suggesting that the protein function at Qo active site is likely to be maintained.

Table 5.5: Active site residue-residue interactions observed to be either reduced, lost or gained in the eight ligand systems across mutant (Y268C, Y268N and Y268S) systems as compared to WT. Residue interactions reduced/lost/gained across all the three mutation ligand systems (coloured in blue), those found in two mutant ligand systems (coloured in green), and interactions reduced/lost in six out of eight ligand systems (coloured in red).

Ligand systems	REDUCED/LOST			GAINED/ENHANCED		
	Y268C	Y268N	Y268S	Y268C	Y268N	Y268S
MMV007638	Ile258-Pro480, Glu261-His242, Tyr263-Ala122, Met133-His242, Met133-Tyr126, Ile145-Ile148, Thr142-Val152	Ile258-Pro480, Ile258-Cys481, Thr142-Ile155, Trp136-Pro255, Trp136-Gln132, Glu261-Tyr126, Met133-His242, Met133-Pro255, Met133-Asn245, Val259-Gln257, Val259-Cys334, Tyr268-Pro260, Ile155-Thr165	Glu261-Tyr126, Ile258-Pro480, Tyr263-Ala122, Ile155-Thr165	Thr142-Ser146, Thr142-Leu145, Trp136-Thr254	Thr142-Leu145, Thr142-Ser146	-
MMV019551	Ile258-Thr254, Ile141-Leu145, Ile141-Phe123, Thr142-Ser146, Thr142-Val161, Glu261-Tyr126, Glu261-His242, Met133-His242, Leu145-Cys156	Leu271-Val284, Met133-His242, Trp262-Leu240, Trp136-Pro255, Trp136-Gln132, Gly137-Pro260, Glu261-Tyr126, Tyr263-Ala122, Met133-Pro255, Leu271-Val284, Met133-His242	Leu144-Ser147, Glu261-Tyr126, Glu261-His242, Leu271-Val284, Pro260-Leu265, Pro260-Val140, Tyr263-Ala122, Tyr268-Pro260	Trp136-Thr254, Leu145-Ile148	Val140-Gly465, Tyr263-Ser83, Leu145-Ile155	Tyr268-Val259, Leu145-Ile155
MMV659004	Thr142-Val161, Ile155-Thr165, Leu144-Leu281, Leu144-Ser147, Pro260-Leu265, Tyr268-Pro260, Ile155-Arg168	Leu271-Val284, Trp262-Trp331, Pro260-Leu265, Tyr268-Pro260, Met116-Leu288	Leu144-Ser147, Leu271-Val284, Thr142-Val161, Trp136-Thr254, Pro260-Leu265, Glu261-Tyr263, Leu271-Lys277, Tyr268-Pro260	Val259-Leu265, Tyr268-Val259	Leu145-Ile148	Tyr268-Val259
MMV676444	-	Thr142-Val161, Trp136-Pro255, Pro260-Leu265, Phe264-Phe123, Met133-Pro255, Met116-Leu288	Val259-Cys334, Val124-Ile141, Ile258-Pro480, Thr142-Val161, Pro260-Leu265, Phe264-Phe123, Tyr268-Pro260	Met133-His242	Ile258-Trp136, Tyr263-Ser83	-
SANC00373	Thr142-Val161, Glu261-Tyr263, Ile141-Phe123, Thr142-Ile155, Glu261-Phe264, Phe264-Ile141, Tyr268-Pro260	Ile258-Pro480, Thr142-Val161, Trp136-Gln132, Val259-Cys334, Ile155-Thr142	Val259-Cys334	Val259-Leu265, Tyr268-Val259, Leu145-Ile148	Val140-Gly465, Leu145-Ile148	Leu285-Gln289, Val259-Leu265, Tyr268-Val259, Leu145-Ile155
SANC00459	Thr142-Val152, Leu145-Ala151, Ile258-Trp136, Leu144-Leu281, Ile141-Cys156, Tyr268-Cys334, Leu145-Cys156, Leu145-Val152	Trp136-Gln132, Pro260-Leu265, Leu271-Leu464, Leu271-Lys277, Val259-Gln257, Val259-Gln335, Tyr268-Cys334, Tyr268-Pro260	Trp136-Gln132, Glu261-His242, Leu144-Lys277, Leu144-Leu281, Tyr268-Cys334	Met133-His242, Val259-Leu265	-	-
SANC00548	Glu261-Tyr263, Ile155-Thr165, Trp262-Leu240, Ile119-Tyr263,	Trp136-Gln132, Met133-His242, Leu144-Leu464, Ile141-Ile155, Thr142-Ile155, Ile155-Thr165	Ile258-Cys466, Leu144-Leu281, Leu285-Leu281, Ile141-Ile155,	Thr142-Leu145, Thr142-Ser146	Thr142-Leu145, Thr142-Ser146	Leu285-Gln289, Leu145-Ile155

	Leu144-Leu281, Ile141-Ile155 , Thr142-Ile155 , Glu261-Leu265 , Glu261-Phe264, Met133-His242		Glu261-Leu265 , Glu261-Val259, Glu261-Tyr263, Ile155-Thr165		Ile155-Leu145	
SANC00725	Leu145-Ala151, Ile119-Tyr268, Ile141-Leu145, Thr142-Tyr159, Thr142-Ser146, Val140-Gly465, Tyr263-Ala122 , Leu145-Val152, Leu145-Cys156, Ile155-Arg168	Ile258-Pro480, Leu285-Leu281, Gly137-Pro260, Tyr263-Ala122 , Met133-Gln257, Phe267-Ser287	Trp136-Gln132 , Pro260-Val140	-	Ile258-Trp136 Tyr263-Ser83	-

Additionally, residue-residue interaction changes around the identified *hubs* (refer to Section **5.3.15**) were analysed in all ligand systems (**Table S5.8**). This was to assess the effect of mutation on residue interactions around the identified hubs. A comparison between WT and each mutant ligand system was performed where reduced/lost and gained/enhanced interactions were reported. Overall, reduced/lost interactions were reported in all ligand systems except MMV007638 and MMV019551. *hubs* Phe264 and Tyr263 were observed to form hydrogen bonds with Glu261 which has been implicated to stabilise Qo substrate binding [201,318,339,340]. These interactions were reported in the systems as follows: Tyr263-Glu261 (MMV659004 and SANC00548) in Y268S mutant only, and Phe264-Glu261 (SANC00373 and SANC00548) in Y268C mutant only. Also, hydrophobic Thr80-Glu236 (SABC00459 and SANC00548 in Y268N and Y268S systems) and Thr80-Ile237 (SANC00459 and SANC00725 in Y268S mutant only) were reported. This shows that the hubs were not affected in all systems, whereas in those that were affected, notable changes were observed in Y268N and Y268C mutant systems. Gained/enhanced interactions were also noted around the hubs including hydrogen bonds Tyr263-Ser83 in Y268N, His320-Pro318 in Y268N and His301-Lys115 (Y268C and Y268N) across all the ligand systems except

in MMV659004 and SANC000459. These interactions formed by His301 and His320 coordinate with [2FE-2S] cluster; thus, the interactions are likely to stabilise the metal coordination sphere and support protein function in the ET process.

PART III

TARGET 2: *PLASMODIUM* INTERFACE

5.3.18 PfCytb-ISP interface pocket prediction analysis

The ET pathway is mainly carried out via the protein's catalytic domain (**Chapter 3: Figure 3.2**). The electrons are transported from Qi to Qo active site found in Cytb through to the [2FE-2S] clusters found in the ISP subunit of the heterodimer protein. Structurally, PfCytb-ISP heterodimer supports the ET process, and at the protein-protein interface there are important residue interactions that support the protein's overall function [341,342]. Therefore, exploring the interface pocket (**Figure 5.28**) can shed light on whether the hit compounds alter protein structural dynamics and function mainly focusing on the Qo active site.

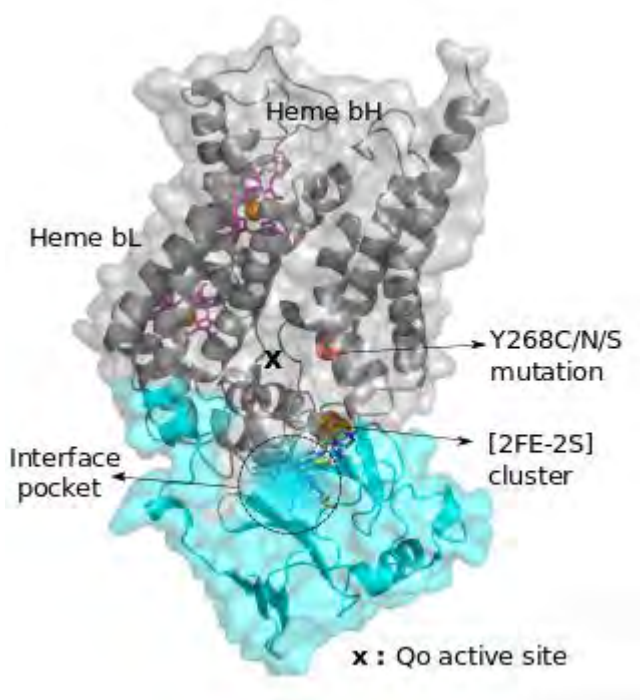


Figure 5.28: Structural diagram of PfCytb-ISP protein showing the potential interface pocket (circled), other structural features, as well as the location of the point mutations (Y268C, Y268N and Y268S), which are near the Qo active site.

Previous studies have reported changes occurring in the active site due to ligand binding in other binding sites besides the active site [80,343–345]. The PfCytb-ISP interface pocket has not been

explored before; thus, this study predicts and targets the pocket to assess the effect of selective compounds on the Qo active site. Additionally, the interface pocket analysis will also determine the effect of the point mutations Y268C/N/S on the ligand-bound systems.

Besides ligand-induced effects on the protein, factors such as mutation, pH and temperature can also be linked to allostery. As such, this concept has been utilised in drug design efforts due to its advantage point. Allosteric sites can be less conserved compared to the active site; thus, they are more selective against the parasite protein and less toxic to the host. Targeting allosteric sites allows the identification of selective hit compounds as supported by previous studies [325,343,344,346,347]. There are several approaches in identifying allosteric sites in either protein models or NMR/x-ray derived protein structures. These approaches include: normal mode analysis (NMA) [299,348], molecular dynamics (MD) [349] and machine learning [298]. Pocket identification was performed using Allosite, FTMap, FTSite and PARS (detailed in the methodology section), tools out of which FTMap was the only one that identified the PfCytb-ISP interface pocket, as shown in **Figure 5.29**.

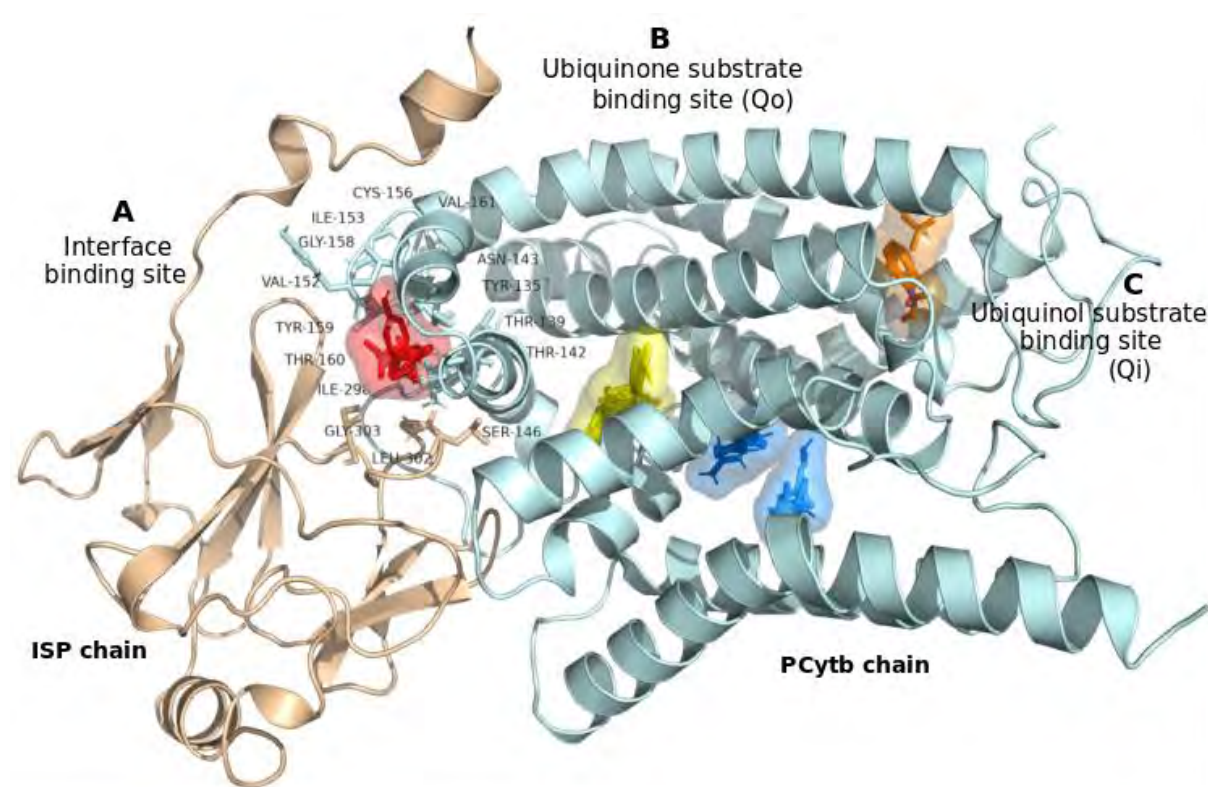


Figure 5.29: Structural representation of the identified pockets in the PfCytb-ISP protein using FTMap. Identified pockets include the canonical Qo, Qi active sites, putative pockets as well as the protein's interface pocket showed as surface and coloured in red. PfCytb and ISP chains, Qo, Qi and the putative pockets are coloured teal, brown, yellow, orange, and blue, respectively.

The interface pocket was identified to be located 19.2 Å away from the Qo active site (**Figure 5.29:** coloured in yellow) and was noted to be characterised by conserved motifs. As such, finding compounds that would bind at the interface and infer changes to the Qo active site (distal site) suggest they could have allosteric properties. The binding pocket identified by FTMap probes composed of cross-cluster1 (probe ID: 1chx, 1ben, 1bdy, 1dfo, 1but, 1phn, 1dfo, 1ths) and cross-cluster 2 (probe ID: 1but, 1dfo, 1phn, 1ths, 1ben, 1ths, 1phn) is seen to be located between alpha-helices of PfCytb and flexible loops of ISP subunits. Further, analysis of the pockets showed residue interactions formed by both non-bonded as well as hydrogen bonds, which were performed within 4 Å of the pocket. Herein, residues Tyr135, Thr139, Thr142, Ser146, Val152, Ile153, Cys156, Gly158, Tyr159, Thr160, Ile298, Leu302 and Gly303 of the interface pocket were identified as shown in **Figure 5.29**. Out of these residues, Thr139 and Thr142 are shared with the Qo active site, showing the two pockets are near each other according to the distance outlined previously. The two residues (Thr139 and Thr142 which formed part of conserved motif 3) and active site residues too, are essential in the stabilisation of the ubiquinone substrate [203]. Some of the remaining pocket residues were part of motif 3 (135, 139, 142, 146, 152, 153, 156, 159, 160). Upon binding of the potential allosteric modulators, these conserved residues are likely to participate in residue interactions with some Qo active site residue to enhance substrate binding. The shared binding site residues evidence the proximity of the two active sites. From this analysis, virtual screening of both natural (SANCDB) and MMV compounds was performed on this pocket, and their dynamics were examined.

5.3.19 Docking output: Thirteen selective compounds were identified to bind in the PfCytb-ISP heterodimer interface

An exploratory research study was performed on the PfCytb-ISP interface with the aim of identifying compounds that could potentially affect the protein's function. This is essential to assess whether the binding of compounds in this interface pocket could transmit binding effects to the functional site (especially the Qo active site). Therefore, MMV and SANCDB compounds were screened and further analysed for their molecular dynamics. **Figure 5.30** shows the compound screening workflow used, and the compounds identified to bind selectively in the PfCytb-ISP interface pocket.

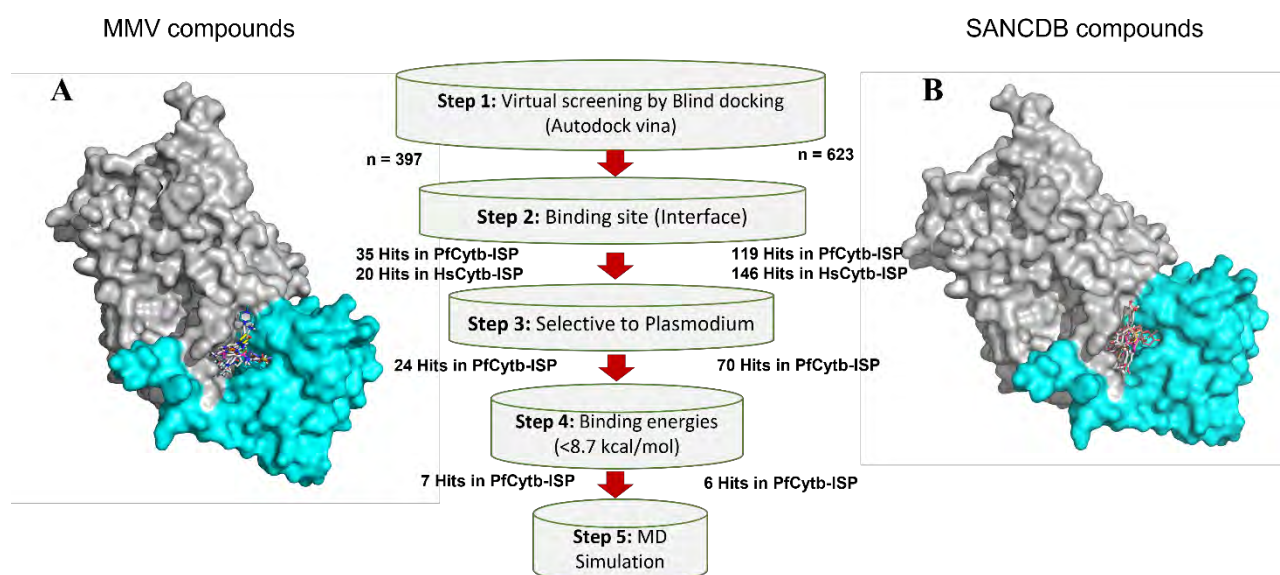


Figure 5.30: Virtual screening workflow showing filtering criteria based on binding site (*Plasmodium* interface), energy ranking and selectivity. The number of hit compounds selected at each filtering steps is indicated. Selective docked (A) MMV and (B) SANCDB compounds highlighted in protein structure.

From the screening of both *Plasmodium* (PfCytb-ISP) and human (HsCytb-ISP) proteins, a total of 13 selective compounds were identified with a binding energy of < -8.7 kcal/mol⁻¹. The lower the binding energy, the greater the binding affinity in the pocket. These compounds and their respective binding energies include: MMV020982 (-8.9 kcal/mol⁻¹), MMV023860 (-8.8 kcal/mol⁻¹), MMV024101 (-8.8 kcal/mol⁻¹), MMV024937 (-9.0 kcal/mol⁻¹), MMV676050 (-9.4 kcal/mol⁻¹),

MMV676599 (-9.3 kcal/mol⁻¹), MMV687700 (-8.9 kcal/mol⁻¹), SANC00257 (-9.7 kcal/mol⁻¹), SANC00264 (-10.2 kcal/mol⁻¹), SANC00347 (-8.9 kcal/mol⁻¹), SANC00559 (-9.2 kcal/mol⁻¹), SANC00587 (-8.9 kcal/mol⁻¹) and SANC00595 (-9.4 kcal/mol⁻¹). These were subjected to MD simulation to assess their effects on the protein dynamics and function. 2D chemical structures of the 13 interface compounds are shown (**Figure 5.31**).

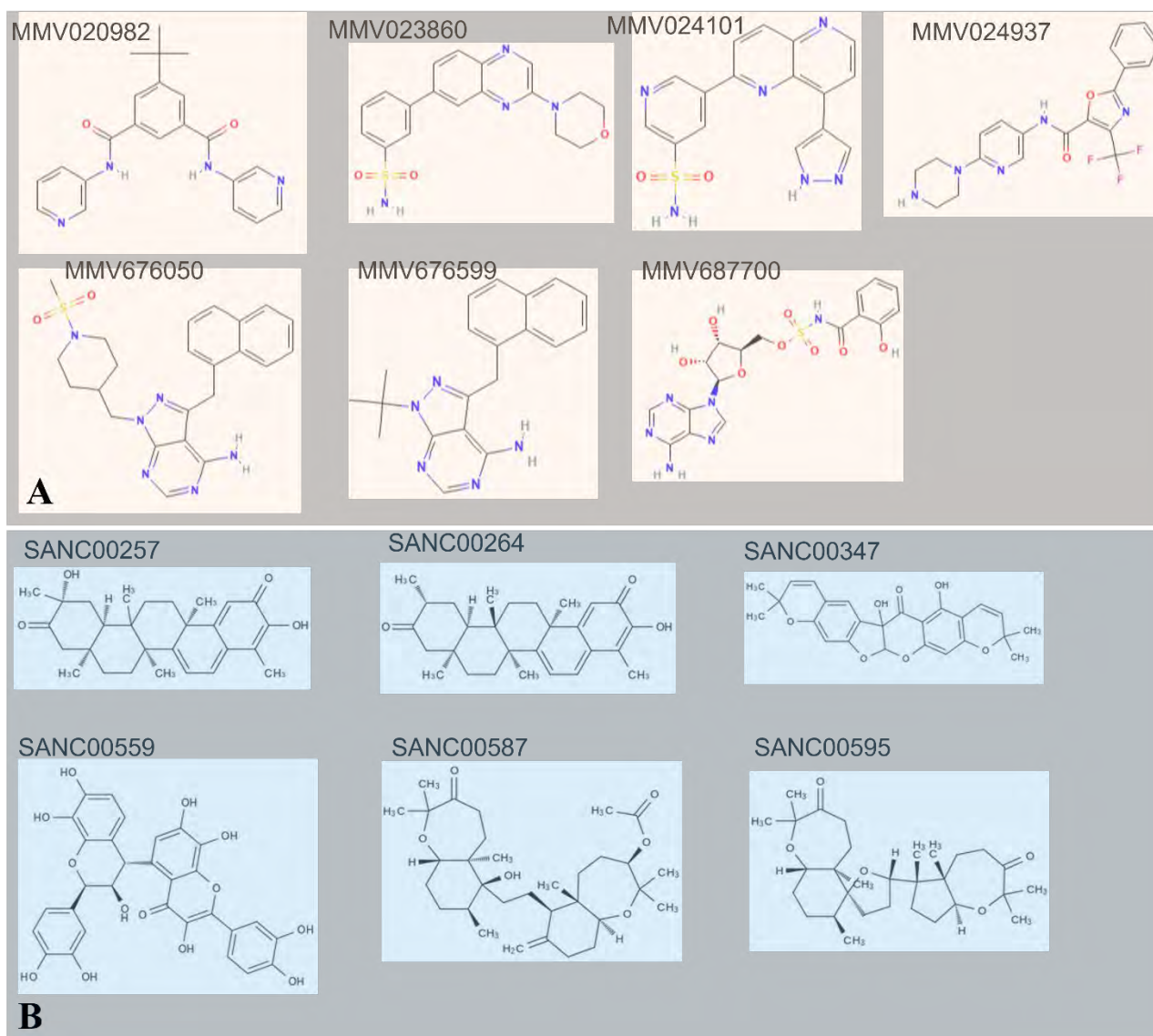


Figure 5.31: 2D structures of (A) MMV and (B) SANCDB compounds that bound in the PfCytb-ISP heterodimer interface with binding energies < -8.7 kcal/mol⁻¹.

5.3.20 Analysis of the stability of the interface selective compounds in the *Plasmodium* interface pocket via molecular dynamics simulations

Firstly, the post-docking analysis was performed on the interface selective compounds for residue molecular interactions formed with the protein around the interface pocket. Molecular interactions including non-bonded (e.g. vdW) and bonded (e.g. hydrogen bonds) interactions have been reported to structurally stabilise protein-ligand complexes [350,351]. As such, the interactions predicted in **Figure 5.29** were used to compare with those formed by the 13 selective compounds. Most of the interactions were formed by all the selective compounds, as shown in **Figure 5.32**.

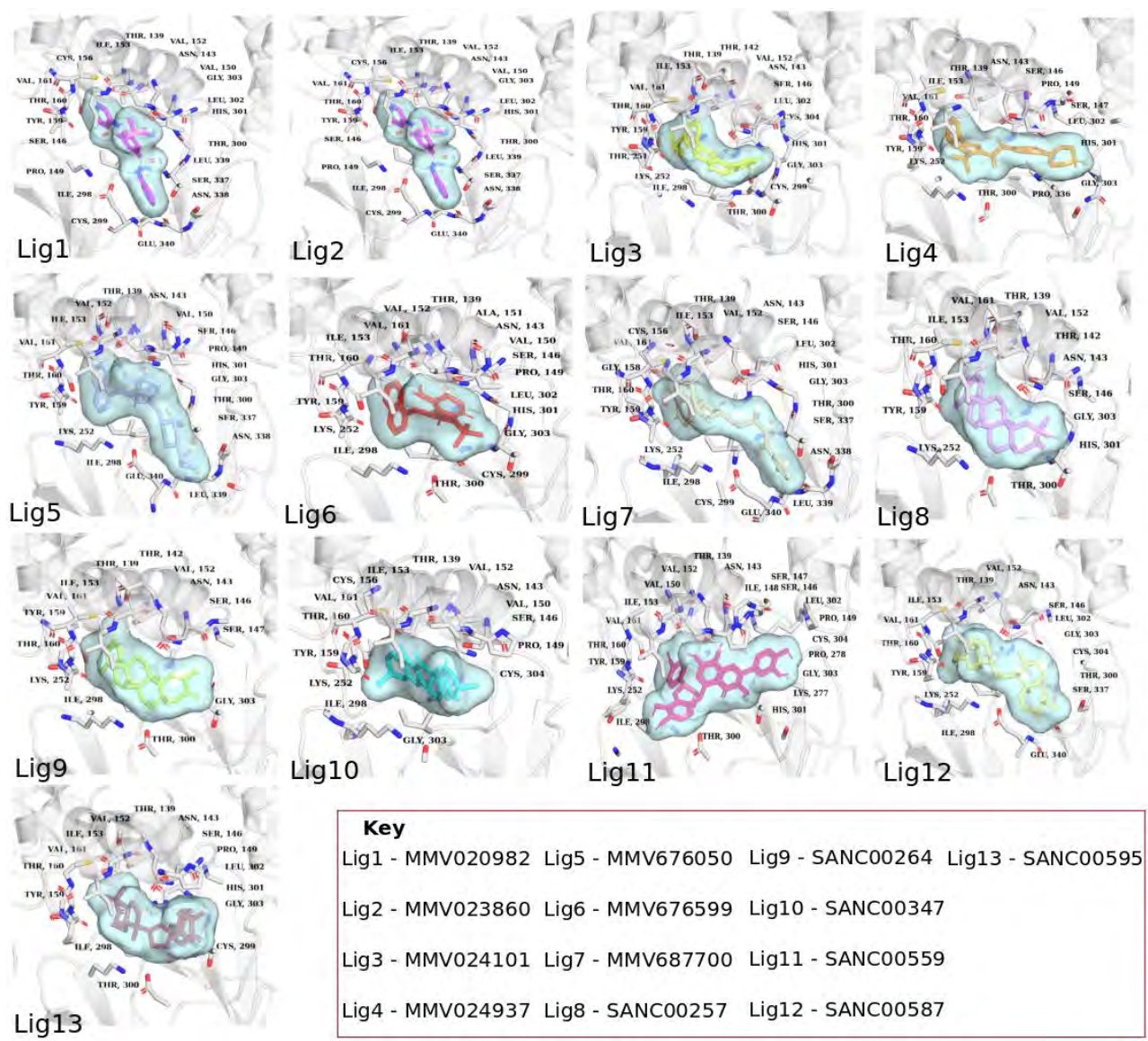


Figure 5.32: 3D ligand interaction network for 13 selective compounds bound in the PfCytb-ISP interface pocket as highlighted in the figure. The protein and the compounds are displayed in cartoon and stick representations, respectively.

Out of these interactions, residues Thr139, Thr142, Asn143, Ser146, Thr300 and His301 formed hydrogen bonds while the rest of the interface residues formed van der Waals (vdW) interactions. Overall, vdW interactions were the majority, and previous studies report their ability to support ligand stability [326,327]. Most of the compounds formed hydrogen bonds with residues Thr139, Thr142, Ser146 and Asn143, while compounds such as MMV024101, MMV024937, MMV687700 and SANC00559 formed hydrogen bonds with residue His301, a [2FE-2S] cluster coordinating residue.

This is because the compounds bound in the pocket with their tails directed towards the cluster. These interactions were formed in static structure, thus evidencing the high likelihood of the ligands remaining stable in the interface pocket.

To further assess their stability, 350 ns MD simulations were performed in both WT and in the presence of Y268C/N/S mutations. Conformational variability of each ligand in WT and mutant states was plotted as shown in **Figure 5.33A**. Using WT as a reference, all compounds showed their stability in the pocket as shown by RMSD values $< 2 \text{ \AA}$, except compounds MMV676050, MMV676599 and SANC00585. Most of the ligand systems showed unimodal RMSD distributions except MMV024101, MMV024937, SANC00257 and SANC00587. The unimodal distribution indicates that the compounds remained stable in the interface pocket. On the contrary, trajectory analysis of those that displayed multimodal distributions confirmed that the compounds remained in the pocket as well despite displaying flipping movements in their structure.

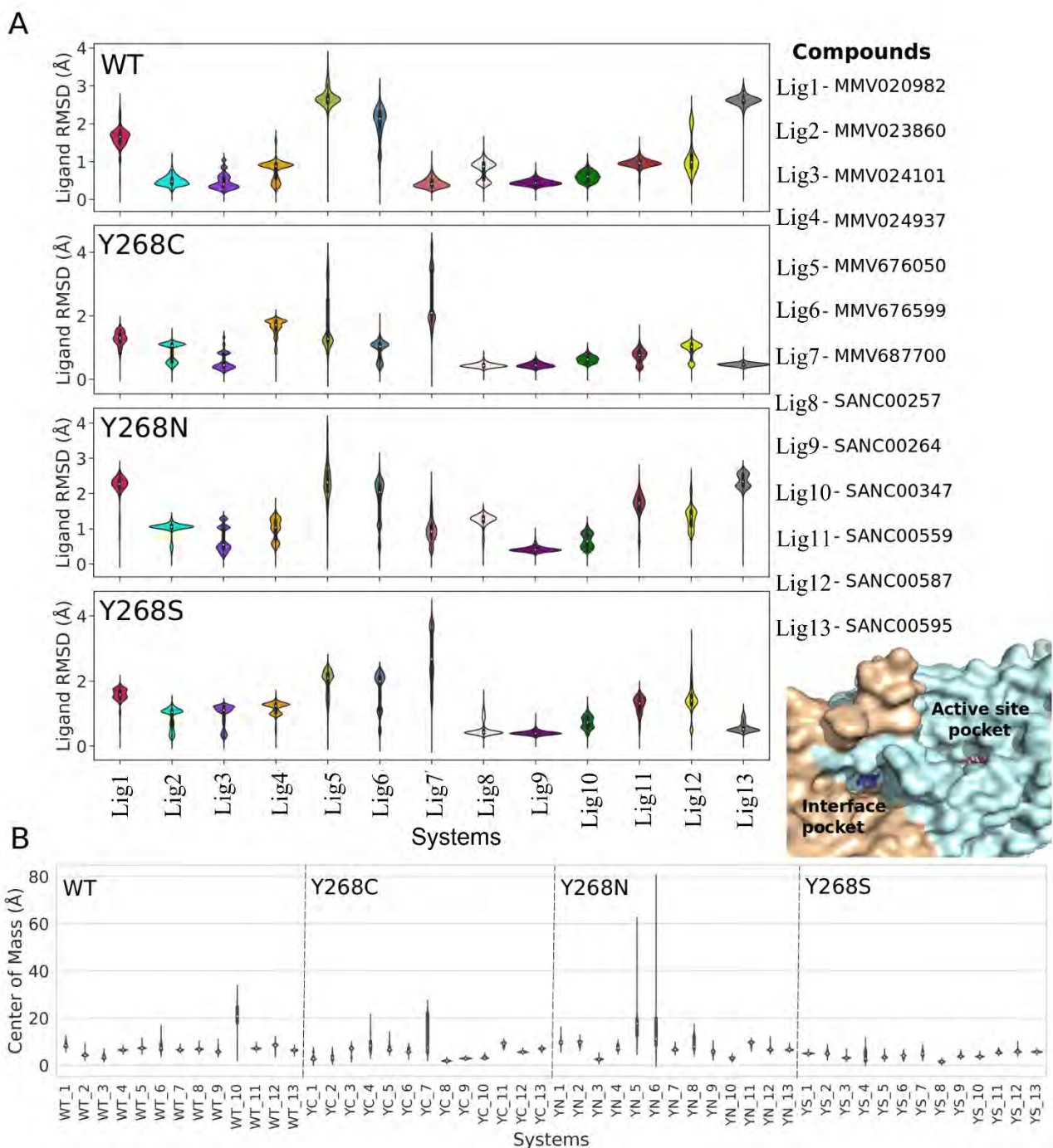


Figure 5.33: Kernel density distributions of (A) Ligand RMSD and (B) COM for 13 interface selective compounds monitored across 350 ns MD simulation. Both ligand RMSD and COM were analysed in a total of 52 ligand systems, assessing their behaviour in the presence of mutations (Y268C, Y268N and Y268S) in comparison to the WT as the reference. Labels Lig1-Lig13 are ligand systems labelled in the figure.

Regarding the most stable ligand conformations, ligand system SANC00264 was observed (**Figure 5.33A**) to have unimodal distribution in not only reference (WT) protein but also in the presence of

three mutations. This depicts its stability during MD simulation and its potential to be the best allosteric modulator as supported by its high binding energy (-10.2 kcal/mol). From these results, when the compounds bound at the protein's interface, changes such as multimodal distributions were reported due to the mutations. This is in agreement with our previous studies which highlighted the effect of mutation on the ATQ drug [141].

Additionally, COM distance between the interface pocket residues and the 13 compounds was performed to confirm the stability of the ligands in the pocket. This was performed for consensus with the ligand RMSD findings. Here, **Figure 5.33B** showed COM distance plots for each of the 13 protein-ligand complexes, which confirmed all compounds were stable in all mutant systems except ligand systems MMV676050 and MMV676599 in Y268N mutant protein. Visual inspection of the two systems showed their slight movement away from the interface pocket and back, suggesting that the two compounds may not be optimal as allosteric modulators.

Hydrogen bond interactions were also investigated as they are known to stabilise ligands at the target site and promote drug efficacy, thus their importance in biological processes [335]. **Figures 5.34** and **5.35** show the protein residues that formed hydrogen bonds (further referred to as h-bonds) with the MMV and SANCDB compounds, respectively. Here, the h-bonds are represented as a fraction (0 – 1) representing their occupancy (0 % - 100 %) during MD simulation. Thus, h-bonds observed in < 10 % of MD simulation time were considered short-lived/weak therefore not significant in providing ligand stability in the interface pocket. For instance, **Figure 5.34a** shows that a hydrogen bond was formed between the nitrogen atom of MMV020982 compound and oxygen atom of Ile298 residue 40% of the simulation time. All systems were reported to form hydrogen bonds with at least one protein residue which support their stability in the interface pocket. Out of all the residues involved in hydrogen bond formation, residues Thr139, Asn143, Ser146 and Val161 were frequently reported

across the ligand systems. Focusing on the 3D ligand interactions formed on the protein static structure after docking (**Section 5.3.20**), these residues were similarly reported.

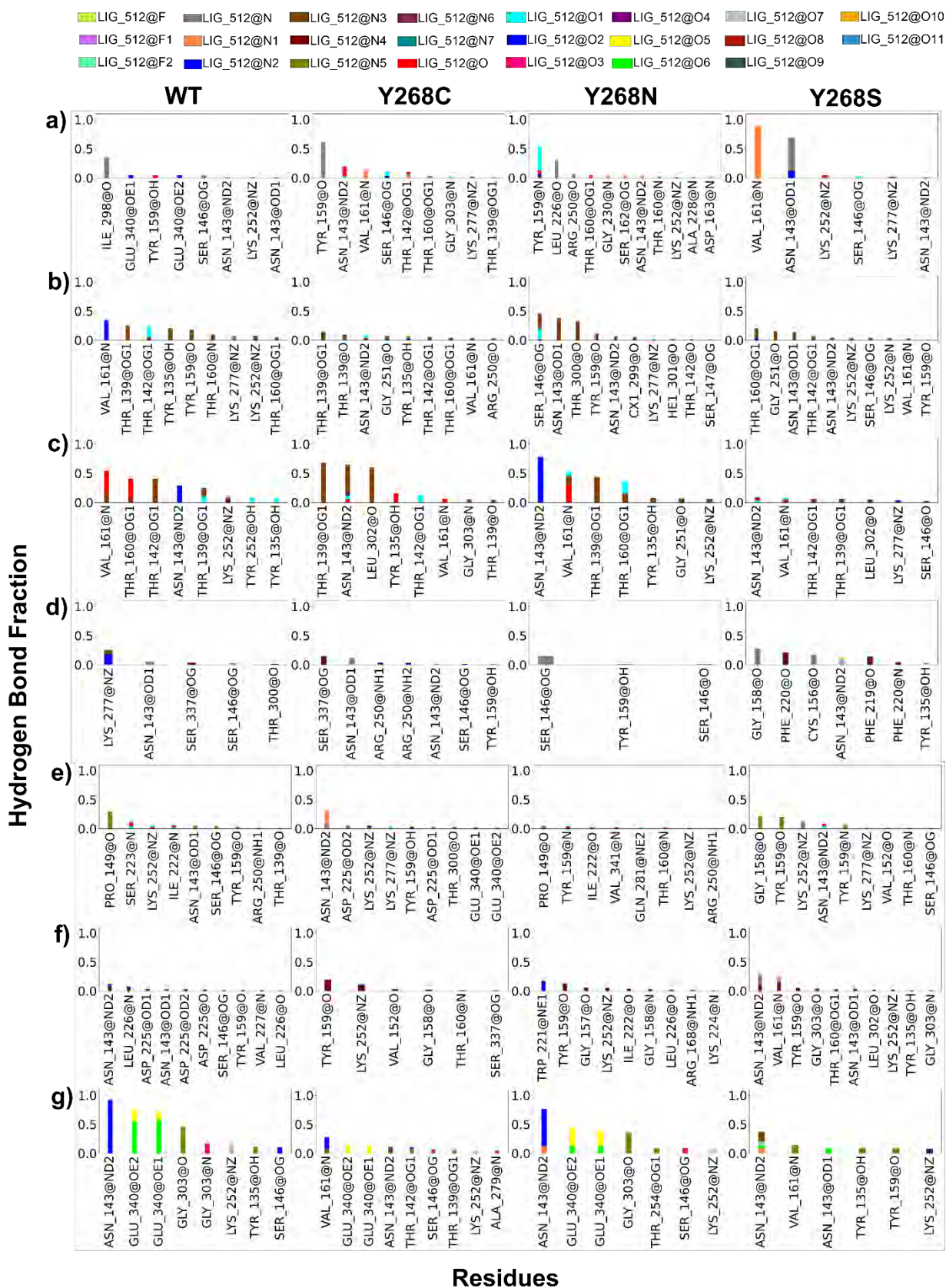


Figure 5.34: A plot of the average number of inter-molecular hydrogen bonds formed between the seven MMV compounds and PfCytb-ISP protein in four protein systems (WT, Y268C, Y268N and Y268S) during 350 ns MD simulations. Compounds labelled a-g representing MMV020982, MMV023860, MMV024101, MMV024937, MMV676050, MMV676599 and MMV687700,

respectively. All the compounds were labelled LIG512 for easy reference in the colour code provided. The colour code is provided which highlights the type of atoms that formed hydrogen bonds with PfCytb-ISP protein.

Altogether, the 13 ligand systems displayed unique profiles in terms of h-bond formation. Regarding h-bonds with the highest occupancy in WT, **Figures 5.34** and **5.35** report all systems formed lead h-bonds with > 10 % occupancy except SANC00347, whose leading h-bond with Asn143 residue was observed 4% of the simulation time. In Y268C protein, all systems reported lead h-bonds with occupancy > 10 %. As for Y268N protein, all systems formed lead h-bonds with > 10 % occupancy except MMV676050 and SANC00264, whose leading h-bonds with Pro149 and Thr139 residues were observed 4% of the simulation time in each case.

hydrogen bond formation was comparable between reference and mutant proteins, suggesting the mutations did not affect the number of h-bond interactions formed in the interface pocket.

5.3.21 Protein backbone flexibility and structural conformations in ligand systems and in the presence of mutations

Herein, reference (WT) and mutant ligand systems were monitored during 350 ns MD simulations and their structural conformations analysed. Using backbone RMSD and Rg, comparison was performed to assess the effect of mutations in the 13 ligand systems. **Figure 5.36A** shows comparable RMSD profiles across all 13 ligand systems averaging 2.7, 2.9, 2.9 and 3.2 Å in WT, Y268C, Y268N and Y268S systems, respectively. In the presence of the three mutations, most of the systems reported relatively higher RMSD than their respective WT systems, structural conformational changes due to the mutations. This is likely to cause structural conformational changes affecting stability and function in the systems. However, it has not yet been established if the observed higher RMSD corresponds to the higher instability predicted values reported in Chapter 3. All systems reported unimodal distribution except MMV024101 (Y268C), MMV023860 and SANC00587 in Y268N system and MMV023860, MMV024101 and SAN00347 in the Y268S mutant system. In terms of the severity of the mutation, we observed that all three mutations can cause conformational changes in the PfCytb-ISP protein. The 13 compounds are also reported to behave differently in the presence of each of the three mutations despite the fact that the mutations occur in the same location, which agrees with previous studies [141].

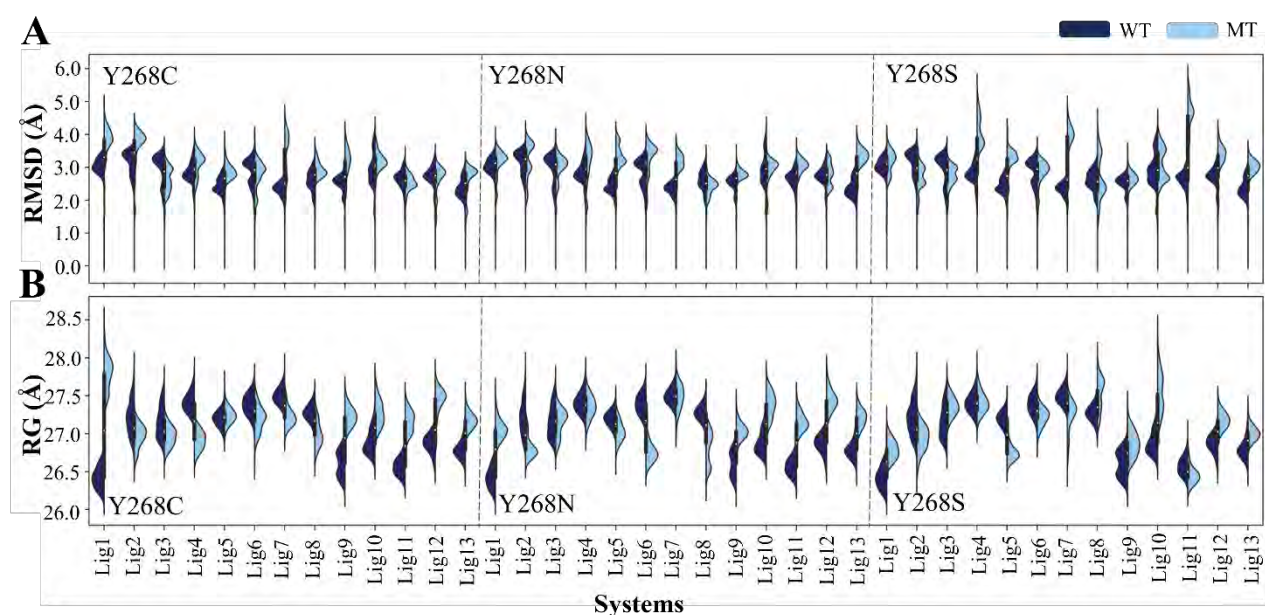


Figure 5.36: Violin plot presentation of (A) RMSD and (B) Rg for 13 ligand systems during 350 ns MD simulations. Ligand systems Lig1-Lig13 represent MMV020982, MMV023860, MMV024101, MMV024937, MMV676050, MMV676599, MMV687700, SANC00257, SANC00264, SANC00347, SANC00559, SANC00587 and SANC00595 respectively. Comparative analysis was performed between each mutant ligand system and its corresponding WT ligand system as a reference; WT: Wildtype in dark blue and MT: Mutant in light blue colours.

Using all-atom Rg (**Figure 5.36B**), the stability of the protein in the 13 ligand systems (WT, Y268C, Y268N and Y268S systems) was further assessed based on its compactness. Here, most of the WT systems showed unimodal distribution except ligand system SANC00264, where bimodal distribution is observed. Despite this observation, this ligand system showed more compactness as compared to its respective mutant proteins. All the ligands displayed a unique profile in the mutant systems showing their varied impact on the protein compactness.

Estimating residue fluctuations in ligand systems compared to the reference and mutant systems was performed using RMSF (**Figure S5.19**). Local residue flexibility was also studied in the 13 ligand systems, including ATQ, and comparative analysis was performed in both reference and mutant proteins. As analysed in PART I and II, the 13 interface compounds report high flexibility

corresponding to the loop regions reported previously [141], while the regions experiencing no flexibility were traced back to the core of the protein including the active site region.

5.3.22 DRN centrality metrics analysis to determine protein communication patterns and ‘hubs’ upon binding of compounds within the PfCytb-ISP interface pocket

Protein residue networks are important in the study of protein communication patterns. Hence, DRN analysis was utilised to identify significant communication residues and analyse the effect of mutations on their communication pattern on the compounds that bound at the heterodimer interface. These findings were then correlated to motif analysis outlined in chapter 3.

According to **Figures S5.20 – S5.24**, residues central to protein function were identified across holo and ligand systems using *Averaged BC, CC, DC, E, and KC* metrics. The figures show the heat maps of the top 4% of each metric representing WT and mutant proteins across the 13 ligand systems calculated in each centrality metric. ‘hubs’, as previously described, were identified across holo and ligand-bound systems (**Table 5.6**). These residues were further discussed under each metric except *EC* metric which reported no *hubs* in the systems.

In *BC* metric (**Table 5.6**), reference (WT) protein reported no hubs, while in the presence of mutations Y268C/N/S, none was reported in Y268C, residue Val85 was reported for (Y268N and Y268S) mutation, while His320 was reported in Y268S mutant (**Table 5.6**). These were *hubs* in the 13 interface compounds including ATQ ligand systems. Motif analysis report residue Val85, a part of conserved motif 1, to increase protein residue communication in the presence of Y268N and Y268S mutations. It is evident that these hubs gained function showing the null effect ligand binding has on protein function. Specifically, we highlighted His320, which is covalently bound to the [2FE-2S] cluster and is thus referred to as a metal coordinating

residue. The cluster, therefore, plays a role in the redox process, which supports ET pathway across the IMM [337].

In *CC* metric (**Table 5.6**), a higher number of *hubs* were identified which highlighted residues linked to all other residues in a network through the shortest path. These were residues Ala122 (WT, Y268C/S), Phe123 (WT, Y268C/N), Phe264 (WT, Y268C/S), Gly137 (Y268C/S), Tyr126 (Y268S) and Tyr263 (WT). The hubs formed part of motif 2 (263, 264) and motif 3 (122, 123, 126, 137). Out of these hubs, Gly137 and Phe264 are Qo active site residues important in substrate binding and stability [107]. On the other hand, Tyr263 forms the conserved PEWY (Pro260-Glu261-Trp262-Tyr263) motif 2, also known as the Qo motif [311]. The PEWY motif found on a loop region of this protein has been reported to participate in substrate binding. Similarly, due to the conserved nature of the motifs/protein, inter-protein residue communication in these systems is likely to be preserved.

Table 5.6: Protein communication ‘*hubs*’ identified from holo, ATQ and 13 ligand systems across five DRN metrics in WT and mutant proteins.

	WT	Y268C	Y268N	Y268S
BC	-	-	Val85	Val85
	-	-	His320	-
CC	Ala122	Ala122	-	Ala122
	Phe123	Phe123	Phe123	-
	Phe264	Phe264	-	Phe264
	Tyr263	-	-	Tyr126
	-	Gly137	-	Gly137
DC	Cys299	Cys299	Cys299	Cys299
	Ser322	-	Ser322	Ser322
EC	-	-	-	-
KC	-	Ile35	-	-
	-	Val85	-	-
	-	Cys299	Cys299	Cys299
	-	Ser322	-	Ser322
	-	Tyr324	-	Tyr324

In *DC* metric (**Table 5.6**), residues Ile35 (Y268C), Val85 (Y268C), Gly131 (Y268S), Cys299 (WT, Y268C, Y268N and Y268S) and Ser322 (WT, Y268N and Y268S) were identified as *hubs*. From these residues, Cys299 is covalently linked to the [2FE-2S] cluster, thus referred to as a metal coordinating residue. In binding to the cluster, the electron transfer process in the protein is maintained. This residue was evident in not only the reference protein but also the mutant proteins, suggesting that the mutations possibility will not affect the cluster's structure and ultimately its participation in electron transfer. These residues formed the following conserved motifs; motif 1(85), motif 3 (131) and motif 5 (35).

In *KC* metric (**Table 5.6**), residues Ile35 (Y268C), Val85 (Y268C), Cys299 (Y268C, Y268N and Y268S), Ser322 (Y268C and Y268S) and Tyr324 (Y268C and Y268S) were identified as *hubs*. As discussed in the previous metric, Cys299 is covalently linked to the [2FE-2S] cluster, supporting the electron transfer process. Residues Ser322 and Tyr324 are located near the metal coordinating residues in support of the cofactor geometrical structure. Residues Val85 and Ile35 were confirmed to be part of conserved motif 1 (85) and motif 5 (35). They were confirmed to be located between the two heme cofactors (heme bL and heme bH) at the transmembrane alpha helices. Because the heme cofactor plays a role in redox reactions during the ET process, the residues are likely to form communication hubs that could ultimately support the ET pathway.

Overall, using the five centrality metrics, the communication hubs showed similarities across the 13 interface compounds and were found to form part of the conserved motifs identified in Chapter 3; mostly motif 2 and 3. The study has gained insights into the communication pattern changes occurring in the 13 ligand systems. Most of the *hubs* highlighted changes occurred not only in the Qo active site but also at the ISP metal coordination sphere (along the ET pathway). Thus, analysis of the hubs from interface binding confirms the potential allosteric modulation of the 13 selective hit compounds. Incorporating mutations and comparing the metrics across all systems has shown that there are distal changes that were implicated in the ubiquinol

substrate binding site. Further, besides the *CC* metric, which reported Qo active site residues Gly137, Tyr263 and Phe264 as *hubs*, there were no other active site residues observed in the other four centrality metrics. Through the *CC* metric, the identified hubs suggest that these ligand systems lost protein communication within the active site. As such, further *in vitro* studies should be performed to confirm these findings in support of future drug discovery studies.

5.3.23 Effect of mutation on residue-residue interaction changes on 13 interface ligand systems; Contact map analysis

The 13 interface compounds were further analysed using weighted residue contact maps [308]. This was performed to investigate any changes at the heterodimer interface upon binding of ligands and whether protein function at the Qo active site would be altered. The residue contact maps were determined by applying a cut-off distance of 6.7 Å for each residue as explicitly detailed in the methodology section. In this section, residue-residue interactions were performed on a total of 28 Qo active site residues for each of the 13 ligand systems in both WT and three mutant proteins ($13 * 4 = 52$ systems in total). According to previous studies, the types of residue-residue interactions consist of vdW, hydrogen bonds and electrostatic bonds [317]. The active site residues are essential in supporting substrate binding, especially Glu261, which formed hydrogen bonds thus promoting substrate stability and protein function [288]. The objective here was to determine the effect of compounds binding at the interface in the presence of mutations to see whether they would cause changes in the Qo active site of the protein. Previous studies have showed that the three mutations' location within the Qo active site pocket compromises residue interactions in both holo and ATQ-bound systems [141].

We compared WT ligand-bound systems (as reference proteins) to their respective mutant ligand systems to obtain information on residue interactions reduced, lost, or gained (**Table 5.7**). Here, all ligand systems reported reduction or loss of at least five residue-residue interactions on average. The interactions observed in at least two mutant systems include: Pro260-Val140, Ile258-Cys481, Pro260-Leu265, Tyr263-Phe123, Phe264-Phe123, Ile119-Tyr263, Gly137-Pro260, Ile258-Val140, Val140-Gly465, Thr142-Cys156, Tyr263-Ala122 and Pro260-Met133. Out of these, hydrophobic interactions Phe264-Phe123, Pro260-Leu265 and Tyr263-Ala122 were similarly reported in the Qo active site ligand systems (PART II). These residue interactions were reported in at least two mutant systems, showing that the mutations cause changes upon binding the 13 compounds in the PfCytb-ISP interface. These results are in agreement with those displayed by ATQ-bound systems [141]. Regarding gained interactions, most of the ligand systems did not gain new interactions except MMV023860, MMV024937, SANC00257 and SANC00559 systems. As such, the results evidence the interface compounds as potential allosteric modulators due to their effect on the Qo active site.

Also, some of the residues experiencing reduction or loss in interactions have been reported in the previous section (**Section 5.3.22**) as communication hubs. These include Ala122 (*BC, CC*), Phe123 (*CC*), Gly137 (*CC*), Phe264 (*BC, CC*) and Tyr263 (*BC, CC*). These residues formed part of conserved motif 2 (263 and 264) and motif 3 (122, 123, 137).

Table 5.7: Residue-residue interactions observed to be either reduced/lost or gained/enhanced in the 13 ligand systems across mutant (Y268C, Y268N and Y268S) systems as compared to WT. Residue interactions reduced/lost/gained across all the three mutation ligand systems (coloured in blue), those found in two mutant ligand systems (coloured in green) and interactions reduced/lost in 9 out 13 ligand systems (coloured in red) are shown.

Ligand systems	REDUCED/LOST			GAINED/ENHANCED		
	Y268C	Y268N	Y268S	Y268C	Y268N	Y268S
MMV020982	Ile258-Cys466, Ile258-Cys481 , Thr142-Cys156, Pro260-Tyr126 , Phe264-Val259, Val140-Gly465, Leu145-Ile155	-	Ile258-Cys481 , Ile258-Gly137, Pro260-Tyr126	-	-	Tyr263-Phe123
MMV023860	Gly137-Pro260, Pro260-Val140 , Glu261-Tyr263	Ile258-Cys481, Trp262-Ser83, Pro260-Leu265	Pro260-Leu265 , Pro260-Val140 , Trp262-Ser83	-	Ile258-Trp136	Tyr263-Phe123
MMV024101	Ile141-Gly137, Thr142-Ala138, Val140-Gly465, Tyr263-Phe123 , Ile155-Arg168	Tyr263-Ala122 , Val259-Leu265, Ile258-Cys481, Trp262-Phe86 , Trp262-Ser83, Ile119-Tyr263 , Ile119-Phe264 , Tyr263-Phe123	Trp262-Ser83, Ile155-Thr165, Ile258-Cys481, Trp262-Phe86 , Ile119-Phe264 , Ile119-Tyr263 , Tyr263-Ala122, Tyr263-Phe123	-	-	
MMV024937	Gly137-Pro260 , Ile119-Tyr263, Pro260-Val140 , Tyr263-Ala122 , Tyr263-Phe123	Leu271-Val284, Trp262-Phe86 , Gly137-Pro260, Pro260-Leu265, Pro260-Val140 , Tyr263-Phe123 , Met133-Gln257	Gly137-Pro260 , Glu261-Leu265, Met133-Gln257, Ile258-Cys481, Trp262-Phe86 , Thr142-Val152, Pro260-Val140 , Tyr263-Phe123	Ile258-Pro255	Ile258-Trp136	-
MMV676050	Ile258-Val140 , Val140-Gly465	Tyr263-Ala122 , Trp262-Ser83, Ile258-Val140 , Val140-Gly465 , Leu145-Val152	Ile258-Val140 , Ile258-Gly137, Trp136-Thr254, Met133-Val259	-	-	-
MMV676599	Gly137-Pro260 , Pro260-Val140 , Glu261-Tyr126, Glu261-Tyr263, Leu145-Val152, Ile155-Arg168	Val259-Leu265, Ile258-Val140, Ile258-Gly137, Gly137-Pro260 , Pro260-Val140 , Leu145-Val152	Leu145-Val152, Ile258-Met133, Ile119-Phe264, Thr142-Thr139, Thr142-Cys156, Thr142-Val152, Thr142-Val161, Pro260-Tyr126	-	-	-
MMV687700	Ile258-Val140 , Ile258-Cys481 , Gly137-Pro260 , Pro260-Val140	Trp262-Ser83, Leu271-Val284, Ile258-Val140 , Ile1258-Cys481 , Ile119-Tyr263, Gly137-Pro260 , Pro260-Val140 , Tyr263-Ala122	Ile258-Val140 , Trp262-Ser83, Trp136-Thr254, Pro260-Val140	-	Tyr263-Ser83	-
SANC00257	Trp136-Met133, Gly137-Pro260, Ile119-Tyr263 , Ile141-Gly137, Thr142-Cys156 , Gly137-Val140, Pro260-Met133 , Glu261-Tyr126, Tyr263-Ala122	-	Met133-Gln257, Ile119-Tyr263 , Thr142-Cys156 , Pro260-Met133 , Pro260-Tyr126, Tyr263-Ala122	Ile258-Pro255	-	Pro260-Val140

SANC00264	Ile155-Arg168, Ile119-Tyr263 , Ile141-Gly137, Gly137-Val140, Val140-Gly465 , Tyr263-Ala122	Val140-Gly465 , Met133-Gln257	Glu261-Leu265, Leu145-Val152, Ile119-Tyr263 , Thr142-Thr139, Tyr263-Ala122 , Ile155-Arg168	-	-	-
SANC00347	Pro260-Met133 , Pro260-Tyr126, Phe264-Val259, Leu145-Ile155	-	Pro260-Met133	-	-	-
SANC00559	Ile155-Arg168	Trp262-Ser83, Val140-Ile258, Ile258-Gly137, Tyr263-Ala122	Gly137-Pro260, Trp262-Ser83, Thr142-Val161, Tyr263-Ala122 , Met133-Val259, Ile155-Arg168	-	Tyr263-Ser83	Pro260-Val140
SANC00587	Ile258-Cys466, Ile141-Gly137, Thr142-Ala138, Val140-Gly465 , Tyr263-Ala122, Leu145-Val152	Tyr263-Ala122 , Phe264-Phe123 Val140-Gly465	Ile258-Met133, Pro260-Leu265	-	-	-
SANC00595	Trp136-Met133, Glu261-Tyr126 Tyr263-Ala122	Ile119-Phe264, Phe264-Phe123	Ile155-Thr165	-	-	-

Notable residue-residue interaction changes around the identified hubs (refer to Section 5.3.22) were calculated and discussed (**Table S5.9**). Across all three mutations, hydrophobic interactions Gly137-Pro260 (MMV023860, MMV024937, MMV676599, MMV687700, SANC00257 and SANC00559), Tyr263-Phe123 (MMV024101 and MMV024937) and Ala122-Tyr263 (in all ligand systems except MMV020982, MMV023860, MMV676599 and SANC00347) were observed to be reduced/lost. This observation suggests that the three mutations are likely to share resistance mechanisms as previously reported [141]. Besides the interactions mentioned before, the hydrophobic interactions Phe264-Ile119 and Tyr263-Ile119 were observed to be reduced/lost in at least two mutant systems. It was further noted that newly formed interactions were reported in Y268N and Y268S mutants only, including Tyr263-Ser83 (MMV687700 and SANC00559) and Tyr263-Phe123 (MMV020982 and MMV023860) hydrophobic interactions. Further, these residues were identified to form part of conserved motifs; motif 2 (263, 264) and motif 3 (122, 137). The key residue Tyr263 and Phe 264 forms a part of the conserved PEWY conserved motif (Motif 2) which is known to facilitate Qo substrate binding

[201]; as such the function in the mentioned ligand systems is likely to be maintained considering the balance in reduced and gained residue interactions in these systems.

5.4 Chapter conclusion

Even though targeting the *P. falciparum* cytochrome *b* remains a complex issue, previous studies have shown that this drug target is still attractive as its inhibition affects cellular respiration consequently leading to parasite cell death. While many studies inhibit the drug target using synthetic drug compounds, this study explores mainly natural compounds reported to have good structural complexity. Here, we have utilised natural South African compounds and clinically approved MMV (Medicines for Malaria Venture) compounds for repositioning purposes. The computational drug discovery approach in this study incorporated important reproducible calculations such as structure-based virtual screening and MD simulations in a phospholipid bilayer membrane that contributed to the reliability of the findings. Further, global and local residue analysis was performed using RMSD, Rg, COM, DRN centrality metrics and weighted residue contact maps to gain insights into this selective drug design study.

From 350 ns MD simulation results, the effect of mutations (Y268C, Y268N and Y268S) on the protein was assessed on holo and ATQ-bound systems. Using *in silico* methods, the three mutations were reported to share a similar resistance mechanism by destabilising the protein thus directly affecting the Qo substrate binding site. Active site residue communication was also reported to increase in the presence of mutations, which suggests the parasite's ability to evade drug action mechanism. These insights were used to screen for selective compounds that can withstand these mutations.

Following virtual screening, this study has identified a total of eight selective hit compounds from Class I (MMV019551, SANC00373 and SANC00459) and Class II (MMV007638, MMV636004, MMV676444, SANC00548 and SANC00725) that could be potential inhibitors of the *P. falciparum* cytochrome *bc₁* complex. Based on docking binding energies, the results

showed that these bound strongly in the *P. falciparum* cytochrome *b* Qo active site. These findings were further supported by MD calculations of these compounds in the human homolog, which showed that all the eight compounds were unstable in the Qo active site. This revealed a distinct selective inhibition of these compounds against the *P. falciparum* cytochrome *b*. The eight compounds were reported to bind and remain stable in the parasite's substrate binding site in the WT and the three mutant models (Y268C, Y268N and Y268S). This was supported by the hydrogen bond and vdW interactions formed and sustained throughout MD simulations. It was also noted through DRN global and local analysis that the eight compounds were likely to inhibit the protein not only by remaining stable in its active site but also by disrupting important residue-residue interactions. Therefore, the findings so far provide sufficient evidence to suggest that the 8 selective hit compounds bind to the Qo active site and have the potential to overcome ATQ drug resistance.

Additionally, 13 interface compounds (MMV020982, MMV023860, MMV024101, MMV024937, MMV676050, MMV676599, MMV687700, SANC00257, SANC00264, SANC00347, SANC00559, SANC00587 and SANC00595) have been identified and observed to have distal effects. Upon their binding at the heterodimer interface, the results showed changes at the Qo active site where compromised interactions were reported within the active site pocket (distal cavity). DRN analysis reported a loss of protein communication because a small number to none of the active site residues were reported as hubs across the five centrality metrics in WT and the mutant systems. Besides this, contact analysis of the 13 ligand systems showed compromised active site residue-residue interactions in the presence of Y268C, Y268N and Y268S mutations, suggesting that the mutations also played a role in the protein-ligand instability which may ultimately affect protein function.

These findings pave the way for further studies such as *in vitro* analysis to ascertain the anti-plasmodial inhibition levels of the 8 potential inhibitors as well as 13 interface compounds of the *P. falciparum* cytochrome *b* drug target.

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

In silico studies involving various approaches were conducted with the main aim of identifying potential inhibitors against Plasmodium falciparum cytochrome b – iron sulphur protein (PfCytb-ISP). Besides the 8 potential compounds that were identified, 13 other compounds were screened and observed to have allosteric properties. This chapter will summarise the major findings in this study and outline any possible recommendations for future research considerations.

Multiple recent studies investigating atovaquone (ATQ) resistance mechanisms have shown that the *Plasmodium falciparum* cytochrome *b* (PfCytb) point mutations (Y268C, Y268N and Y268S) confer ATQ drug resistance. With this looming drug resistance, novel compounds are needed that can withstand ATQ-linked mutations. Also, computational approaches incorporating phospholipid bilayer would help design efficacious drugs, considering this approach has not been previously implemented. With these knowledge gaps and study justification, this work has identified 8 potential Qo active site hit compounds and 13 potential allosteric interface compounds from both the South African Natural Compounds Database (SANCDB) and Medicines for Malaria Venture (MMV). These compounds reported stability during MD in not only the WT but also all three mutant systems. For the first time, this work presents a selective inhibitory study performed on the PfCytb-iron sulphur protein (ISP) (PfCytb-ISP) drug target incorporated into a POPC: POPE phospholipid bilayer. From this, the study gained insights into the molecular resistance mechanisms linked to ATQ drug. Here, various *in silico* approaches such as molecular docking, molecular dynamics (MD), principal component analysis (PCA), dynamic residue network (DRN) and residue contact maps were employed.

Firstly, a rigorous analysis of the cytochrome *b* sequence and structure was performed. Generally, when various organisms were compared at sequence level, the study showed a high level of conservation within the Qo active site across most species. A comparative analysis focusing on *P. falciparum* and human homolog was performed using physiochemical properties, phylogenetics and motif calculations. The results showed distinctive features between *P. falciparum* and its human homolog at a sequence and structural level. The differences were observed in other regions away from the Qo active site. However, there is a key structural difference that was noted in the Qo active site pocket. For instance, a four-residue was reported in *Plasmodium* spp. which was reported to cause a 0.83 Å displacement in the Qo substrate binding site. Such unique region can be exploited to develop more selective drugs against the

Plasmodium parasite. This suggests that the *Plasmodium*'s Qo active site pocket experiences a change that could potentially screen for selective drug compounds, which was confirmed by this study (Chapter 5). Also, this study identified five unique motifs whose function is still unknown as they are not located in regions of known functional importance. Further studies are warranted to ascertain the function of these motifs and whether they could be explored in drug selectivity. However, at a residue level, the differences observed between *P. falciparum* and the human homolog will contribute to the identification of selective antimalarial hit compounds.

Secondly, for accurate description of the MD simulations, metal force field parameters were derived and validated through 350 ns MD simulations. Accurate force field parameters ensured that molecular systems were correctly set up *in silico* to decipher protein dynamics. In this study, Fe²⁺ metal parameters were derived for a total of three catalytic cofactors (b-type hemes and [2FE-2S] cluster). The force field parameters were validated through all-atom MD simulations of the PfCytb-ISP protein embedded in a POPC: POPE phospholipid bilayer. The validation was confirmed by monitoring metal bond distances throughout MD simulations, which were confirmed to remain within the range of expected metal bond distances. The derived force field parameters are transferable and can be implemented in other metalloproteins that share similar coordination environments to the PfCytb-ISP protein.

Thirdly, the resistance mechanism in holo and ligand-bound systems was studied through 350 ns MD simulations. A total of three mutations were investigated after previous studies associated ATQ treatment failures with Y268C, Y268N and Y268S, with the latter being the most prevalent. Various *in silico* methods were used including global analysis (root mean square deviation (RMSD), radius of gyration (Rg), centre of mass (COM)) as well as local analysis (DRN). The effect of mutations was evaluated by focusing on local residue level through DRN and weighted residue contact maps analysis, which also showed that Y268C and Y268N mutation share a similar resistance mechanism to Y268S mutation. In holo and ATQ-bound systems, DRN and

contact maps analysis showed that the three mutations disrupting protein communication by directly affecting the Qo substrate binding site. However, through weighted residue contact maps, gained residue interactions were similarly reported in the mutant systems, which could suggest a fitness cost of the parasites to available drugs resulting in parasites' ability to evade drug action mechanism due to the compensatory nature of the mutations. This warrants identification of novel compounds in replacement of the ATQ drug which has already faced resistance. Specifically, the screening of novel inhibitors that can withstand all three mutations was realised by the computational studies conducted in this work.

Fourthly, through a 350 ns all-atom MD simulation, eight selective compounds from both natural South African compounds and clinically approved (Medicines for Malaria Venture) compounds were identified. These compounds showed stability in the reference protein as well as in all three mutant systems. This *in silico* drug discovery study incorporated myriad approaches detailed previously in order to obtain reliable findings. Both global and local residue analyses including but not limited to RMSD, Rg, COM, DRN centrality metrics and weighted residue contact maps were performed in order to gain insights on the selective drug design studies. Their docking binding energies showed that these bound strongly in the *P. falciparum* cytochrome *b* Qo active site via hydrogen bonds and vdW interactions formed which sustained the compounds during MD simulations. DRN global and local analysis showed that the eight compounds were likely to inhibit the protein not only by remaining stable in its active site but also by disrupting important residue-residue interactions.

Lastly, 13 selective hit compounds were identified in the protein's heterodimer interface where their ligand dynamics were analysed through 350 ns all-atom MD simulation. The findings showed that once these selective compounds were bound at the interface, changes were induced at a distal site, specifically at the Qo active site. Using the DRN, it was noted that protein communication was affected in both reference and mutant systems. There were also a high

number of reduced/lost residue interactions in each of these mutant ligand systems. This is evidence that the mutations played a role in disrupting protein communication, thus the compounds could potentially have allosteric effects on the protein. This suggests that these compounds were likely to affect the proper functioning of the protein (ET pathway) and could be explored as potential allosteric modulators.

Wet laboratory validation of these *in silico* studies to ascertain the anti-plasmodial inhibition levels of the eight potential inhibitors of the *P. falciparum* cytochrome *b* drug target is recommended. Also, through *in vitro* and possibly *in vivo* studies, the allosteric properties of the 13 interface compounds could also be validated.

BIBLIOGRAPHY

1. World Health Organization World Malaria Report 2020: 20 Years of Global Progress and Challenges. 2020.
2. Talapko, J.; Škrlec, I.; Alebić, T.; Jukić, M.; Včev, A. Malaria: The Past and the Present. *Microorganisms* **2019**, *7*, 179, doi:10.3390/microorganisms7060179.
3. Douglas Gollin; Christian Zimmermann *Malaria: Disease Impacts and Long-Run Income Differences*; Department of Economics, Williams College, 2008;
4. Yang, T.; Otilie, S.; Istvan, E.; Godinez-Macias, K.; Lukens, A.; Baragaña, B.; Campo, B.; Walpole, C.; Niles, J.; Chibale, K.; et al. MalDA, Accelerating Malaria Drug Discovery. *Trends Parasitol.* **2021**, *37*, doi:10.1016/j.pt.2021.01.009.
5. Schumacher, R.-F.; Spinelli, E. Malaria in Children. *Mediterr. J. Hematol. Infect. Dis.* **2012**, *4*, e2012073–e2012073, doi:10.4084/MJHID.2012.073.
6. Sachs, J.; Malaney, P. The Economic and Social Burden of Malaria. *Nature* **2002**, *415*, 680–685, doi:10.1038/415680a.
7. Di Gennaro, F.; Marotta, C.; Locantore, P.; Pizzol, D.; Putoto, G. Malaria and COVID-19: Common and Different Findings. *Trop. Med. Infect. Dis.* **2020**, *5*, 141, doi:10.3390/tropicalmed5030141.
8. Nganyewo, N.; Olwal, C.; Udoakang, A.; Amenga-Etego, L.; Amambua-Ngwa, A. Pausing the Fight Against Malaria to Combat the COVID-19 Pandemic in Africa: Is the Future of Malaria Bleak? *Front. Microbiol.* **2020**, *11*, 1476, doi:10.3389/fmicb.2020.01476.
9. Wang, J.; Xu, C.; Wong, Y.; He, Y.; Adegnik, A.; Kremsner, P.; Agnandji, S.; Sall, A.; Liang, Z.; Qiu, C.; et al. Preparedness Is Essential for Malaria-Endemic Regions during the COVID-19 Pandemic. *The Lancet* **2020**, *395*, doi:10.1016/S0140-6736(20)30561-4.

10. White, N. Plasmodium Knowlesi: The Fifth Human Malaria Parasite. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **2008**, *46*, 172–173, doi:10.1086/524889.
11. Assefa, S.; Lim, C.; Preston, M.D.; Duffy, C.W.; Nair, M.B.; Adroub, S.A.; Kadir, K.A.; Goldberg, J.M.; Neafsey, D.E.; Divis, P.; et al. Population Genomic Structure and Adaptation in the Zoonotic Malaria Parasite *Plasmodium Knowlesi*. *Proc. Natl. Acad. Sci.* **2015**, *112*, 13027, doi:10.1073/pnas.1509534112.
12. Idro, R.; Marsh, K.; John, C.C.; Newton, C.R.J. Cerebral Malaria: Mechanisms of Brain Injury and Strategies for Improved Neurocognitive Outcome. *Pediatr. Res.* **2010**, *68*, 267–274, doi:10.1203/PDR.0b013e3181eee738.
13. Josling, G.A.; Russell, T.J.; Venezia, J.; Orchard, L.; van Biljon, R.; Painter, H.J.; Llinás, M. Dissecting the Role of PfAP2-G in Malaria Gametocytogenesis. *Nat. Commun.* **2020**, *11*, 1503, doi:10.1038/s41467-020-15026-0.
14. Cowell, A.N.; Winzeler, E.A. The Genomic Architecture of Antimalarial Drug Resistance. *Brief. Funct. Genomics* **2019**, *18*, 314–328, doi:10.1093/bfpg/elz008.
15. Pöschl, B.; Waneesorn, J.; Thekisoe, O.; Chutipongvivate, S.; Karanis, P.; Panagiotis, K. Comparative Diagnosis of Malaria Infections by Microscopy, Nested PCR, and LAMP in Northern Thailand. *Am. J. Trop. Med. Hyg.* **2010**, *83*, 56–60, doi:10.4269/ajtmh.2010.09-0630.
16. Zimmerman, P.A.; Howes, R.E. Malaria Diagnosis for Malaria Elimination. *Curr. Opin. Infect. Dis.* **2015**, *28*.
17. Trampuz, A.; Jereb, M.; Muzlovic, I.; Rajesh, ms Clinical Review: Severe Malaria. *Crit. Care Lond. Engl.* **2003**, *7*, 315–323, doi:10.1186/cc2183.
18. Gitta, B.; Kilian, N. Diagnosis of Malaria Parasites *Plasmodium* Spp. in Endemic Areas: Current Strategies for an Ancient Disease. *BioEssays News Rev. Mol. Cell. Dev. Biol.* **2020**, *42*, e1900138, doi:10.1002/bies.201900138.

19. Lucchi, N.W.; Demas, A.; Narayanan, J.; Sumari, D.; Kabanywany, A.; Kachur, S.P.; Barnwell, J.W.; Udhayakumar, V. Real-Time Fluorescence Loop Mediated Isothermal Amplification for the Diagnosis of Malaria. *PloS One* **2010**, *5*, e13733, doi:10.1371/journal.pone.0013733.
20. Baral, R.; Levin, A.; Odero, C.; Pecenka, C.; Tabu, C.; Mwendo, E.; Bonsu, G.; Bawa, J.; Dadzie, J.F.; Charo, J.; et al. Costs of Continuing RTS,S/AS01E Malaria Vaccination in the Three Malaria Vaccine Pilot Implementation Countries. *PloS One* **2021**, *16*, e0244995, doi:10.1371/journal.pone.0244995.
21. Gosling, R.; von Seidlein, L. The Future of the RTS,S/AS01 Malaria Vaccine: An Alternative Development Plan. *PLoS Med.* **2016**, *13*, e1001994, doi:10.1371/journal.pmed.1001994.
22. Mahmoudi, S.; Keshavarz, H. Efficacy of Phase 3 Trial of RTS, S/AS01 Malaria Vaccine: The Need for an Alternative Development Plan. *Hum. Vaccines Immunother.* **2017**, *13*, 2098–2101, doi:10.1080/21645515.2017.1295906.
23. Wangdi, K.; Furuya-Kanamori, L.; Clark, J.; Barendregt, J.; Gatton, M.; Banwell, C.; Kelly, G.; Doi, S.; Clements, A. Comparative Effectiveness of Malaria Prevention Measures: A Systematic Review and Network Meta-Analysis. *Parasit. Vectors* **2018**, *11*, 210, doi:10.1186/s13071-018-2783-y.
24. Asingizwe, D.; Poortvliet, P.M.; Koenraadt, C.J.M.; van Vliet, A.J.H.; Ingabire, C.M.; Mutesa, L.; Leeuwis, C. Role of Individual Perceptions in the Consistent Use of Malaria Preventive Measures: Mixed Methods Evidence from Rural Rwanda. *Malar. J.* **2019**, *18*, 270, doi:10.1186/s12936-019-2904-x.
25. Thomsen, E.K.; Koimbu, G.; Pulford, J.; Jamea-Maiasa, S.; Ura, Y.; Keven, J.B.; Siba, P.M.; Mueller, I.; Hetzel, M.W.; Reimer, L.J. Mosquito Behavior Change After Distribution of Bednets Results in Decreased Protection Against Malaria Exposure. *J. Infect. Dis.* **2017**, *215*, 790–797, doi:10.1093/infdis/jiw615.
26. Moiroux, N.; Gomez, M.B.; Pennetier, C.; Elanga, E.; Djènontin, A.; Chandre, F.; Djègbé, I.; Guis, H.; Corbel, V. Changes in Anopheles Funestus Biting Behavior Following Universal Coverage of

Long-Lasting Insecticidal Nets in Benin. *J. Infect. Dis.* **2012**, *206*, 1622–1629,

doi:10.1093/infdis/jis565.

27. Cooke, M.K.; Kahindi, S.C.; Oriango, R.M.; Owaga, C.; Ayoma, E.; Mabuka, D.; Nyangau, D.; Abel, L.; Atieno, E.; Awuor, S.; et al. ‘A Bite before Bed’: Exposure to Malaria Vectors Outside the Times of Net Use in the Highlands of Western Kenya. *Malar. J.* **2015**, *14*, 259, doi:10.1186/s12936-015-0766-4.
28. Lindsay, S.W.; Thomas, M.B.; Kleinschmidt, I. Threats to the Effectiveness of Insecticide-Treated Bednets for Malaria Control: Thinking beyond Insecticide Resistance. *Lancet Glob. Health* **2021**, *9*, e1325–e1331, doi:10.1016/S2214-109X(21)00216-3.
29. Vinit, R.; Timinao, L.; Bubun, N.; Katusele, M.; Robinson, L.J.; Kaman, P.; Sakur, M.; Makita, L.; Reimer, L.; Schofield, L.; et al. Decreased Bioefficacy of Long-Lasting Insecticidal Nets and the Resurgence of Malaria in Papua New Guinea. *Nat. Commun.* **2020**, *11*, 3646, doi:10.1038/s41467-020-17456-2.
30. Shibeshi, M.A.; Kifle, Z.D.; Atnafie, S.A. Antimalarial Drug Resistance and Novel Targets for Antimalarial Drug Discovery. *Infect. Drug Resist.* **2020**, *13*, 4047–4060, doi:10.2147/IDR.S279433.
31. Egwu, C.O.; Tsamesidis, I.; Pério, P.; Augereau, J.-M.; Benoit Vical, F.; Reybier, K. Superoxide: A Major Role in the Mechanism of Action of Essential Antimalarial Drugs. *Free Radic. Biol. Med.* **2021**.
32. Uzor, P.F.; Prasasty, V.D.; Agubata, C.O. Natural Products as Sources of Antimalarial Drugs. *Evid. Based Complement. Alternat. Med.* **2020**, *2020*, 9385125, doi:10.1155/2020/9385125.
33. Ginsburg, H.; Deharo, E. A Call for Using Natural Compounds in the Development of New Antimalarial Treatments – an Introduction. *Malar. J.* **2011**, *10*, S1, doi:10.1186/1475-2875-10-S1-S1.
34. Wells, T.N.C. Natural Products as Starting Points for Future Anti-Malarial Therapies: Going Back to Our Roots? *Malar. J.* **2011**, *10 Suppl 1*, S3, doi:10.1186/1475-2875-10-S1-S3.

35. Narayanan, A.; Sindhe, K.; Rai, L. Drug Resistance in Plasmodium Sp. and Novel Antimalarial Natural Products-Emerging Trends. In; 2019; pp. 95–108 ISBN 978-981-13-6919-3.
36. Ashley, E.A.; Phyo, A.P. Drugs in Development for Malaria. *Drugs* **2018**, *78*, 861–879, doi:10.1007/s40265-018-0911-9.
37. Erhirhie, E.O. Antimalarial Therapies and Infertility: A Comprehensive Review. *Toxicol. Int. Former. Indian J. Toxicol.* **2016**, *23*, 107–111.
38. Staines, H.M.; Burrow, R.; Teo, B.H.-Y.; Chis Ster, I.; Kremsner, P.G.; Krishna, S. Clinical Implications of Plasmodium Resistance to Atovaquone/Proguanil: A Systematic Review and Meta-Analysis. *J. Antimicrob. Chemother.* **2018**, *73*, 581–595, doi:10.1093/jac/dkx431.
39. Nam, T.-G.; McNamara, C.; Dharia, N.; Meister, S.; Bonamy, G.; Plouffe, D.; Kato, N.; McCormack, S.; Bursulaya, B.; Ke, H.; et al. A Chemical Genomic Analysis of Decoquinatone, a Plasmodium Falciparum Cytochrome b Inhibitor. *ACS Chem. Biol.* **2011**, *6*, 1214–1222, doi:10.1021/cb200105d.
40. Sibley, C.H.; Hyde, J.E.; Sims, P.F.; Plowe, C.V.; Kublin, J.G.; Mberu, E.K.; Cowman, A.F.; Winstanley, P.A.; Watkins, W.M.; Nzila, A.M. Pyrimethamine-Sulfadoxine Resistance in Plasmodium Falciparum: What Next? *Trends Parasitol.* **2001**, *17*, 582–588, doi:10.1016/S1471-4922(01)02085-2.
41. Rénia, L.; Goh, Y.S. Malaria Parasites: The Great Escape. *Front. Immunol.* **2016**, *7*, 463–463, doi:10.3389/fimmu.2016.00463.
42. Bopp, S.E.R.; Manary, M.J.; Bright, A.T.; Johnston, G.L.; Dharia, N.V.; Luna, F.L.; McCormack, S.; Plouffe, D.; McNamara, C.W.; Walker, J.R.; et al. Mitotic Evolution of Plasmodium Falciparum Shows a Stable Core Genome but Recombination in Antigen Families. *PLoS Genet.* **2013**, *9*, e1003293, doi:10.1371/journal.pgen.1003293.
43. Finnan, D. Artemisia: Madagascar's Coronavirus Cure or Covid-19 Quackery? 2020.
44. Principi, N.; Esposito, S. Chloroquine or Hydroxychloroquine for Prophylaxis of COVID-19. *Lancet Infect. Dis.* **2020**, *20*, doi:10.1016/S1473-3099(20)30296-6.

45. Srivastava, I.; Rottenberg, H.; Vaidya, A. Atovaquone, a Broad Spectrum Antiparasitic Drug, Collapses Mitochondrial Membrane Potential in a Malarial Parasite. *J. Biol. Chem.* **1997**, *272*, 3961–3966, doi:10.1074/jbc.272.7.3961.
46. Srivastava, I.K.; Vaidya, A.B. A Mechanism for the Synergistic Antimalarial Action of Atovaquone and Proguanil. *Antimicrob. Agents Chemother.* **1999**, *43*, 1334–1339, doi:10.1128/AAC.43.6.1334.
47. Baggish Aaron L.; Hill David R. Antiparasitic Agent Atovaquone. *Antimicrob. Agents Chemother.* **2002**, *46*, 1163–1173, doi:10.1128/AAC.46.5.1163-1173.2002.
48. Nixon, G.L.; Moss, D.M.; Shone, A.E.; Lalloo, D.G.; Fisher, N.; O'Neill, P.M.; Ward, S.A.; Biagini, G.A. Antimalarial Pharmacology and Therapeutics of Atovaquone. *J. Antimicrob. Chemother.* **2013**, *68*, 977–985.
49. Cheng, G.; Hardy, M.; Topchyan, P.; Zander, R.; Volberding, P.; Cui, W.; Kalyanaraman, B. Potent Inhibition of Tumour Cell Proliferation and Immunoregulatory Function by Mitochondria-Targeted Atovaquone. *Sci. Rep.* **2020**, *10*, 17872, doi:10.1038/s41598-020-74808-0.
50. Korsinczky, M.; Chen, N.; Kotecka, B.; Saul, A.; Rieckmann, K.; Cheng, Q. Mutations in Plasmodium Falciparum Cytochrome b That Are Associated with Atovaquone Resistance Are Located at a Putative Drug-Binding Site. *Antimicrob. Agents Chemother.* **2000**, *44*, 2100–2108.
51. Blanshard, A.; Hine, P. Atovaquone-Proguanil for Treating Uncomplicated Plasmodium Falciparum Malaria. *Cochrane Database Syst. Rev.* **2021**, *1*, CD004529, doi:10.1002/14651858.CD004529.pub3.
52. National Academies of Sciences, Engineering; Medicine *Assessment of Long-Term Health Effects of Antimalarial Drugs When Used for Prophylaxis*; Savitz, D.A., Styka, A.N., Eds.; The National Academies Press: Washington, DC, 2020; ISBN 978-0-309-67210-8.
53. Dong, C.K.; Urgaonkar, S.; Cortese, J.F.; Gamo, F.-J.; Garcia-Bustos, J.F.; Lafuente, M.J.; Patel, V.; Ross, L.; Coleman, B.I.; Derbyshire, E.R.; et al. Identification and Validation of Tetracyclic

- Benzothiazepines as Plasmodium Falciparum Cytochrome Bc1 Inhibitors. *Chem. Biol.* **2011**, *18*, 1602–1610, doi:10.1016/j.chembiol.2011.09.016.
54. Srivastava, I.K.; Morrissey, J.M.; Darrouzet, E.; Daldal, F.; Vaidya, A.B. Resistance Mutations Reveal the Atovaquone-Binding Domain of Cytochrome b in Malaria Parasites. *Mol. Microbiol.* **1999**, *33*, 704–711, doi:10.1046/j.1365-2958.1999.01515.x.
 55. Mather, M.W.; Darrouzet, E.; Valkova-Valchanova, M.; Cooley, J.W.; McIntosh, M.T.; Daldal, F.; Vaidya, A.B. Uncovering the Molecular Mode of Action of the Antimalarial Drug Atovaquone Using a Bacterial System. *J. Biol. Chem.* **2005**, *280*, 27458–27465, doi:10.1074/jbc.M502319200.
 56. Wojnarski, M.; Lon, C.; Vanachayangkul, P.; Gosi, P.; Sok, S.; Rachmat, A.; Harrison, D.; Berjohn, C.M.; Spring, M.; Chaoratanakawee, S.; et al. Atovaquone-Proguanil in Combination With Artesunate to Treat Multidrug-Resistant P. Falciparum Malaria in Cambodia: An Open-Label Randomized Trial. *Open Forum Infect. Dis.* **2019**, *6*, doi:10.1093/ofid/ofz314.
 57. Färnert, A.; Lindberg, J.; Gil, P.; Swedberg, G.; Berqvist, Y.; Thapar, M.M.; Lindegårdh, N.; Berezcky, S.; Björkman, A. Evidence of Plasmodium Falciparum Malaria Resistant to Atovaquone and Proguanil Hydrochloride: Case Reports. *BMJ* **2003**, *326*, 628–629, doi:10.1136/bmj.326.7390.628.
 58. Wichmann, O.; Muehlen, M.; Gruss, H.; Mockenhaupt, F.P.; Suttorp, N.; Jelinek, T. Malarone Treatment Failure Not Associated with Previously Described Mutations in the Cytochrome b Gene. *Malar. J.* **2004**, *3*, 14, doi:10.1186/1475-2875-3-14.
 59. di Rago, J.P.; Coppée, J.Y.; Colson, A.M. Molecular Basis for Resistance to Myxothiazol, Mucidin (Strobilurin A), and Stigmatellin. Cytochrome b Inhibitors Acting at the Center o of the Mitochondrial Ubiquinol-Cytochrome c Reductase in Saccharomyces Cerevisiae. *J. Biol. Chem.* **1989**, *264*, 14543–14548.
 60. Fivelman, Q.L.; Butcher, G.A.; Adagu, I.S.; Warhurst, D.C.; Pasvol, G. Malarone Treatment Failure and in Vitro Confirmation of Resistance of Plasmodium Falciparum Isolate from Lagos, Nigeria. *Malar. J.* **2002**, *1*, 1–1, doi:10.1186/1475-2875-1-1.

61. Schwartz, E.; Bujanover, S.; Kain, K.C. Genetic Confirmation of Atovaquone-Proguanil-Resistant *Plasmodium Falciparum* Malaria Acquired by a Nonimmune Traveler to East Africa. *Clin. Infect. Dis.* **2003**, *37*, 450–451, doi:10.1086/375599.
62. Kessl, J.J.; Ha, K.H.; Merritt, A.K.; Lange, B.B.; Hill, P.; Meunier, B.; Meshnick, S.R.; Trumppower, B.L. Cytochrome b Mutations That Modify the Ubiquinol-Binding Pocket of the Cytochrome Bc1 Complex and Confer Anti-Malarial Drug Resistance in *Saccharomyces Cerevisiae*. *J. Biol. Chem.* **2005**, *280*, 17142–17148, doi:10.1074/jbc.M500388200.
63. Musset, L.; Bouchaud, O.; Matheron, S.; Massias, L.; Bras, J.L. Clinical Atovaquone-Proguanil Resistance of *Plasmodium Falciparum* Associated with Cytochrome b Codon 268 Mutations. *Microbes Infect.* **2006**, *8*, 2599–2604, doi:10.1016/j.micinf.2006.07.011.
64. Fisher, N.; Abd Majid, R.; Antoine, T.; Al-Helal, M.; Warman, A.J.; Johnson, D.J.; Lawrenson, A.S.; Ranson, H.; O'Neill, P.M.; Ward, S.A.; et al. Cytochrome b Mutation Y268S Conferring Atovaquone Resistance Phenotype in Malaria Parasite Results in Reduced Parasite Bc1 Catalytic Turnover and Protein Expression. *J. Biol. Chem.* **2012**, *287*, 9731–9741, doi:10.1074/jbc.M111.324319.
65. Ingasia, L.A.; Akala, H.M.; Imbuga, M.O.; Opot, B.H.; Eyase, F.L.; Johnson, J.D.; Bulimo, W.D.; Kamau, E. Molecular Characterization of the Cytochrome b Gene And In Vitro Atovaquone Susceptibility of *Plasmodium Falciparum* Isolates from Kenya. *Antimicrob. Agents Chemother.* **2015**, *59*, 1818–1821, doi:10.1128/aac.03956-14.
66. *Assessment of Long-Term Health Effects of Antimalarial Drugs When Used for Prophylaxis*; Styka, A.N., Savitz, D.A., Eds.; National Academies Press (US): Washington (DC), 2020; ISBN 0-309-67210-4.
67. Sykes, J.E.; Papich, M.G. Antiprotozoal Drugs. In *Canine and Feline Infectious Diseases*; Elsevier Inc., 2013; pp. 97–104 ISBN 978-1-4377-0795-3.

68. Winter, R.; Kelly, J.; Smilkstein, M.; Dodean, R.; Hinrichs, D.; Riscoe, M. Antimalarial Quinolones: Synthesis, Potency, and Mechanistic Studies. *Exp. Parasitol.* **2008**, *118*, 487–497, doi:10.1016/j.exppara.2007.10.016.
69. Ranson, H.; Lissenden, N. Insecticide Resistance in African Anopheles Mosquitoes: A Worsening Situation That Needs Urgent Action to Maintain Malaria Control. *Spec. Issue Vectors* **2016**, *32*, 187–196, doi:10.1016/j.pt.2015.11.010.
70. Roper, C.; Pearce, R.; Nair, S.; Sharp, B.; Nosten, F.; Anderson, T. Intercontinental Spread of Pyrimethamine-Resistant Malaria. *Science* **2004**, *305*, 1124, doi:10.1126/science.1098876.
71. Amponsah, A.O.; Vosper, H.; Marfo, A.F.A. Patient Related Factors Affecting Adherence to Antimalarial Medication in an Urban Estate in Ghana. *Malar. Res. Treat.* **2015**, *2015*, 452539–452539, doi:10.1155/2015/452539.
72. Hatherley, R.; Brown, D.K.; Musyoka, T.M.; Penkler, D.L.; Faya, N.; Lobb, K.A.; Tastan Bishop, Ö. SANCDB: A South African Natural Compound Database. *J. Cheminformatics* **2015**, *7*, 29, doi:10.1186/s13321-015-0080-8.
73. Diallo, B.N.; Glenister, M.; Musyoka, T.M.; Lobb, K.; Tastan Bishop, Ö. SANCDB: An Update on South African Natural Compounds and Their Readily Available Analogs. *J. Cheminformatics* **2021**, *13*, 37, doi:10.1186/s13321-021-00514-2.
74. Horizny, C. The Drug Discovery Process. 2019.
75. Trigg, D.J.; Taylor, J.B. *Comprehensive Medicinal Chemistry II*; Elsevier Science, 2006; ISBN 978-0-08-044513-7.
76. Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E.W. Computational Methods in Drug Discovery. *Pharmacol. Rev.* **2014**, *66*, 334–395, doi:10.3762/bjoc.12.267.
77. Achan, J.; Talisuna, A.O.; Erhart, A.; Yeka, A.; Tibenderana, J.K.; Baliraine, F.N.; Rosenthal, P.J.; D'Alessandro, U. Quinine, an Old Anti-Malarial Drug in a Modern World: Role in the Treatment of Malaria. *Malar. J.* **2011**, *10*, 144, doi:10.1186/1475-2875-10-144.

78. Boateng, R.A.; Tasthan Bishop, Ö.; Musyoka, T.M. Characterisation of Plasmodial Transketolases and Identification of Potential Inhibitors: An in Silico Study. *Malar. J.* **2020**, *19*, 442, doi:10.1186/s12936-020-03512-1.
79. Khairallah, A.; Ross, C.J.; Tasthan Bishop, Ö. Probing the Structural Dynamics of the Plasmodium Falciparum Tunneling-Fold Enzyme 6-Pyruvoyl Tetrahydropterin Synthase to Reveal Allosteric Drug Targeting Sites. *Front. Mol. Biosci.* **2020**, *7*, 575196, doi:10.3389/fmolb.2020.575196.
80. Nyamai, D.; Tasthan Bishop, O. Identification of Selective Novel Hits against Plasmodium Falciparum Prolyl TRNA Synthetase Active Site and a Predicted Allosteric Site Using In Silico Approaches. *Int. J. Mol. Sci.* **2020**, *21*, 3803, doi:10.3390/ijms21113803.
81. Diallo, B.N.; Swart, T.; Hoppe, H.C.; Tasthan Bishop, Ö.; Lobb, K. Potential Repurposing of Four FDA Approved Compounds with Antiplasmodial Activity Identified through Proteome Scale Computational Drug Discovery and in Vitro Assay. *Sci. Rep.* **2021**, *11*, 1413, doi:10.1038/s41598-020-80722-2.
82. Notredame, C.; Higgins, D.G.; Heringa, J. T-Coffee: A Novel Method for Fast and Accurate Multiple Sequence Alignment. *J. Mol. Biol.* **2000**, *302*, 205–217, doi:10.1006/jmbi.2000.4042.
83. Pei, J.; Kim, B.-H.; Grishin, N.V. PROMALS3D: A Tool for Multiple Protein Sequence and Structure Alignments. *Nucleic Acids Res.* **2008**, *36*, 2295–2300, doi:10.1093/nar/gkn072.
84. Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice. *Nucleic Acids Res.* **1994**, *22*, 4673–4680, doi:10.1093/nar/22.22.4673.
85. Edgar, R.C. MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797, doi:10.1093/nar/gkh340.
86. Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066, doi:10.1093/nar/gkf436.

87. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; et al. Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539, doi:10.1038/msb.2011.75.
88. Henikoff, S.; Henikoff, J.G. Amino Acid Substitution Matrices from Protein Blocks. *Proc. Natl. Acad. Sci.* **1992**, *89*, 10915, doi:10.1073/pnas.89.22.10915.
89. Pirovano, W.; Heringa, J. Multiple Sequence Alignment. *Methods Mol. Biol. Clifton NJ* **2008**, *452*, 143–161, doi:10.1007/978-1-60327-159-2_7.
90. Katoh, K.; Kuma, K.; Toh, H.; Miyata, T. MAFFT Version 5: Improvement in Accuracy of Multiple Sequence Alignment. *Nucleic Acids Res.* **2005**, *33*, 511–518, doi:10.1093/nar/gki198.
91. Berger, M.P.; Munson, P.J. A Novel Randomized Iterative Strategy for Aligning Multiple Protein Sequences. *Comput. Appl. Biosci. CABIOS* **1991**, *7*, 479–484, doi:10.1093/bioinformatics/7.4.479.
92. Gotoh, O. Optimal Alignment between Groups of Sequences and Its Application to Multiple Sequence Alignment. *Comput. Appl. Biosci. CABIOS* **1993**, *9*, 361–370, doi:10.1093/bioinformatics/9.3.361.
93. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.A.; Clamp, M.; Barton, G.J. Jalview Version 2—a Multiple Sequence Alignment Editor and Analysis Workbench. *Bioinformatics* **2009**, *25*, 1189–1191, doi:10.1093/bioinformatics/btp033.
94. Stamatakis, A. RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313, doi:10.1093/bioinformatics/btu033.
95. Adachi, J.; Hasegawa, M. Model of Amino Acid Substitution in Proteins Encoded by Mitochondrial DNA. *J. Mol. Evol.* **1996**, *42*, 459–468, doi:10.1007/BF02498640.
96. Jones, D.T.; Taylort, W.R.; Thornton, J.M. A New Approach to Protein Fold Recognition. *Nature* **1992**, *358*, 86–89, doi:10.1038/358086a0.

97. Dimmic, M.W.; Rest, J.S.; Mindell, D.P.; Goldstein, R.A. RtREV: An Amino Acid Substitution Matrix for Inference of Retrovirus and Reverse Transcriptase Phylogeny. *J. Mol. Evol.* **2002**, *55*, 65–73, doi:10.1007/s00239-001-2304-y.
98. Le, S.Q.; Gascuel, O. An Improved General Amino Acid Replacement Matrix. *Mol. Biol. Evol.* **2008**, *25*, 1307–1320, doi:10.1093/molbev/msn067.
99. Whelan, S.; Goldman, N. A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. *Mol. Biol. Evol.* **2001**, *18*, 691–699, doi:10.1093/oxfordjournals.molbev.a003851.
100. Izquierdo-Carrasco, F.; Smith, S.A.; Stamatakis, A. Algorithms, Data Structures, and Numerics for Likelihood-Based Phylogenetic Inference of Huge Trees. *BMC Bioinformatics* **2011**, *12*, 470, doi:10.1186/1471-2105-12-470.
101. Faya, N.; Penkler, D.L.; Tastan Bishop, Ö. Human, Vector and Parasite Hsp90 Proteins: A Comparative Bioinformatics Analysis. *FEBS Open Bio* **2015**, *5*, 916–927, doi:10.1016/j.fob.2015.11.003.
102. Hillis, D.M.; Bull, J.J. An Empirical Test of Bootstrapping as a Method for Assessing Confidence in Phylogenetic Analysis. *Syst. Biol.* **1993**, *42*, 182–192, doi:10.1093/sysbio/42.2.182.
103. Huson, D.H.; Richter, D.C.; Rausch, C.; DeZulian, T.; Franz, M.; Rupp, R. Dendroscope: An Interactive Viewer for Large Phylogenetic Trees. *BMC Bioinformatics* **2007**, *8*, 460–460, doi:10.1186/1471-2105-8-460.
104. Zhang, Y. I-TASSER Server for Protein 3D Structure Prediction. *BMC Bioinformatics* **2008**, *9*, 40, doi:10.1186/1471-2105-9-40.
105. Wu, S.; Zhang, Y. LOMETS: A Local Meta-Threading-Server for Protein Structure Prediction. *Nucleic Acids Res.* **2007**, *35*, 3375–3382, doi:10.1093/nar/gkm251.
106. Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: Protein Structure and Function Prediction. *Nat. Methods* **2015**, *12*, 7–8, doi:10.1038/nmeth.3213.

107. Akhoun, B.A.; Singh, K.P.; Varshney, M.; Gupta, S.K.; Shukla, Y.; Gupta, S.K. Understanding the Mechanism of Atovaquone Drug Resistance in Plasmodium Falciparum Cytochrome b Mutation Y268S Using Computational Methods. *PloS One* **2014**, *9*, e110041.
108. Spassov, V.Z.; Bashford, D. Multiple-Site Ligand Binding to Flexible Macromolecules: Separation of Global and Local Conformational Change and an Iterative Mobile Clustering Approach. *J. Comput. Chem.* **1999**, *20*, 1091–1111, doi:10.1002/(SICI)1096-987X(199908)20:11<1091::AID-JCC1>3.0.CO;2-3.
109. de Ruyck, J.; Brysbaert, G.; Blossey, R.; Lensink, M.F. Molecular Docking as a Popular Tool in Drug Design, an in Silico Travel. *Adv. Appl. Bioinforma. Chem. AABC* **2016**, *9*, 1–11, doi:10.2147/AABC.S105289.
110. Meng, X.-Y.; Zhang, H.-X.; Mezei, M.; Cui, M. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Curr. Comput. Aided Drug Des.* **2011**, *7*, 146–157, doi:10.2174/157340911795677602.
111. Verdonk, M.L.; Cole, J.C.; Hartshorn, M.J.; Murray, C.W.; Taylor, R.D. Improved Protein-Ligand Docking Using GOLD. *Proteins* **2003**, *52*, 609–623, doi:10.1002/prot.10465.
112. Morris, G.M.; Huey, R.; Lindstrom, W.; Sanner, M.F.; Belew, R.K.; Goodsell, D.S.; Olson, A.J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791, doi:10.1002/jcc.21256.
113. Friesner, R.A.; Banks, J.L.; Murphy, R.B.; Halgren, T.A.; Klicic, J.J.; Mainz, D.T.; Repasky, M.P.; Knoll, E.H.; Shelley, M.; Perry, J.K.; et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749, doi:10.1021/jm0306430.
114. McConkey, B.J.; Sobolev, V.; Edelman, M. The Performance of Current Methods in Ligand–Protein Docking. *Curr. Sci.* **2002**, *83*, 845–856.
115. Koshland, D.E.J. Correlation of Structure and Function in Enzyme Action. *Science* **1963**, *142*, 1533–1541, doi:10.1126/science.142.3599.1533.

116. Hammes, G.G. Multiple Conformational Changes in Enzyme Catalysis. *Biochemistry* **2002**, *41*, 8221–8228, doi:10.1021/bi0260839.
117. Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267*, 727–748, doi:10.1006/jmbi.1996.0897.
118. Oshiro, C.M.; Kuntz, I.D.; Dixon, J.S. Flexible Ligand Docking Using a Genetic Algorithm. *J. Comput. Aided Mol. Des.* **1995**, *9*, 113–130, doi:10.1007/BF00124402.
119. Kitchen, D.B.; Decornez, H.; Furr, J.R.; Bajorath, J. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nat. Rev. Drug Discov.* **2004**, *3*, 935–949, doi:10.1038/nrd1549.
120. Eberhardt, J.; Santos-Martins, D.; Tillack, A.F.; Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* **2021**, *61*, 3891–3898, doi:10.1021/acs.jcim.1c00203.
121. Durrant, J.D.; McCammon, J.A. Molecular Dynamics Simulations and Drug Discovery. *BMC Biol.* **2011**, *9*, 71, doi:10.1186/1741-7007-9-71.
122. Hollingsworth, S.A.; Dror, R.O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99*, 1129–1143, doi:10.1016/j.neuron.2018.08.011.
123. Hospital, A.; Goñi, J.R.; Orozco, M.; Gelpí, J.L. Molecular Dynamics Simulations: Advances and Applications. *Adv. Appl. Bioinforma. Chem. AABC* **2015**, *8*, 37–47, doi:10.2147/AABC.S70333.
124. Karplus, M.; McCammon, J.A. Molecular Dynamics Simulations of Biomolecules. *Nat. Struct. Biol.* **2002**, *9*, 646–652, doi:10.1038/nsb0902-646.
125. McCammon, J.A.; Gelin, B.R.; Karplus, M. Dynamics of Folded Proteins. *Nature* **1977**, *267*, 585–590, doi:10.1038/267585a0.
126. Cornell, W.D.; Cieplak, P.; Bayly, C.I.; Gould, I.R.; Merz, K.M.; Ferguson, D.M.; Spellmeyer, D.C.; Fox, T.; Caldwell, J.W.; Kollman, P.A. A Second Generation Force Field for the Simulation

- of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197, doi:10.1021/ja00124a002.
127. Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174, doi:10.1002/jcc.20035.
 128. Brooks, B.R.; Brucoleri, R.E.; Olafson, B.D.; States, D.J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187–217, doi:10.1002/jcc.540040211.
 129. Christen, M.; Hünenberger, P.H.; Bakowies, D.; Baron, R.; Bürgi, R.; Geerke, D.P.; Heinz, T.N.; Kastenholz, M.A.; Kräutler, V.; Oostenbrink, C.; et al. The GROMOS Software for Biomolecular Simulation: GROMOS05. *J. Comput. Chem.* **2005**, *26*, 1719–1751, doi:10.1002/jcc.20303.
 130. Stone, J.E.; Hallock, M.J.; Phillips, J.C.; Peterson, J.R.; Luthey-Schulten, Z.; Schulten, K. Evaluation of Emerging Energy-Efficient Heterogeneous Computing Platforms for Biomolecular and Cellular Simulation Workloads. *IEEE Int. Symp. Parallel Distrib. Process. Workshop Phd Forum Proc. IEEE Int. Symp. Parallel Distrib. Process. Workshop Phd Forum* **2016**, *2016*, 89–100, doi:10.1109/IPDPSW.2016.130.
 131. Salomon-Ferrer, R.; Götz, A.W.; Poole, D.; Le Grand, S.; Walker, R.C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878–3888, doi:10.1021/ct400314y.
 132. Di Paola, L.; De Ruvo, M.; Paci, P.; Santoni, D.; Giuliani, A. Protein Contact Networks: An Emerging Paradigm in Chemistry. *Chem. Rev.* **2013**, *113*, 1598–1613, doi:10.1021/cr3002356.
 133. Süel, G.M.; Lockless, S.W.; Wall, M.A.; Ranganathan, R. Evolutionarily Conserved Networks of Residues Mediate Allosteric Communication in Proteins. *Nat. Struct. Biol.* **2003**, *10*, 59–69, doi:10.1038/nsb881.
 134. Soundararajan, V.; Raman, R.; Raguram, S.; Sasisekharan, V.; Sasisekharan, R. Atomic Interaction Networks in the Core of Protein Domains and Their Native Folds. *PloS One* **2010**, *5*, e9391, doi:10.1371/journal.pone.0009391.

135. Dokholyan Nikolay V.; Li Lewyn; Ding Feng; Shakhnovich Eugene I. Topological Determinants of Protein Folding. *Proc. Natl. Acad. Sci.* **2002**, *99*, 8637–8641, doi:10.1073/pnas.122076099.
136. Yadav, M.; Igarashi, M.; Yamamoto, N. Dynamic Residue Interaction Network Analysis of the Oseltamivir Binding Site of N1 Neuraminidase and Its H274Y Mutation Site Conferring Drug Resistance in Influenza A Virus. *PeerJ* **2021**, *9*, e11552, doi:10.7717/peerj.11552.
137. Sheik Amamuddy, O.; Afriyie Boateng, R.; Barozi, V.; Wavinya Nyamai, D.; Tastan Bishop, Ö. Novel Dynamic Residue Network Analysis Approaches to Study Allosteric Modulation: SARS-CoV-2 Mpro and Its Evolutionary Mutations as a Case Study. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 6431–6455, doi:10.1016/j.csbj.2021.11.016.
138. Okeke, C.J.; Musyoka, T.M.; Sheik Amamuddy, O.; Barozi, V.; Tastan Bishop, Ö. Allosteric Pockets and Dynamic Residue Network Hubs of Falcipain 2 in Mutations Including Those Linked to Artemisinin Resistance. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 5647–5666, doi:10.1016/j.csbj.2021.10.011.
139. Xia, D.; Esser, L.; Tang, W.-K.; Zhou, F.; Zhou, Y.; Yu, L.; Yu, C.-A. Structural Analysis of Cytochrome Bc1 Complexes: Implications to the Mechanism of Function. *Biochim. Biophys. Acta BBA-Bioenerg.* **2013**, *1827*, 1278–1294.
140. Zhu, J.; Egawa, T.; Yeh, S.-R.; Yu, L.; Yu, C.-A. Simultaneous Reduction of Iron-Sulfur Protein and Cytochrome BL during Ubiquinol Oxidation in Cytochrome Bc1 Complex. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 4864–4869, doi:10.1073/pnas.0607812104.
141. Chebon-Bore, L.; Sanyanga, T.A.; Manyumwa, C.V.; Khairallah, A.; Tastan Bishop, Ö. Decoding the Molecular Effects of Atovaquone Linked Resistant Mutations on Plasmodium Falciparum Cytb-ISP Complex in the Phospholipid Bilayer Membrane. *Int. J. Mol. Sci.* **2021**, *22*, doi:10.3390/ijms22042138.
142. Ma, H.-W.; Yang, S.; Yu, L.; Yu, C.-A. Formation of Engineered Intersubunit Disulfide Bond in Cytochrome Bc1 Complex Disrupts Electron Transfer Activity in the Complex. *Biochim. Biophys. Acta BBA-Bioenerg.* **2008**, *1777*, 317–326.

143. Merbitz-Zahradnik, T.; Zwicker, K.; Nett, J.H.; Link, T.A.; Trumpower, B.L. Elimination of the Disulfide Bridge in the Rieske Iron- Sulfur Protein Allows Assembly of the [2Fe-2S] Cluster into the Rieske Protein but Damages the Ubiquinol Oxidation Site in the Cytochrome Bc 1 Complex. *Biochemistry* **2003**, *42*, 13637–13645.
144. Song, Z.; Clain, J.; Iorga, B.I.; Vallières, C.; Lalève, A.; Fisher, N.; Meunier, B. Interplay between the Hinge Region of Iron Sulphur Protein and the Qo Site in the Bc1 Complex — Analysis of Plasmodium-like Mutations in the Yeast Enzyme. *Biochim. Biophys. Acta BBA - Bioenerg.* **2015**, *1847*, 1487–1494, doi:10.1016/j.bbabbio.2015.08.005.
145. Crofts, A.R.; Berry, E.A. Structure and Function of the Cytochrome Bc1 Complex of Mitochondria and Photosynthetic Bacteria. *Curr. Opin. Struct. Biol.* **1998**, *8*, 501–509, doi:10.1016/S0959-440X(98)80129-2.
146. Crofts, A.R.; Barquera, B.; Gennis, R.B.; Kuras, R.; Guergova-Kuras, M.; Berry, E.A. Mechanism of Ubiquinol Oxidation by the Bc1 Complex: Different Domains of the Quinol Binding Pocket and Their Role in the Mechanism and Binding of Inhibitors. *Biochemistry* **1999**, *38*, 15807–15826, doi:10.1021/bi990962m.
147. Saraste, M. Oxidative Phosphorylation at the Fin de Siècle. *Science* **1999**, *283*, 1488–1493, doi:10.1126/science.283.5407.1488.
148. Hunte, C.; Solmaz, S.; Palsdóttir, H.; Wenz, T. A Structural Perspective on Mechanism and Function of the Cytochrome Bc (1) Complex. *Results Probl. Cell Differ.* **2008**, *45*, 253–278, doi:10.1007/400_2007_042.
149. Kao, W.-C.; Hunte, C. The Molecular Evolution of the Qo Motif. *Genome Biol. Evol.* **2014**, *6*, 1894–1910, doi:10.1093/gbe/evu147.
150. Brandt, U.; Trumpower, B. The Protonmotive Q Cycle in Mitochondria and Bacteria. *Crit. Rev. Biochem. Mol. Biol.* **1994**, *29*, 165–197.
151. Trumpower, B.L. Cytochrome Bc1 Complexes of Microorganisms. *Microbiol. Rev.* **1990**, *54*, 101–129.

152. Crofts, A.R. The Cytochrome bc₁ Complex: Function in the Context of Structure. *Annu. Rev. Physiol.* **2004**, *66*, 689–733, doi:10.1146/annurev.physiol.66.032102.150251.
153. Seddigh, S.; Darabi, M. Functional, Structural, and Phylogenetic Analysis of Mitochondrial Cytochrome b (Cytb) in Insects. *Mitochondrial DNA Part DNA Mapp. Seq. Anal.* **2018**, *29*, 236–249, doi:10.1080/24701394.2016.1275596.
154. Castresana, J. Cytochrome b Phylogeny and the Taxonomy of Great Apes and Mammals. *Mol. Biol. Evol.* **2001**, *18*, 465–471, doi:10.1093/oxfordjournals.molbev.a003825.
155. UniProt Consortium The Universal Protein Resource (UniProt). *Nucleic Acids Res.* **2008**, *36*, D190–5, doi:10.1093/nar/gkm895.
156. The Plasmodium Genome Database Collaborative PlasmoDB: An Integrative Database of the Plasmodium Falciparum Genome. Tools for Accessing and Analyzing Finished and Unfinished Sequence Data. The Plasmodium Genome Database Collaborative. *Nucleic Acids Res.* **2001**, *29*, 66–69, doi:10.1093/nar/29.1.66.
157. Sayers, E.W.; Barrett, T.; Benson, D.A.; Bryant, S.H.; Canese, K.; Chetvernin, V.; Church, D.M.; DiCuccio, M.; Edgar, R.; Federhen, S.; et al. Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2009**, *37*, D5–15, doi:10.1093/nar/gkn741.
158. McGuffin, L.J.; Bryson, K.; Jones, D.T. The PSIPRED Protein Structure Prediction Server. *Bioinforma. Oxf. Engl.* **2000**, *16*, 404–405, doi:10.1093/bioinformatics/16.4.404.
159. Kyte, J.; Doolittle, R.F. A Simple Method for Displaying the Hydropathic Character of a Protein. *J. Mol. Biol.* **1982**, *157*, 105–132, doi:10.1016/0022-2836(82)90515-0.
160. Bru, C.; Courcelle, E.; Carrère, S.; Beausse, Y.; Dalmar, S.; Kahn, D. The ProDom Database of Protein Domain Families: More Emphasis on 3D. *Nucleic Acids Res.* **2005**, *33*, D212–D215, doi:10.1093/nar/gki034.
161. Korsinczky, M.; Chen, N.; Kotecka, B.; Saul, A.; Rieckmann, K.; Cheng, Q. Mutations in Plasmodium Falciparum Cytochrome b That Are Associated with Atovaquone Resistance Are Located at a Putative Drug-Binding Site. *Antimicrob. Agents Chemother.* **2000**, *44*, 2100–2108.

162. Ingasia, L.A.; Akala, H.M.; Imbuga, M.O.; Opot, B.H.; Eyase, F.L.; Johnson, J.D.; Bulimo, W.D.; Kamau, E. Molecular Characterization of the Cytochrome b Gene And In Vitro Atovaquone Susceptibility of Plasmodium Falciparum Isolates from Kenya. *Antimicrob. Agents Chemother.* **2015**, *59*, 1818–1821, doi:10.1128/aac.03956-14.
163. Musset, L.; Bouchaud, O.; Matheron, S.; Massias, L.; Bras, J.L. Clinical Atovaquone-Proguanil Resistance of Plasmodium Falciparum Associated with Cytochrome b Codon 268 Mutations. *Microbes Infect.* **2006**, *8*, 2599–2604, doi:10.1016/j.micinf.2006.07.011.
164. Brown, D.K.; Tastan Bishop, Ö. HUMA: A Platform for the Analysis of Genetic Variation in Humans. *Hum Mutat* **2018**, *39*, 40–51, doi:10.1002/humu.23334.
165. Capriotti, E.; Calabrese, R.; Casadio, R. Predicting the Insurgence of Human Genetic Diseases Associated to Single Point Protein Mutations with Support Vector Machines and Evolutionary Information. *Bioinforma. Oxf. Engl.* **2006**, *22*, 2729–2734, doi:10.1093/bioinformatics/btl423.
166. Rodrigues, C.H.; Pires, D.E.; Ascher, D.B. DynaMut: Predicting the Impact of Mutations on Protein Conformation, Flexibility and Stability. *Nucleic Acids Res.* **2018**, *46*, W350–W355, doi:10.1093/nar/gky300.
167. Capriotti, E.; Fariselli, P.; Casadio, R. I-Mutant2.0: Predicting Stability Changes upon Mutation from the Protein Sequence or Structure. *Nucleic Acids Res.* **2005**, *33*, W306–W310, doi:10.1093/nar/gki375.
168. Cheng, J.; Randall, A.; Baldi, P. Prediction of Protein Stability Changes for Single-site Mutations Using Support Vector Machines. *Proteins Struct. Funct. Bioinforma.* **2005**, *62*, 1125–1132, doi:10.1002/prot.20810.
169. Adzhubei, I.A.; Schmidt, S.; Peshkin, L.; Ramensky, V.E.; Gerasimova, A.; Bork, P.; Kondrashov, A.S.; Sunyaev, S.R. A Method and Server for Predicting Damaging Missense Mutations. *Nat. Methods* **2010**, *7*, 248–249, doi:10.1038/nmeth0410-248.

170. Choi, Y.; Sims, G.E.; Murphy, S.; Miller, J.R.; Chan, A.P. Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PloS One* **2012**, *7*, e46688, doi:10.1371/journal.pone.0046688.
171. Shihab, H.A.; Gough, J.; Cooper, D.N.; Stenson, P.D.; Barker, G.L.A.; Edwards, K.J.; Day, I.N.M.; Gaunt, T.R. Predicting the Functional, Molecular, and Phenotypic Consequences of Amino Acid Substitutions Using Hidden Markov Models. *Hum. Mutat.* **2013**, *34*, 57–65, doi:10.1002/humu.22225.
172. UniProt Consortium The Universal Protein Resource (UniProt). *Nucleic Acids Res.* **2008**, *36*, D190—5, doi:10.1093/nar/gkm895.
173. Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: A Unified Platform for Automated Protein Structure and Function Prediction. *Nat. Protoc.* **2010**, *5*, 725–738, doi:10.1038/nprot.2010.5.
174. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242, doi:10.1093/nar/28.1.235.
175. DeLano, W.L.; others Pymol: An Open-Source Molecular Graphics Tool. *CCP4 Newsl. Protein Crystallogr.* **2002**, *40*, 82–92.
176. Lovell, S.C.; Davis, I.W.; Arendall, W.B.; de Bakker, P.I.W.; Word, J.M.; Prisant, M.G.; Richardson, J.S.; Richardson, D.C. Structure Validation by Calpha Geometry: Phi,Psi and Cbeta Deviation. *Proteins* **2003**, *50*, 437—450, doi:10.1002/prot.10286.
177. Laskowski, R.; MacArthur, M.; Moss, D.; Thornton, J. PROCHECK: A Program to Check the Stereochemical Quality of Protein Structures. *J. Appl. Crystallogr.* **1993**, *26*, 283—291, doi:10.1107/s0021889892009944.
178. BIOVIA, D.S. Discovery Studio Modeling Environment, Release 2017, San Diego: DassaultSystèmes, 2016. *Available From Accessed 1 Sept. 2016* **2016**.

179. Case, D.A.; Betz, R.; Cerutti, D.; Cheatham, T.; Darden, T.; Duke, R.; Giese, T.; Gohlke, H.; Goetz, A.; Homeyer, N.; et al. AMBER 2016 Reference Manual. *Univ. Calif. San Franc.* **2016**, 1–923.
180. Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.; Li, W.W.; Noble, W.S. MEME SUITE: Tools for Motif Discovery and Searching. *Nucleic Acids Res.* **2009**, *37*, W202–208, doi:10.1093/nar/gkp335.
181. Bailey, T.L.; Gribskov, M. Methods and Statistics for Combining Motif Match Scores. *J. Comput. Biol. J. Comput. Mol. Cell Biol.* **1998**, *5*, 211–221, doi:10.1089/cmb.1998.5.211.
182. Bailey, T.L.; Gribskov, M. Combining Evidence Using P-Values: Application to Sequence Homology Searches. *Bioinforma. Oxf. Engl.* **1998**, *14*, 48–54, doi:10.1093/bioinformatics/14.1.48.
183. Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E.L.L. Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes¹¹ Edited by F. Cohen. *J. Mol. Biol.* **2001**, *305*, 567–580, doi:10.1006/jmbi.2000.4315.
184. Hulo, N.; Bairoch, A.; Bulliard, V.; Cerutti, L.; De Castro, E.; Langendijk-Genevaux, P.S.; Pagni, M.; Sigrist, C.J.A. The PROSITE Database. *Nucleic Acids Res.* **2006**, *34*, D227–D230, doi:10.1093/nar/gkj063.
185. Finn, R.D.; Bateman, A.; Clements, J.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Heger, A.; Hetherington, K.; Holm, L.; Mistry, J.; et al. Pfam: The Protein Families Database. *Nucleic Acids Res.* **2014**, *42*, D222–D230, doi:10.1093/nar/gkt1223.
186. Schultz, J.; Milpetz, F.; Bork, P.; Ponting, C.P. SMART, a Simple Modular Architecture Research Tool: Identification of Signaling Domains. *Proc. Natl. Acad. Sci.* **1998**, *95*, 5857, doi:10.1073/pnas.95.11.5857.
187. Xia Di; Yu Chang-An; Kim Hoeon; Xia Jia-Zhi; Kachurin Anatoly M.; Zhang Li; Yu Linda; Deisenhofer Johann Crystal Structure of the Cytochrome Bc₁ Complex from Bovine Heart Mitochondria. *Science* **1997**, *277*, 60–66, doi:10.1126/science.277.5322.60.

188. Widger, W.R.; Cramer, W.A.; Herrmann, R.G.; Trebst, A. Sequence Homology and Structural Similarity between Cytochrome b of Mitochondrial Complex III and the Chloroplast B6-f Complex: Position of the Cytochrome b Hemes in the Membrane. *Proc. Natl. Acad. Sci. U. S. A.* **1984**, *81*, 674–678, doi:10.1073/pnas.81.3.674.
189. Solmaz, S.R.; Hunte, C. Structure of Complex III with Bound Cytochrome c in Reduced State and Definition of a Minimal Core Interface for Electron Transfer. *J. Biol. Chem.* **2008**, *283*, 17542–17549, doi:10.1074/jbc.M710126200.
190. Iwata, S.; Lee, J.W.; Okada, K.; Lee, J.K.; Iwata, M.; Rasmussen, B.; Link, T.A.; Ramaswamy, S.; Jap, B.K. Complete Structure of the 11-Subunit Bovine Mitochondrial Cytochrome Bc1 Complex. *Science* **1998**, *281*, 64–71, doi:10.1126/science.281.5373.64.
191. Huang, L.-S.; Cobessi, D.; Tung, E.Y.; Berry, E.A. Binding of the Respiratory Chain Inhibitor Antimycin to the Mitochondrial Bc1 Complex: A New Crystal Structure Reveals an Altered Intramolecular Hydrogen-Bonding Pattern. *J. Mol. Biol.* **2005**, *351*, 573–597, doi:10.1016/j.jmb.2005.05.053.
192. Suresh, S.; Turley, S.; Oppendoes, F.R.; Michels, P.A.; Hol, W.G. A Potential Target Enzyme for Trypanocidal Drugs Revealed by the Crystal Structure of NAD-Dependent Glycerol-3-Phosphate Dehydrogenase from *Leishmania Mexicana*. *Structure* **2000**, *8*, 541–552, doi:https://doi.org/10.1016/S0969-2126(00)00135-0.
193. Söding, J.; Biegert, A.; Lupas, A.N. The HHpred Interactive Server for Protein Homology Detection and Structure Prediction. *Nucleic Acids Res.* **2005**, *33*, W244–W248, doi:10.1093/nar/gki408.
194. Hatherley, R.; Brown, D.K.; Glenister, M.; Tastan Bishop, Ö. PRIMO: An Interactive Homology Modeling Pipeline. *PloS One* **2016**, *11*, e0166698–e0166698, doi:10.1371/journal.pone.0166698.
195. Haddad, Y.; Adam, V.; Heger, Z. Ten Quick Tips for Homology Modeling of High-Resolution Protein 3D Structures. *PLOS Comput. Biol.* **2020**, *16*, e1007449, doi:10.1371/journal.pcbi.1007449.

196. Iglesias, J.; Saen-oon, S.; Soliva, R.; Guallar, V. Computational Structure-Based Drug Design: Predicting Target Flexibility. *WIREs Comput. Mol. Sci.* **2018**, *8*, e1367, doi:10.1002/wcms.1367.
197. Wu, D.; Wu, Z. Superimposition of Protein Structures with Dynamically Weighted RMSD. *J. Mol. Model.* **2010**, *16*, 211—222, doi:10.1007/s00894-009-0538-6.
198. Ramachandran, G.; Ramakrishnan, C.; Sasisekharan, V. Stereochemistry of Polypeptide Chain Configurations. *J. Mol. Biol.* **1963**, *7*, 95—99, doi:10.1016/s0022-2836(63)80023-6.
199. Bermudez, M.; Mortier, J.; Rakers, C.; Sydow, D.; Wolber, G. More than a Look into a Crystal Ball: Protein Structure Elucidation Guided by Molecular Dynamics Simulations. *Drug Discov. Today* **2016**, *21*, 1799—1805, doi:10.1016/j.drudis.2016.07.001.
200. Kim, H.J.; Khalimonchuk, O.; Smith, P.M.; Winge, D.R. Structure, Function, and Assembly of Heme Centers in Mitochondrial Respiratory Complexes. *Biochim. Biophys. Acta* **2012**, *1823*, 1604—1616, doi:10.1016/j.bbamcr.2012.04.008.
201. Kessl, J.J.; Lange, B.B.; Merbitz-Zahradnik, T.; Zwicker, K.; Hill, P.; Meunier, B.; Pálsdóttir, H.; Hunte, C.; Meshnick, S.; Trumpower, B.L. Molecular Basis for Atovaquone Binding to the Cytochrome Bc 1 Complex *. *J. Biol. Chem.* **2003**, *278*, 31312—31318, doi:10.1074/jbc.M304042200.
202. Fisher, N.; Abd Majid, R.; Antoine, T.; Al-Helal, M.; Warman, A.J.; Johnson, D.J.; Lawrenson, A.S.; Ranson, H.; O'Neill, P.M.; Ward, S.A.; et al. Cytochrome b Mutation Y268S Conferring Atovaquone Resistance Phenotype in Malaria Parasite Results in Reduced Parasite Bc1 Catalytic Turnover and Protein Expression. *J. Biol. Chem.* **2012**, *287*, 9731—9741, doi:10.1074/jbc.M111.324319.
203. Kessl, J.J.; Ha, K.H.; Merritt, A.K.; Lange, B.B.; Hill, P.; Meunier, B.; Meshnick, S.R.; Trumpower, B.L. Cytochrome b Mutations That Modify the Ubiquinol-Binding Pocket of the Cytochrome Bc1 Complex and Confer Anti-Malarial Drug Resistance in *Saccharomyces Cerevisiae*. *J. Biol. Chem.* **2005**, *280*, 17142—17148, doi:10.1074/jbc.M500388200.

204. Wichmann, O.; Muehlberger, N.; Jelinek, T.; Alifrangis, M.; Hoffmann, G.P.; hlen, M.M.; Grobusch, M.P.; Gascon, J.; Matteelli, A.; Laferl, H.; et al. Screening for Mutations Related to Atovaquone/ Proguanil Resistance in Treatment Failures and Other Imported Isolates of Plasmodium Falciparum in Europe. *J. Infect. Dis.* **2004**, *190*, 1541–1546, doi:10.1086/424469.
205. Fisher, N.; Meunier, B. Molecular Basis of Resistance to Cytochrome Bc 1 Inhibitors. *FEMS Yeast Res.* **2008**, *8*, 183–192, doi:10.1111/j.1567-1364.2007.00328.x.
206. Binkowski, T.A.; Naghibzadeh, S.; Liang, J. CASTp: Computed Atlas of Surface Topography of Proteins. *Nucleic Acids Res.* **2003**, *31*, 3352–3355, doi:10.1093/nar/gkg512.
207. Tian, W.; Chen, C.; Lei, X.; Zhao, J.; Liang, J. CASTp 3.0: Computed Atlas of Surface Topography of Proteins. *Nucleic Acids Res.* **2018**, *46*, W363–W367, doi:10.1093/nar/gky473.
208. Thompson, J.D.; Linard, B.; Lecompte, O.; Poch, O. A Comprehensive Benchmark Study of Multiple Sequence Alignment Methods: Current Challenges and Future Perspectives. *PloS One* **2011**, *6*, e18093, doi:10.1371/journal.pone.0018093.
209. Van Noorden, R.; Maher, B.; Nuzzo, R. The Top 100 Papers. *Nature* **2014**, *514*, 550–553, doi:10.1038/514550a.
210. Sodero, A.C.R.; Abraham-Vieira, B.; Torres, P.H.M.; Pascutti, P.G.; Garcia, C.R.; Ferreira, V.F.; Rocha, D.R. da; Ferreira, S.B.; Silva, F.P., Jr Insights into Cytochrome Bc1 Complex Binding Mode of Antimalarial 2-Hydroxy-1,4-Naphthoquinones through Molecular Modelling. *Mem. Inst. Oswaldo Cruz* **2017**, *112*, 299–308, doi:10.1590/0074-02760160417.
211. Schreiber, G.; Fersht, A.R. Energetics of Protein-Protein Interactions: Analysis Ofthe Barnase-Barstar Interface by Single Mutations and Double Mutant Cycles. *J. Mol. Biol.* **1995**, *248*, 478–486, doi:10.1016/S0022-2836(95)80064-6.
212. Wells, J. *Additivity of Mutational Effects in Proteins.*; Biochemistry 29:, 1990;
213. Lo, T.; Komar-Panicucci S; Sherman F, McLendon G,; Brayer GD *Structural and Functional Effects of Multiple Mutations at Distal Sites in Cytochrome c.*; 1995; Vol. Biochemistry 34;

214. Eliane Barbosa Evanovich dos Santos ED1 - Zubaida Yousaf Phylogenetics. In *Recent Advances in Phylogenetics*; IntechOpen: Rijeka, 2019; p. Ch. 1 ISBN 978-1-78985-890-7.
215. Felsenstein, J. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution* **1985**, *39*, 783–791, doi:10.2307/2408678.
216. Efron, B.; Halloran, E.; Holmes, S. Bootstrap Confidence Levels for Phylogenetic Trees. *Proc. Natl. Acad. Sci.* **1996**, *93*, 13429, doi:10.1073/pnas.93.23.13429.
217. Stamatakis, A. RAxML-VI-HPC: Maximum Likelihood-Based Phylogenetic Analyses with Thousands of Taxa and Mixed Models. *Bioinformatics* **2006**, *22*, 2688–2690, doi:10.1093/bioinformatics/btl446.
218. Skelton, F.S.; Lunan, K.D.; Folkers, K.; Schnell, J.V.; Siddiqui, W.A.; Geiman, Q.M. Biosynthesis of Ubiquinones by Malarial Parasites. I. Isolation of [¹⁴C]Ubiquinones from Cultures of Rhesus Monkey Blood Infected with Plasmodium Knowlesi. *Biochemistry* **1969**, *8*, 1284–1287, doi:10.1021/bi00831a064.
219. Schnell, J.; Siddiqui, W.; Geiman, Q. Biosynthesis of Coenzymes Q by Malarial Parasites. 2. Coenzyme Q Synthesis in Blood Cultures of Monkeys Infected with Malarial Parasites (Plasmodium Falciparum and P. Knowlesi). *J Med Chem* **1971**, *14*, 1026–1029, doi:10.1021/jm00293a002.
220. Osyczka, A.; Zhang, H.; Mathé, C.; Rich, P.R.; Moser, C.C.; Dutton, P.L. Role of the PEWY Glutamate in Hydroquinone–Quinone Oxidation–Reduction Catalysis in the Qo Site of Cytochrome Bc1. *Biochemistry* **2006**, *45*, 10492–10503, doi:10.1021/bi060013a.
221. Musyoka, T.M.; Njuguna, J.N.; Tasthan Bishop, Ö. Comparing Sequence and Structure of Falcipains and Human Homologs at Prodomain and Catalytic Active Site for Malarial Peptide Based Inhibitor Design. *Malar. J.* **2019**, *18*, 159, doi:10.1186/s12936-019-2790-2.
222. Kuzmanic, A.; Pritchard, R.B.; Hansen, D.F.; Gervasio, F.L. Importance of the Force Field Choice in Capturing Functionally Relevant Dynamics in the von Willebrand Factor. *J. Phys. Chem. Lett.* **2019**, *10*, 1928–1934, doi:10.1021/acs.jpcclett.9b00517.

223. Boldt, D.H. New Perspectives on Iron: An Introduction. *Am. J. Med. Sci.* **1999**, *318*, 207–212, doi:10.1097/00000441-199910000-00001.
224. Hu, L.; Ryde, U. Comparison of Methods to Obtain Force-Field Parameters for Metal Sites. *J. Chem. Theory Comput.* **2011**, *7*, 2452–2463, doi:10.1021/ct100725a.
225. Li, P.; Song, L.F.; Merz, K.M. Systematic Parameterization of Monovalent Ions Employing the Nonbonded Model. *J. Chem. Theory Comput.* **2015**, *11*, 1645–1657, doi:10.1021/ct500918t.
226. Stote, R.H.; Karplus, M. Zinc Binding in Proteins and Solution: A Simple but Accurate Nonbonded Representation. *Proteins Struct. Funct. Bioinforma.* **1995**, *23*, 12–31, doi:10.1002/prot.340230104.
227. Pang, Y.P.; Xu, K.; Yazal, J.E.; Prendergas, F.G. Successful Molecular Dynamics Simulation of the Zinc-Bound Farnesyltransferase Using the Cationic Dummy Atom Approach. *Protein Sci. Publ. Protein Soc.* **2000**, *9*, 1857–1865.
228. Kollman, P.A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; et al. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000**, *33*, 889–897, doi:10.1021/ar000033j.
229. Hoops, S.C.; Anderson, K.W.; Merz, K.M. Force Field Design for Metalloproteins. *J. Am. Chem. Soc.* **1991**, *113*, 8262–8270, doi:10.1021/ja00022a010.
230. Ma, H.-W.; Yang, S.; Yu, L.; Yu, C.-A. Formation of Engineered Intersubunit Disulfide Bond in Cytochrome Bc1 Complex Disrupts Electron Transfer Activity in the Complex. *Biochim. Biophys. Acta BBA - Bioenerg.* **2008**, *1777*, 317–326, doi:10.1016/j.bbabbio.2008.01.005.
231. Merbitz-Zahradnik, T.; Zwicker, K.; Nett, J.H.; Link, T.A.; Trumpower, B.L. Elimination of the Disulfide Bridge in the Rieske Iron–Sulfur Protein Allows Assembly of the [2Fe-2S] Cluster into the Rieske Protein but Damages the Ubiquinol Oxidation Site in the Cytochrome Bc1 Complex. *Biochemistry* **2003**, *42*, 13637–13645, doi:10.1021/bi035344r.

232. Stehling, O.; Lill, R. The Role of Mitochondria in Cellular Iron–Sulfur Protein Biogenesis: Mechanisms, Connected Processes, and Diseases. *Cold Spring Harb. Perspect. Biol.* **2013**, *5*, doi:10.1101/cshperspect.a011312.
233. Zhang, Z.; Huang, L.; Shulmeister, V.M.; Chi, Y.-I.; Kim, K.K.; Hung, L.-W.; Crofts, A.R.; Berry, E.A.; Kim, S.-H. Electron Transfer by Domain Movement in Cytochrome Bc1. *Nature* **1998**, *392*, 677.
234. Carvalho, A.T.P.; Swart, M. Electronic Structure Investigation and Parametrization of Biologically Relevant Iron–Sulfur Clusters. *J. Chem. Inf. Model.* **2014**, *54*, 613–620, doi:10.1021/ci400718m.
235. Shahrokh, K.; Orendt, A.; Yost, G.S.; Cheatham, T.E., 3rd Quantum Mechanically Derived AMBER-Compatible Heme Parameters for Various States of the Cytochrome P450 Catalytic Cycle. *J. Comput. Chem.* **2012**, *33*, 119–133, doi:10.1002/jcc.21922.
236. Ortiz de Montellano, P.R. Hydrocarbon Hydroxylation by Cytochrome P450 Enzymes. *Chem. Rev.* **2010**, *110*, 932–948, doi:10.1021/cr9002193.
237. Adam, S.; Knapp-Mohammady, M.; Yi, J.; Bondar, A.-N. Revised CHARMM Force Field Parameters for Iron-Containing Cofactors of Photosystem II. *J. Comput. Chem.* **2018**, *39*, 7–20, doi:10.1002/jcc.24918.
238. Iwata, S.; Saynovits, M.; Link, T.A.; Michel, H. Structure of a Water Soluble Fragment of the ‘Rieske’ Iron–Sulfur Protein of the Bovine Heart Mitochondrial Cytochrome Bc1 Complex Determined by MAD Phasing at 1.5 Å Resolution. *Structure* **1996**, *4*, 567–579, doi:10.1016/S0969-2126(96)00062-7.
239. Tendwa, M.B.; Chebon-Bore, L.; Lobb, K.; Musyoka, T.M.; Tastan Bishop, Ö. 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* **2021**, *26*, doi:10.3390/molecules26102929.

240. Myers, J.; Grothaus, G.; Narayanan, S.; Onufriev, A. A Simple Clustering Algorithm Can Be Accurate Enough for Use in Calculations of PKs in Macromolecules. *Proteins Struct. Funct. Bioinforma.* **2006**, *63*, 928–938, doi:10.1002/prot.20922.
241. Schmidtman, M.; Coster, P.; Henry, P.F.; Ting, V.P.; Weller, M.T.; Wilson, C.C. Determining Hydrogen Positions in Crystal Engineered Organic Molecular Complexes by Joint Neutron Powder and Single Crystal X-Ray Diffraction. *CrystEngComm* **2014**, *16*, 1232–1236, doi:10.1039/C3CE42070A.
242. Grimsley, G.R.; Scholtz, J.M.; Pace, C.N. A Summary of the Measured PK Values of the Ionizable Groups in Folded Proteins. *Protein Sci. Publ. Protein Soc.* **2009**, *18*, 247–251, doi:10.1002/pro.19.
243. Pace, C.N.; Grimsley, G.R.; Scholtz, J.M. Protein Ionizable Groups: PK Values and Their Contribution to Protein Stability and Solubility. *J. Biol. Chem.* **2009**, *284*, 13285–13289, doi:10.1074/jbc.R800080200.
244. Anandakrishnan, R.; Aguilar, B.; Onufriev, A. H++ 3.0: Automating PK Prediction and the Preparation of Biomolecular Structures for Atomistic Molecular Modeling and Simulations. *Nucleic Acids Res.* **2012**, *40*, W537–41, doi:10.1093/nar/gks375.
245. Gordon, J.C.; Myers, J.B.; Foltz, T.; Shoja, V.; Heath, L.S.; Onufriev, A. H++: A Server for Estimating PKas and Adding Missing Hydrogens to Macromolecules. *Nucleic Acids Res.* **2005**, *33*, W368–W371, doi:10.1093/nar/gki464.
246. Fisher, S.J.; Wilkinson, J.; Henchman, R.H.; Helliwell, J.R. An Evaluation Review of the Prediction of Protonation States in Proteins versus Crystallographic Experiment. *Crystallogr. Rev.* **2009**, *15*, 231–259, doi:10.1080/08893110903213700.
247. Schrödinger *Schrödinger Release 2019–3, Maestro*; New York, 2019;
248. Abraham, M.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1*, doi:10.1016/j.softx.2015.06.001.
249. Nielsen, A.; Holder, A. Gauss View 3.0, User's Reference. *Gaussian Inc Pittsburgh PA* **2003**.

250. Harding, M.M. Small Revisions to Predicted Distances around Metal Sites in Proteins. *Acta Crystallogr. Sect. D* **2006**, *62*, 678–682, doi:10.1107/S0907444906014594.
251. Li, P.; Merz, K.M. MCPB.Py: A Python Based Metal Center Parameter Builder. *J. Chem. Inf. Model.* **2016**, *56*, 599–604, doi:10.1021/acs.jcim.5b00674.
252. Seminario, J.M. Calculation of Intramolecular Force Fields from Second-Derivative Tensors. *Int. J. Quantum Chem.* **1996**, *60*, 1271–1277, doi:10.1002/(SICI)1097-461X(1996)60:7<1271::AID-QUA8>3.0.CO;2-W.
253. Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Petersson, G.; Nakatsuji, H.; et al. Gaussian 09, Revision A. 02, Gaussian, Inc., Wallingford, CT, 2016. *Chem. Chem. Phys. Accept. Manuscr.*
254. Riley, K.E.; Op't Holt, B.T.; Merz, K.M. Critical Assessment of the Performance of Density Functional Methods for Several Atomic and Molecular Properties. *J. Chem. Theory Comput.* **2007**, *3*, 407–433, doi:10.1021/ct600185a.
255. Oda, A.; Yamaotsu, N.; Hirono, S. New AMBER Force Field Parameters of Heme Iron for Cytochrome P450s Determined by Quantum Chemical Calculations of Simplified Models. *J. Comput. Chem.* **2005**, *26*, 818–826, doi:10.1002/jcc.20221.
256. Jakalian, A.; Bush, B.L.; Jack, D.B.; Bayly, C.I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132–146, doi:10.1002/(SICI)1096-987X(20000130)21:2<132::AID-JCC5>3.0.CO;2-P.
257. Jakalian, A.; Jack, D.B.; Bayly, C.I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.* **2002**, *23*, 1623–1641, doi:10.1002/jcc.10128.
258. Bayly, C.I.; Cieplak, P.; Cornell, W.; Kollman, P.A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269–10280, doi:10.1021/j100142a004.

259. Kang, Y.K.; Scheraga, H.A. An Efficient Method for Calculating Atomic Charges of Peptides and Proteins from Electronic Populations. *J. Phys. Chem. B* **2008**, *112*, 5470–5478, doi:10.1021/jp711484f.
260. Riley, K.E.; Op't Holt, B.T.; Merz, K.M., Jr Critical Assessment of the Performance of Density Functional Methods for Several Atomic and Molecular Properties. *J. Chem. Theory Comput.* **2007**, *3*, 407–433, doi:10.1021/ct600185a.
261. Schott-Verdugo, S.; Gohlke, H. PACKMOL-Memgen: A Simple-to-Use, Generalized Workflow for Membrane-Protein–Lipid-Bilayer System Building. *J. Chem. Inf. Model.* **2019**, *59*, 2522–2528, doi:10.1021/acs.jcim.9b00269.
262. Schott-Verdugo, S.; Gohlke, H. PACKMOL-Memgen: A Simple-to-Use, Generalized Workflow for Membrane-Protein–Lipid-Bilayer System Building. *J. Chem. Inf. Model.* **2019**, *59*, 2522–2528, doi:10.1021/acs.jcim.9b00269.
263. Schenkel, L.C.; Bakovic, M. Formation and Regulation of Mitochondrial Membranes. *Int. J. Cell Biol.* **2014**.
264. Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys* **1983**, *79*, 926–935, doi:10.1063/1.445869.
265. Lomize, M.A.; Pogozheva, I.D.; Joo, H.; Mosberg, H.I.; Lomize, A.L. OPM Database and PPM Web Server: Resources for Positioning of Proteins in Membranes. *Nucleic Acids Res.* **2012**, *40*, D370–D376, doi:10.1093/nar/gkr703.
266. Dickson, C.J.; Madej, B.D.; Skjevik, Å.A.; Betz, R.M.; Teigen, K.; Gould, I.R.; Walker, R.C. Lipid14: The Amber Lipid Force Field. *J. Chem. Theory Comput.* **2014**, *10*, 865–879, doi:10.1021/ct4010307.
267. Wan, Q.; Parks, J.M.; Hanson, B.L.; Fisher, S.Z.; Ostermann, A.; Schrader, T.E.; Graham, D.E.; Coates, L.; Langan, P.; Kovalevsky, A. Direct Determination of Protonation States and

- Visualization of Hydrogen Bonding in a Glycoside Hydrolase with Neutron Crystallography. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 12384–12389, doi:10.1073/pnas.1504986112.
268. Anandakrishnan, R.; Aguilar, B.; Onufriev, A. H++ 3.0: Automating PK Prediction and the Preparation of Biomolecular Structures for Atomistic Molecular Modeling and Simulations. *Nucleic Acids Res.* **2012**, *40*, W537–41, doi:10.1093/nar/gks375.
269. Iwata, S.; Saynovits, M.; Link, T.A.; Michel, H. Structure of a Water Soluble Fragment of the ‘Rieske’ Iron–Sulfur Protein of the Bovine Heart Mitochondrial Cytochrome Bc₁ Complex Determined by MAD Phasing at 1.5 Å Resolution. *Structure* **1996**, *4*, 567–579, doi:https://doi.org/10.1016/S0969-2126(96)00062-7.
270. Rovira, C.; Carloni, P.; Parrinello, M. The Iron–Sulfur Bond in Cytochrome c. *J. Phys. Chem. B* **1999**, *103*, 7031–7035, doi:10.1021/jp990832n.
271. Shahrokh, K.; Orendt, A.; Yost, G.S.; Cheatham, T.E., 3rd Quantum Mechanically Derived AMBER-Compatible Heme Parameters for Various States of the Cytochrome P450 Catalytic Cycle. *J. Comput. Chem.* **2012**, *33*, 119–133, doi:10.1002/jcc.21922.
272. Johansson, M.P.; Kaila, V.R.I.; Laakkonen, L. Charge Parameterization of the Metal Centers in Cytochrome c Oxidase. *J. Comput. Chem.* **2008**, *29*, 753–767, doi:10.1002/jcc.20835.
273. Autenrieth, F.; Tajkhorshid, E.; Baudry, J.; Luthey-Schulten, Z. Classical Force Field Parameters for the Heme Prosthetic Group of Cytochrome c. *J. Comput. Chem.* **2004**, *25*, 1613–1622, doi:10.1002/jcc.20079.
274. Adam, S.; Knapp-Mohammady, M.; Yi, J.; Bondar, A. Revised CHARMM Force Field Parameters for Iron-containing Cofactors of Photosystem II. *J. Comput. Chem.* **2017**, *39*, doi:10.1002/jcc.24918.
275. Purich, D.L. *Enzyme Kinetics: Catalysis and Control: A Reference of Theory and Best-Practice Methods*; Elsevier, 2010;

276. Kang, Y.K.; Scheraga, H.A. An Efficient Method for Calculating Atomic Charges of Peptides and Proteins from Electronic Populations. *J. Phys. Chem. B* **2008**, *112*, 5470–5478, doi:10.1021/jp711484f.
277. Konidala, K.K.; Bommu, U.D.; Yeguvapalli, S.; Pabbaraju, N. In Silico Insights into Prediction and Analysis of Potential Novel Pyrrolopyridine Analogs against Human MAPKAPK-2: A New SAR-Based Hierarchical Clustering Approach. *3 Biotech* **2018**, *8*, 385, doi:10.1007/s13205-018-1405-x.
278. Leelananda, S.P.; Lindert, S. Computational Methods in Drug Discovery. *Beilstein J Org Chem* **2016**, *12*, 2694–2718, doi:10.3762/bjoc.12.267.
279. Charman, S.A.; Arbe-Barnes, S.; Bathurst, I.C.; Brun, R.; Campbell, M.; Charman, W.N.; Chiu, F.C.K.; Chollet, J.; Craft, J.C.; Creek, D.J.; et al. Synthetic Ozonide Drug Candidate OZ439 Offers New Hope for a Single-Dose Cure of Uncomplicated Malaria. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 4400–4405, doi:10.1073/pnas.1015762108.
280. Nilsen, A.; LaCrue, A.N.; White, K.L.; Forquer, I.P.; Cross, R.M.; Marfurt, J.; Mather, M.W.; Delves, M.J.; Shackleford, D.M.; Saenz, F.E.; et al. Quinolone-3-Diarylethers: A New Class of Antimalarial Drug. *Sci. Transl. Med.* **2013**, *5*, 177ra37, doi:10.1126/scitranslmed.3005029.
281. Stickles, A.M.; Smilkstein, M.J.; Morrissey, J.M.; Li, Y.; Forquer, I.P.; Kelly, J.X.; Pou, S.; Winter, R.W.; Nilsen, A.; Vaidya, A.B.; et al. Atovaquone and ELQ-300 Combination Therapy as a Novel Dual-Site Cytochrome Bc1 Inhibition Strategy for Malaria. *Antimicrob. Agents Chemother.* **2016**, *60*, 4853–4859, doi:10.1128/AAC.00791-16.
282. Song, Z.; Iorga, B.I.; Mounkoro, P.; Fisher, N.; Meunier, B. The Antimalarial Compound ELQ-400 Is an Unusual Inhibitor of the Bc1 Complex, Targeting Both Qo and Qi Sites. *FEBS Lett.* **2018**, *592*, 1346–1356, doi:10.1002/1873-3468.13035.
283. Siregar, J.E.; Kurisu, G.; Kobayashi, T.; Matsuzaki, M.; Sakamoto, K.; Mi-ichi, F.; Watanabe, Y.; Hirai, M.; Matsuoka, H.; Syafruddin, D.; et al. Direct Evidence for the Atovaquone Action on the

Plasmodium Cytochrome Bc1 Complex. *SI Kazuyuki Tanabe's Gedenkschr.* **2015**, *64*, 295–300, doi:10.1016/j.parint.2014.09.011.

284. David, K.P.; Alifrangis, M.; Salanti, A.; Vestergaard, L.S.; Rønn, A.; Bygbjerg, I.B. Atovaquone/Proguanil Resistance in Africa: A Case Report. *Scand. J. Infect. Dis.* **2003**, *35*, 897–898.
285. Berry, A.; Senescau, A.; Lelièvre, J.; Benoit-Vical, F.; Fabre, R.; Marchou, B.; Magnaval, J.F. Prevalence of Plasmodium Falciparum Cytochrome b Gene Mutations in Isolates Imported from Africa, and Implications for Atovaquone Resistance. *Trans. R. Soc. Trop. Med. Hyg.* **2006**, *100*, 986–988, doi:10.1016/j.trstmh.2006.02.004.
286. Sutherland, C.J.; Laundy, M.; Price, N.; Burke, M.; Fivelman, Q.L.; Pasvol, G.; Klein, J.L.; Chiodini, P.L. Mutations in the Plasmodium Falciparum Cytochrome b Gene Are Associated with Delayed Parasite Recrudescence in Malaria Patients Treated with Atovaquone-Proguanil. *Malar. J.* **2008**, *7*, 240, doi:10.1186/1475-2875-7-240.
287. Srivastava, I.K.; Morrissey, J.M.; Darrouzet, E.; Daldal, F.; Vaidya, A.B. Resistance Mutations Reveal the Atovaquone-Binding Domain of Cytochrome b in Malaria Parasites. *Mol. Microbiol.* **1999**, *33*, 704–711, doi:10.1046/j.1365-2958.1999.01515.x.
288. Kessl, J.J.; Meshnick, S.R.; Trumpower, B.L. Modeling the Molecular Basis of Atovaquone Resistance in Parasites and Pathogenic Fungi. *Trends Parasitol.* **2007**, *23*, 494–501, doi:10.1016/j.pt.2007.08.004.
289. Fisher, N.; Abd Majid, R.; Antoine, T.; Al-Helal, M.; Warman, A.J.; Johnson, D.J.; Lawrenson, A.S.; Ranson, H.; O'Neill, P.M.; Ward, S.A.; et al. Cytochrome b Mutation Y268S Conferring Atovaquone Resistance Phenotype in Malaria Parasite Results in Reduced Parasite Bc1 Catalytic Turnover and Protein Expression. *J. Biol. Chem.* **2012**, *287*, 9731–9741, doi:10.1074/jbc.M111.324319.

290. Sommer, B. Membrane Packing Problems: A Short Review on Computational Membrane Modeling Methods and Tools. *Comput. Struct. Biotechnol. J.* **2013**, *5*, e201302014, doi:10.5936/csbj.201302014.
291. Javanainen, M.; Martinez-Seara, H. Efficient Preparation and Analysis of Membrane and Membrane Protein Systems. *Biochim. Biophys. Acta BBA-Biomembr.* **2016**, *1858*, 2468–2482, doi:10.1016/j.bbamem.2016.02.036.
292. Goossens, K.; De Winter, H. Molecular Dynamics Simulations of Membrane Proteins: An Overview. *J. Chem. Inf. Model.* **2018**, *58*, 2193–2202, doi:10.1021/acs.jcim.8b00639.
293. Lindahl, E.; Sansom, M.S. Membrane Proteins: Molecular Dynamics Simulations. *Curr. Opin. Struct. Biol.* **2008**, *18*, 425–431, doi:10.1016/j.sbi.2008.02.003.
294. Brown, D.K.; Tastan Bishop, Ö. Role of Structural Bioinformatics in Drug Discovery by Computational SNP Analysis: Analyzing Variation at the Protein Level. *Glob. Heart* **2017**, *12*, 151–161, doi:10.1016/j.gheart.2017.01.009.
295. Brown, D.K.; Sheik Amamuddy, O.; Tastan Bishop, Ö. Structure-Based Analysis of Single Nucleotide Variants in the Renin-Angiotensinogen Complex. *Glob. Heart* **2017**, *12*, 121–132, doi:10.1016/j.gheart.2017.01.006.
296. Kozakov, D.; Grove, L.E.; Hall, D.R.; Bohnuud, T.; Mottarella, S.E.; Luo, L.; Xia, B.; Beglov, D.; Vajda, S. The FTMap Family of Web Servers for Determining and Characterizing Ligand-Binding Hot Spots of Proteins. *Nat. Protoc.* **2015**, *10*, 733–755, doi:10.1038/nprot.2015.043.
297. Ngan, C.-H.; Hall, D.R.; Zerbe, B.; Grove, L.E.; Kozakov, D.; Vajda, S. FTSite: High Accuracy Detection of Ligand Binding Sites on Unbound Protein Structures. *Bioinforma. Oxf. Engl.* **2012**, *28*, 286–287, doi:10.1093/bioinformatics/btr651.
298. Huang, W.; Lu, S.; Huang, Z.; Liu, X.; Mou, L.; Luo, Y.; Zhao, Y.; Liu, Y.; Chen, Z.; Hou, T.; et al. Allosite: A Method for Predicting Allosteric Sites. *Bioinformatics* **2013**, *29*, 2357–2359, doi:10.1093/bioinformatics/btt399.

299. Panjkovich, A.; Daura, X. PARS: A Web Server for the Prediction of Protein Allosteric and Regulatory Sites. *Bioinformatics* **2014**, *30*, 1314–1315, doi:10.1093/bioinformatics/btu002.
300. Huang, B.; Schroeder, M. LIGSITEcsc: Predicting Ligand Binding Sites Using the Connolly Surface and Degree of Conservation. *BMC Struct. Biol.* **2006**, *6*, 19, doi:10.1186/1472-6807-6-19.
301. Kim, S.; Thiessen, P.A.; Bolton, E.E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B.A.; et al. PubChem Substance and Compound Databases. *Nucleic Acids Res.* **2016**, *44*, D1202–D1213, doi:10.1093/nar/gkv951.
302. Hentschel, C.C. The Medicines for Malaria Venture. *Nature* **2002**, *415*, 715–715, doi:10.1038/415715a.
303. Trott, O.; Olson, A.J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461, doi:10.1002/jcc.21334.
304. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–8, 27–8, doi:10.1016/0263-7855(96)00018-5.
305. Roe, D.R.; Cheatham, T.E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095, doi:10.1021/ct400341p.
306. Varadwaj, P.R.; Varadwaj, A.; Marques, H.M.; Yamashita, K. Significance of Hydrogen Bonding and Other Noncovalent Interactions in Determining Octahedral Tilting in the CH₃NH₃PbI₃ Hybrid Organic-Inorganic Halide Perovskite Solar Cell Semiconductor. *Sci. Rep.* **2019**, *9*, 50, doi:10.1038/s41598-018-36218-1.
307. Lee, H.R.; Helquist, S.A.; Kool, E.T.; Johnson, K.A. Importance of Hydrogen Bonding for Efficiency and Specificity of the Human Mitochondrial DNA Polymerase. *J. Biol. Chem.* **2008**, *283*, 14402–14410, doi:10.1074/jbc.M705007200.

308. Sheik Amamuddy, O.; Glenister, M.; Tastan Bishop, O. MDM-TASK-Web: A Web Platform for Protein Dynamic Residue Networks and Modal Analysis. *bioRxiv* **2021**, 2021.01.29.428734, doi:10.1101/2021.01.29.428734.
309. Jolliffe, I.T.; Cadima, J. Principal Component Analysis: A Review and Recent Developments. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **2016**, *374*, 20150202, doi:10.1098/rsta.2015.0202.
310. Sanyanga, T.A.; Nizami, B.; Tastan Bishop, Ö. Mechanism of Action of Non-Synonymous Single Nucleotide Variations Associated with α -Carbonic Anhydrase II Deficiency. *Molecules* **2019**, *24*, doi:10.3390/molecules24213987.
311. Kao, W.-C.; Hunte, C. The Molecular Evolution of the Qo Motif. *Genome Biol. Evol.* **2014**, *6*, 1894–1910, doi:10.1093/gbe/evu147.
312. Song, Z.; Clain, J.; Iorga, B.I.; Yi, Z.; Fisher, N.; Meunier, B. Saccharomyces Cerevisiae-Based Mutational Analysis of the BcI Complex Qo Site Residue 279 to Study the Trade-off between Atovaquone Resistance and Function. *Antimicrob. Agents Chemother.* **2015**, *59*, 4053–4058, doi:10.1128/AAC.00710-15.
313. Carlson, H.A.; McCammon, J.A. Accommodating Protein Flexibility in Computational Drug Design. *Mol. Pharmacol.* **2000**, *57*, 213–218.
314. Wang, Y.; Zhao, T.; Wei, D.; Strandberg, E.; Ulrich, A.S.; Ulmschneider, J.P. How Reliable Are Molecular Dynamics Simulations of Membrane Active Antimicrobial Peptides? *Biochim. Biophys. Acta* **2014**, *1838*, 2280–2288, doi:10.1016/j.bbamem.2014.04.009.
315. Akhoun, B.A.; Singh, K.P.; Varshney, M.; Gupta, S.K.; Shukla, Y.; Gupta, S.K. Understanding the Mechanism of Atovaquone Drug Resistance in Plasmodium Falciparum Cytochrome b Mutation Y268S Using Computational Methods. *PloS One* **2014**, *9*, e110041, doi:10.1371/journal.pone.0110041.
316. Srivastava, A.; Bala, S.; Motomura, H.; Kohda, D.; Tama, F.; Miyashita, O. Conformational Ensemble of an Intrinsically Flexible Loop in Mitochondrial Import Protein Tim21 Studied by

- Modeling and Molecular Dynamics Simulations. *Biochim. Biophys. Acta Gen. Subj.* **2020**, *1864*, 129417, doi:10.1016/j.bbagen.2019.129417.
317. Brown, D.K.; Penkler, D.L.; Sheik Amamuddy, O.; Ross, C.; Atilgan, A.R.; Atilgan, C.; Tastan Bishop, Ö. MD-TASK: A Software Suite for Analyzing Molecular Dynamics Trajectories. *Bioinformatics* **2017**, *33*, 2768–2771, doi:10.1093/bioinformatics/btx349.
318. Birth, D.; Kao, W.-C.; Hunte, C. Structural Analysis of Atovaquone-Inhibited Cytochrome Bc1 Complex Reveals the Molecular Basis of Antimalarial Drug Action. *Nat. Commun.* **2014**, *5*, 4029, doi:10.1038/ncomms5029.
319. Palsdottir, H.; Hunte, C. Lipids in Membrane Protein Structures. *Biochim. Biophys. Acta* **2004**, *1666*, 2–18, doi:10.1016/j.bbamem.2004.06.012.
320. Wichmann, O.; Muehlen, M.; Gruss, H.; Mockenhaupt, F.P.; Suttorp, N.; Jelinek, T. Malarone Treatment Failure Not Associated with Previously Described Mutations in the Cytochrome b Gene. *Malar. J.* **2004**, *3*, 14, doi:10.1186/1475-2875-3-14.
321. Amusengeri, A.; Tata, R.B.; Tastan Bishop, Ö. Understanding the Pyrimethamine Drug Resistance Mechanism via Combined Molecular Dynamics and Dynamic Residue Network Analysis. *Molecules* **2020**, *25*, 904, doi:10.3390/molecules25040904.
322. Sanyanga, T.A.; Tastan Bishop, Ö. Structural Characterization of Carbonic Anhydrase VIII and Effects of Missense Single Nucleotide Variations to Protein Structure and Function. *Int. J. Mol. Sci.* **2020**, *21*, 2764, doi:10.3390/ijms21082764.
323. Kao, W.-C.; Hunte, C. The Molecular Evolution of the Qo Motif. *Genome Biol. Evol.* **2014**, *6*, 1894–1910, doi:10.1093/gbe/evu147.
324. Gennis, R.; Barquera, B.; Hacker, B.; Van Doren, S.; Arnaud, S.; Crofts, A.; Davidson, E.; Gray, K.; Daldal, F. The Bc1 Complexes of Rhodobacter Sphaeroides and Rhodobacter Capsulatus. *J. Bioenerg. Biomembr.* **1993**, *25*, 195—209, doi:10.1007/bf00762582.

325. Penkler, D.L.; Atilgan, C.; Tastan Bishop, Ö. Allosteric Modulation of Human Hsp90 α Conformational Dynamics. *J. Chem. Inf. Model.* **2018**, *58*, 383–404, doi:10.1021/acs.jcim.7b00630.
326. Maier, R.M.; Pepper, I.L.; Gerba, C.P. 1945- *Environmental Microbiology*; Academic Press: San Diego, Calif., 2000; ISBN 0-12-497570-4.
327. Barratt, E.; Bingham, R.J.; Warner, D.J.; Laughton, C.A.; Phillips, S.E.V.; Homans, S.W. Van Der Waals Interactions Dominate Ligand-Protein Association in a Protein Binding Site Occluded from Solvent Water. *J. Am. Chem. Soc.* **2005**, *127*, 11827–11834, doi:10.1021/ja0527525.
328. Tougan, T.; Toya, Y.; Uchihashi, K.; Horii, T. Application of the Automated Haematology Analyzer XN-30 for Discovery and Development of Anti-Malarial Drugs. *Malar. J.* **2019**, *18*, 8, doi:10.1186/s12936-019-2642-0.
329. van Zyl, R.L.; Khan, F.; Edwards, T.J.; Drewes, S.E. Antiplasmodial Activities of Some Abietane Diterpenes from the Leaves of Five Plectranthus Species. *South Afr. J. Sci.* **2008**, *104*, 62–64.
330. van Heerden, F.R. Hoodia Gordonii: A Natural Appetite Suppressant. *J. Ethnopharmacol.* **2008**, *119*, 434–437, doi:10.1016/j.jep.2008.08.023.
331. Fu, X.; Li, X.-C.; Smillie, T.J.; Carvalho, P.; Mabusela, W.; Syce, J.; Johnson, Q.; Folk, W.; Avery, M.A.; Khan, I.A. Cycloartane Glycosides from Sutherlandia Frutescens. *J. Nat. Prod.* **2008**, *71*, 1749–1753, doi:10.1021/np800328r.
332. McPhail, K.L.; Davies-Coleman, M.T.; Starmer, J. Sequestered Chemistry of the Arminacean Nudibranch Leminda Millecra in Algoa Bay, South Africa. *J. Nat. Prod.* **2001**, *64*, 1183–1190, doi:10.1021/np010085x.
333. Musyoka, T.; Bishop, Ö.T. South African Abietane Diterpenoids and Their Analogs as Potential Antimalarials: Novel Insights from Hybrid Computational Approaches. *Molecules* **2019**, *24*, doi:10.3390/molecules24224036.
334. Panigrahi, S.K. Strong and Weak Hydrogen Bonds in Protein-Ligand Complexes of Kinases: A Comparative Study. *Amino Acids* **2008**, *34*, 617–633, doi:10.1007/s00726-007-0015-4.

335. Patil, R.; Das, S.; Stanley, A.; Yadav, L.; Sudhakar, A.; Varma, A.K. Optimized Hydrophobic Interactions and Hydrogen Bonding at the Target-Ligand Interface Leads the Pathways of Drug-Designing. *PloS One* **2010**, *5*, e12029–e12029, doi:10.1371/journal.pone.0012029.
336. Zhang, Z.; Huang, L.; Shulmeister, V.M.; Chi, Y.-I.; Kim, K.K.; Hung, L.-W.; Crofts, A.R.; Berry, E.A.; Kim, S.-H. Electron Transfer by Domain Movement in Cytochrome Bc 1. *Nature* **1998**, *392*, 677–684.
337. Carvalho, A.T.P.; Swart, M. Electronic Structure Investigation and Parametrization of Biologically Relevant Iron–Sulfur Clusters. *J. Chem. Inf. Model.* **2014**, *54*, 613–620, doi:10.1021/ci400718m.
338. Cramer, W.A.; Hasan, S.S.; Yamashita, E. The Q Cycle of Cytochrome Bc Complexes: A Structure Perspective. *Biochim. Biophys. Acta BBA - Bioenerg.* **2011**, *1807*, 788–802, doi:10.1016/j.bbabbio.2011.02.006.
339. Hunte, C.; Koepke, J.; Lange, C.; Roßmanith, T.; Michel, H. Structure at 2.3 Å Resolution of the Cytochrome Bc1 Complex from the Yeast *Saccharomyces Cerevisiae* Co-Crystallized with an Antibody Fv Fragment. *Structure* **2000**, *8*, 669–684, doi:10.1016/S0969-2126(00)00152-0.
340. Lancaster, C.R.D.; Hunte, C.; Kelley, J.; Trumpower, B.L.; Ditchfield, R. A Comparison of Stigmatellin Conformations, Free and Bound to the Photosynthetic Reaction Center and the Cytochrome Bc1 Complex. *J. Mol. Biol.* **2007**, *368*, 197–208, doi:10.1016/j.jmb.2007.02.013.
341. Planas-Iglesias, J.; Bonet, J.; García-García, J.; Marín-López, M.A.; Feliu, E.; Oliva, B. Understanding Protein-Protein Interactions Using Local Structural Features. *J. Mol. Biol.* **2013**, *425*, 1210–1224, doi:10.1016/j.jmb.2013.01.014.
342. Elangovan, A.; Davis, M.; Verspoor, K. Assigning Function to Protein-Protein Interactions: A Weakly Supervised BioBERT Based Approach Using PubMed Abstracts 2020.
343. Amusengeri, A.; Astl, L.; Lobb, K.; Verkhivker, G.M.; Tastan Bishop, Ö. Establishing Computational Approaches Towards Identifying Malarial Allosteric Modulators: A Case Study of *Plasmodium Falciparum* Hsp70s. *Int. J. Mol. Sci.* **2019**, *20*, doi:10.3390/ijms20225574.

344. Amusengeri, A.; Tastan Bishop, Ö. Discorhabdin N, a South African Natural Compound, for Hsp72 and Hsc70 Allosteric Modulation: Combined Study of Molecular Modeling and Dynamic Residue Network Analysis. *Mol. Basel Switz.* **2019**, *24*, doi:10.3390/molecules24010188.
345. Penkler, D.L.; Tastan Bishop, Ö. Modulation of Human Hsp90 α Conformational Dynamics by Allosteric Ligand Interaction at the C-Terminal Domain. *Sci. Rep.* **2019**, *9*, 1600, doi:10.1038/s41598-018-35835-0.
346. Wenthur, C.J.; Gentry, P.R.; Mathews, T.P.; Lindsley, C.W. Drugs for Allosteric Sites on Receptors. *Annu. Rev. Pharmacol. Toxicol.* **2014**, *54*, 165–184, doi:10.1146/annurev-pharmtox-010611-134525.
347. Ayyildiz, M.; Celiker, S.; Ozhelvaci, F.; Akten, E.D. Identification of Alternative Allosteric Sites in Glycolytic Enzymes for Potential Use as Species-Specific Drug Targets. *Front. Mol. Biosci.* **2020**, *7*, 88, doi:10.3389/fmolb.2020.00088.
348. Greener, J.G.; Sternberg, M.J.E. AlloPred: Prediction of Allosteric Pockets on Proteins Using Normal Mode Perturbation Analysis. *BMC Bioinformatics* **2015**, *16*, 335, doi:10.1186/s12859-015-0771-1.
349. Weinkam, P.; Pons, J.; Sali, A. Structure-Based Model of Allostery Predicts Coupling between Distant Sites. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 4875–4880, doi:10.1073/pnas.1116274109.
350. Ferreira de Freitas, R.; Schapira, M. A Systematic Analysis of Atomic Protein-Ligand Interactions in the PDB. *MedChemComm* **2017**, *8*, 1970–1981, doi:10.1039/c7md00381a.
351. Bronowska K, A. *Thermodynamics of Ligand-Protein Interactions: Implications for Molecular Design*; 2011; Vol. Thermodynamics-Interaction Studies-Solids, Liquids and Gases;

SUPPLEMENTARY DATA

Table S3.1: Retrieved cytochrome *b* sequences from a total of 72 organisms including *Plasmodium* and human homolog. # names used for convenience.

		Query coverage	Short name	E value	% Identity	Sequence length	Accession no.
1	P.falciparum 3D7(PlasmoDB, Query)					376	mal_mito_3
2	Plasmodium reichenowi	100%	Pre_MT-CYB	0	99%	376	NP_057946.1
3	Plasmodium gaboni	100%	Pgab_MT-CYB	0	98%	376	ACR56350.2
4	Plasmodium vivax	100%	Pvi_MT-CYB	0	91%	382	AGS77576.1
5	Plasmodium knowlesi	100%	Pkn_MT-CYB	0	91%	376	ACJ65364.1
6	Plasmodium ovale	100%	Pov_MT-CYB	0	91%	376	BAG68535.1
7	Plasmodium fragile	100%	Pfr_MT-CYB	0	91%	376	BAI82347.1
8	Plasmodium cynomolgi	100%	Pcy_MT-CYB	0	91%	376	AND46784.1
9	Plasmodium inui	100%	Pin_MT-CYB	0	91%	376	BAI82272.1
10	Plasmodium malariae	100%	Pma_MT-CYB	0	90%	380	AVN53206.1
11	Plasmodium chabaudi	91%	Pch_MT-CYB	0	90%	345	AAM47792.1
12	Plasmodium vinckei	100%	Pvin_MT-CYB	0	90%	376	BAK18851.1
13	Plasmodium berghei	100%	Pbe_MT-CYB	0	89%	376	AAD37977.1
14	Plasmodium yoeli	91%	Pyo_MT-CYB	0	89%	345	AAM47793.1
15	Isospora manorinae	93%	Is.man_MT-CYB	1,00E-148	80%	359	YP_009350000.1
16	Haemoproteus columbae	93%	Hae.col_MT-CYB	0	79%	375	YP_002734944.1
17	Leucocytozoon fringillinarum	100%	Le.fri_MT-CYB	0,00E+00	75%	376	YP_002734963.1
18	Leucocytozoon majoris	100%	Le.maj_MT-CYB	0	75%	376	YP_002734960.1
19	Eimeria praecox	93%	Ei.pra_MT-CYB	5,00E-145	59%	359	AOX48549.1

20	Sarcocystis speeri	65%	Sa.spe_MT-CYB	4,00E-105	59%	247	AKT76131.1
21	Eimeria acervulina	93%	Ei.ace_MT-CYB	2,00E-143	58%	359	AOX48558.1
22	Eimeria falciformis	93%	Ei.fal_MT-CYB	1,00E-145	58%	359	APU51309.1
23	Eimeria tenella	93%	Ei.ten_MT-CYB	3,00E-144	58%	359	ADW23647.1
24	Neospora caninum	93%	N.can_MT-CYB	4,00E-129	54%	368	QLF97291.1
25	Toxoplasma gondii	93%	Tg_MT-CYB	2E-127	54%	368	AOT06086.1
26	Cytauxzoon felis	79%	C.fel_MT-CYB	2,00E-88	49%	363	AGF95349.1
27	Babesia rodhaini	84%	Ba.rod_MT-CYB	1E-82	46%	384	BAM68229.1
28	Theileria equi	79%	Th.equ_MT-CYB	2,00E-75	46%	363	BAI66182.1
29	Rickettsia typhi	93%	R.typ_MT-CYB	4E-87	45%	398	WP_011190728.1
30	Rickettsia prowazekii	93%	R.pro_MT-CYB	6,00E-91	45%	398	WP_004599371.1
31	Babesia ovata	79%	Ba.ova_MT-CYB	2,00E-81	45%	363	BAV78785.1
32	Gracilaria edulis	94%	Gr.edu_MT-CYB	8,00E-93	44%	380	YP_009490373.1
33	Sitophilus oryzae	90%	Si.ory_MT-CYB	1,00E-89	44%	379	YP_009271152.1
34	Babesia bovis	75%	Ba.bov_MT-CYB	4,00E-83	44%	363	ACS68793.1
35	Theileria annulata	73%	Th.ann_MT-CYB	9,00E-71	44%	363	AVV68815.1
36	Homo sapiens	94%	H.sap_MT-CYB	2E-79	43%	378	ABR98433.1
37	Saccharomyces cerevisiae	80%	S.cer_MT-CYB	1E-73	43%	385	YP_009144709.1
38	Babesia gibsoni	95%	Ba.gib_MT-CYB	3E-82	43%	363	AKL78760.1
39	Aedes aegypti	92%	Ae.aeg_MT-CYB	2E-82	43%	378	YP_009389271.1
40	Anaplasma centrale	91%	Ana.cen_MT-CYB	2E-92	43%	408	WP_012880629.1
41	Anaplasma marginale	91%	Ana.mar_MT-CYB	2E-92	43%	406	WP_037352393.1
42	Vermamoeba vermiformis	90%	V.ver_MT-CYB	1E-87	43%	388	AKT93955.1
43	Theileria parva	79%	Th.par_MT-CYB	2E-57	43%	386	YP_001994287.1
44	Cyanidioschyzon merolae	92%	Cy.mer_MTCYB	3E-94	43%	382	NP_059352.1
45	Gracilaria changii	94%	Gr.cha_MT-CYB	4E-93	43%	380	YP_009365195.1
46	Agarophyton chilense	94%	Ag.chi_MT-CYB	5E-93	43%	380	YP_009130767.1

47	Agarophyton tenuistipitatum	94%	Ag.ten_MT-CYB	8E-93	43%	379	YP_009490423.1
48	Psammoneis japonica	97%	Ps.ja_MT-CYB	4E-95	42%	415	YP_009495538.1
49	Triparma laevis	94%	Tri.lae_MT-CYB	5E-96	42%	387	YP_009163731.1
50	Gracilaria spinulosa	92%	Gr.spi_MT-CYB	1E-91	42%	380	YP_009490398.1
51	Tetrahymena thermophila	92%	Te.the_MT-CYB	6E-81	42%	380	ABU68799.1
52	Gracilaria salicornia	92%	Gr.sal_MT-CYB	1E-91	42%	379	YP_009019504.1
53	Zeylanicobdella arugamensis	92%	Z.aru_MT-CYB	4E-96	42%	374	YP_009400210.1
54	Babesia orientalis	88%	Ba.ori_MT-CYB	3E-81	42%	363	AHB82181.1
55	Anopheles sinensis	92%	An.sin_MT-CYB	3E-73	42%	378	YP_009169982.1
56	Acanthamoeba castellanii	92%	Ac.cas_MT-CYB	2E-84	42%	385	AKT93989.1
57	Anopheles melas	92%	An.mel_MT-CYB	1E-72	41%	378	YP_009176726.1
58	Chattonella marina	93%	Ch.mar_MT-CYB	4E-95	41%	386	YP_003434220.1
59	Acanthamoeba polyphaga	93%	Ac.pol_MT-CYB	8E-85	41%	385	ALG41020.1
60	Babesia microti	83%	Ba.mic_MT-CYB	1E-76	41%	375	YP_009305217.1
61	Theileria orientalis	79%	Th.ori_MT-CYB	1E-75	41%	363	BAI66179.1
62	Anopheles gambiae	92%	An.gam_MT-CYB	4E-73	41%	378	AVM85220.1
63	Anopheles atroparvus	92%	An.atr_MT-CYB	1E-73	41%	378	YP_009176648.1
64	Anopheles darlingi	92%	An.dar_MT-CYB	2E-73	41%	378	YP_003734752.1
65	Anopheles farauti	92%	An.far_MT-CYB	7E-73	41%	378	YP_007626977.1
66	Anopheles maculatus	92%	An.mac_MT-CYB	3E-72	41%	378	YP_009176713.1
67	Anopheles minimus	92%	An.min_MT-CYB	4E-73	41%	378	YP_009176752.1
68	Anopheles stephensi	92%	An.ste_MT-CYB	7E-73	41%	378	ANE11051.1
69	Schistosoma hematobium	75%	Sc.hae_MT-CYB	1E-46	38%	367	YP_626521.1
70	Schistosoma curassoni	80%	Sc.cur_MT-CYB	8E-50	36%	367	BAV81456.1
71	Schistosoma mansoni	77%	Sc.man_MT-CYB	5E-45	36%	364	NP_066212.2
72	Schistosoma japonicum	81%	Sc.jap_MT-CYB	6E-47	35%	370	AFX73134.1

73	Trypanosoma rangeli	67%	Tr.ran_MT-CYB	5E-25	33%	368	ESL04761.1
74	Trypanosoma vivax	66%	Tr.viv_MT-CYB	8,00E-22	29%	370	AJO53308.1
		Query coverage	Short name	E value	% Identity	Sequence length	Accession no.
1	Toxoplasma gondii	93%	T.gondii	2,00E-127	54%	368	AOT06086.1
2	Gallus gallus	92%	G.gallus	2,00E-84	43%	380	AIE44496.1
3	Bos taurus	92%	B.taurus	4,00E-82	44%	379	ALM02696.1

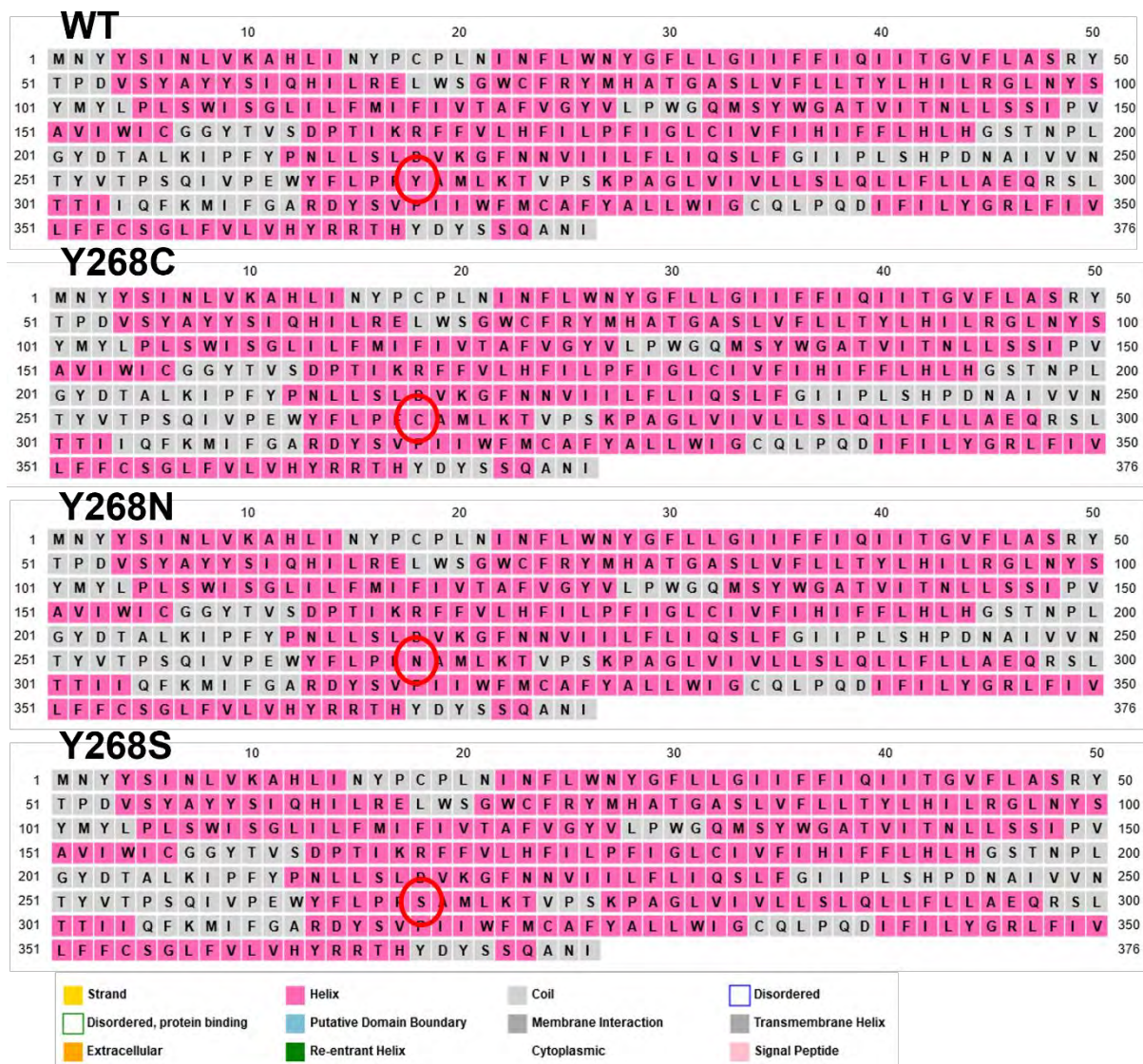


Figure S3.1: Secondary structure prediction of the three mutations (Y268C/N/S) in comparison to the wild-type. The red circle highlights the mutation location while the key indicates the type of secondary structures.

[illegible]

Figure S3.2: Multiple sequence alignment results generated by PROMALS3D which uses ClustalX colouring scheme. Here, the alignments colour is based on amino acid profile (residue type versus their alignment position). Hydrophobic (blue), Positive charge (red), Negative charge (magenta), Polar (green), Cysteines (pink), Glycines (orange), Prolines (yellow), Aromatic (cyan) and unconserved/gaps (white).

Motif 1	Pfa_MT-CYB	47	1.50e-50	FIQIITGVFL ASRYT DDVSY AYYSIQ ILRELWSG WCFR Y M ATGASLV FLT Y L ILRG LNYSYMYLPL
	H.sap_MT-CYB	52	1.12e-45	ILQITTGLFL AM YST DASTAFSSIA ITRDVNYG WIIRYL ANGASMFFICLFL IGRG LYYGSFLYSE
Motif 2	Pfa_MT-CYB	241	1.55e-41	IQSLFGIIPL S P DN AI VNT V T SQIV EW Y FL FYAMLKT V SK A L VIVLLSLQLL
	H.sap_MT-CYB	251	2.71e-38	TLTLFSPDLL GD DN Y TLAN LNT P IK EW FL FAY TILRSV NKLG GV LALLLSILIL
Motif 3	Pfa_MT-CYB	108	1.56e-40	NYSYMYLPLS WISGLILEMIFIVTAFVGYVLF WGQMS WGATVITNLLSSI PVAVIWICGG
	H.sap_MT-CYB	113	1.27e-41	YYGSFLYSET WNIGIILLLATMATAFMGYVLF WGQMSFWGATVITNLLSAI PYIGTDLVQW
Motif 4	Pfa_MT-CYB	152	8.67e-30	TNLLSSIPVA VIWICGGYTVSD TIKREFFVL FIL FI G LCIVFIHIF
	H.sap_MT-CYB	161	1.04e-29	SAIPYIGTDL VQWIWGGYSVDS TLTRFFTF FIL FII AALATLHLLF
Motif 5	Pfa_MT-CYB	15	1.71e-28	SINLVKAHLI NY C LN INFLWN YGFL LGIIFFIQIITG VFLASRYTPD
	H.sap_MT-CYB	20	3.37e-25	LMKLNHSLI DL T SNISAWWNFGSLLGACIL LIQITTG LFLAMHYS PD
Motif 6	Pfa_MT-CYB	182	7.36e-20	LHFILPFI ~~GL~~ CIVFI IFFL LGSTN LG YDTALKIP
	H.sap_MT-CYB	191	5.02e-15	FHFILPFI ~~IA~~ ALATL LLFL ETGSNN LG ITSHSDKI
Motif 7	Pfa_MT-CYB	328	6.50e-27	PIIWFMCAFY ALLWIGCQL QDIFILYGR LFI VLF FC SG LEVLVHYRRT
	H.sap_MT-CYB	334	3.60e-19	LYWLLAADLL ILTWIGGQ VSY FTIIGQVASVLYFTTI LILMPTISLI
Motif 8	Pfa_MT-CYB	202	6.57e-21	HLHGSTNPLG YD TALKI FY N LLSL LDVKG F NNVILFLIQ
	H.sap_MT-CYB	212	1.98e-16	ETGSNNPLGI TS SDKITF SY YTIKDAL GL LLFILSLMTL
Motif 11	Pfa_MT-CYB	97	1.33e-09	LLTYLHILRG LNYSYMYLPLS WISGLILEMI
	H.sap_MT-CYB	102	1.28e-07	ICLFLHIGRG LYYGSFLYSET WNIGIILLLA

Figure S3.3: Regular expressions of predicted conserved motifs comparing PfCytb and HsCytb proteins. The residues are coloured by amino acid category: Pink-H; Magenta-DE; Red-KR; Blue-ACFILVM; Orange-G; Yellow-P; Cyan-Y; Green-NQST

Table S4.1: Titratable residues and pKa values in the PfCytb-ISP protein complex. Reproduced from [141].

Residue	pK (int)	pKa_(1/2)	Residue	pK (int)	pKa_(1/2)	Residue	pK (int)	pKa_(1/2)
TYR-2	11.160	>12.000	CYS-310	10.463	10.189	ASP-473	6.271	<0.000
CYS-4	11.058	>12.000	TYR-313	12.143	>12.000	ARG-477	11.789	>12.000
TYR-14	11.330	>12.000	CYS-320	9.510	9.361	ARG-479	11.287	>12.000
ARG-35	12.118	>12.000	ASP-325	3.876	2.734	GLU-488	5.623	1.812
TYR-36	14.652	>12.000	TYR-330	10.027	10.072	TYR-492	12.303	>12.000
ASP-39	3.584	2.076	ARG-332	12.244	>12.000	GLU-493	4.930	3.562
TYR-42	10.429	10.636	CYS-340	9.619	9.393	ASP-496	4.608	3.001
TYR-44	11.718	>12.000	HIS-348	5.819	4.919	GLU-497	4.451	2.924
TYR-45	10.057	10.160	TYR-349	10.003	9.952	LYS-501	10.182	>12.000
HIS-49	6.067	6.658	ARG-350	13.576	>12.000	CTASN-361	5.192	4.681
ARG-52	12.409	>12.000	ARG-351	11.992	>12.000	ASP-473	6.271	<0.000
GLU-53	4.706	2.437	HIS-353	2.975	<0.000	ARG-477	11.789	>12.000
CYS-59	9.050	8.825	TYR-354	9.973	9.929	ARG-479	11.287	>12.000
ARG-61	10.818	>12.000	ASP-355	5.352	3.234			
TYR-62	11.376	11.657	TYR-356	9.834	10.132			
HID-64	6.239	4.746	NTASN-1	6.908	5.029			
TYR-76	11.774	>12.000	NT CYS-362	8.126	7.935			
HIS-78	4.667	<0.000	LYS-363	10.201	10.024			
ARG-81	10.864	>12.000	HIS-366	6.496	5.836			
TYR-85	12.449	10.926	LYS-372	10.203	10.941			
TYR-87	12.909	>12.000	ASP-373	3.921	2.741			
TYR-89	10.095	9.951	GLU-381	5.545	5.074			
TYR-112	10.773	>12.000	ASP-383	3.867	1.768			
TYR-121	12.813	>12.000	ARG-385	12.067	>12.000			
CYS-142	11.459	11.532	GLU-391	4.212	3.079			
TYR-145	12.226	9.966	HIS-392	5.725	4.122			
ASP-149	4.262	0.551	LYS-396	10.095	10.531			
LYS-153	9.714	>12.000	ARG-398	12.462	>12.000			

ARG-154	12.600	>12.000	LYS-400	9.934	>12.000	
HIS-159	5.792	2.608	LYS-405	10.189	>12.000	
CYS-168	9.273	9.182	HID-406	7.143	7.022	
HID-173	6.246	4.463	ARG-407	12.854	>12.000	
HIS-178	6.066	5.662	GLU-410	5.172	4.492	
HIS-180	6.564	5.832	ASP-411	4.318	<0.000	
TYR-188	11.056	10.993	ARG-414	12.066	>12.000	
ASP-189	3.965	2.709	LYS-416	9.838	>12.000	
LYS-193	9.978	10.823	GLU-417	5.511	2.963	
TYR-197	9.328	9.343	ASP-418	5.434	0.436	
ASP-204	4.730	3.064	ASP-419	3.755	2.706	
LYS-206	10.360	10.876	LYS-420	10.393	11.725	
HID-228	4.530	4.797	ARG-426	12.923	>12.000	
ASP-230	3.544	2.014	ASP-427	3.237	1.368	
TYR-238	12.779	>12.000	ASP-431	7.220	<0.000	
GLU-247	6.434	3.415	ASP-433	3.770	3.138	
TYR-249	14.109	>12.000	ARG-434	12.729	>12.000	
LYS-258	11.439	>12.000	LYS-437	9.841	>12.000	
LYS-263	10.896	11.739	GLU-439	5.317	3.630	
GLU-282	4.497	2.110	CYS-447	8.775	>12.000	
ARG-284	12.469	>12.000	HIS-449	3.925	<0.000	
LYS-293	10.144	10.054	TYR-460	10.496	11.004	
ARG-299	13.129	>12.000	TYR-463	14.864	>12.000	
ASP-300	4.115	2.877	CYS-465	7.745	6.648	
TYR-301	10.462	11.616	HID-468	5.114	1.246	

Table S4.2. Parameter and coordinate files for both Heme & [2FE-2S) cluster model B3LYP/6-31G* containing optimized coordinates and derived atomic partial charges (AMBER parameter files included also):

- (a) Protoporphyrin IX containing FE; Heme (HEM.mol2)
- (b) Heme proximal histidine ligand (His78.mol2)
- (c) Heme distal histidine ligand (His173.mol2)
- (d) [2FE-2S] proximal histidine ligand (His301.mol2)
- (e) [2FE-2S] distal histidine ligand (His320.mol2)
- (f) [2FE-2S] proximal cysteine ligand (Cys299.mol2)
- (g) [2FE-2S] distal cysteine ligand (Cys317.mol2)
- (h) An AMBER parameter (*.frcmod) file

(a) Protoporphyrin IX Heme octa-coordinate and its derived atomic charges.

HEM.mol2

@<TRIPOS>MOLECULE

HEM

72 76 1 0 0

SMALL

RESP Charge

@<TRIPOS>ATOM

1	CHA	56.5750	66.3360	63.4640	ce	1	HEM	-0.053091
2	CHB	60.1540	67.7760	66.4260	ce	1	HEM	-0.142569
3	CHC	57.4820	66.6740	70.2620	ce	1	HEM	-0.109350
4	CHD	53.9860	65.2570	67.2620	ce	1	HEM	-0.139521
5	C1A	57.8200	66.8080	63.9350	cc	1	HEM	-0.074815
6	C2A	58.9610	67.3250	63.1290	cc	1	HEM	-0.097541
7	C3A	59.9930	67.6700	63.9390	cd	1	HEM	0.042618
8	C4A	59.4420	67.4480	65.2850	cd	1	HEM	-0.055780
9	CMA	61.3210	68.2390	63.5190	c3	1	HEM	-0.105332
10	CAA	58.9710	67.6000	61.6440	c3	1	HEM	-0.022660
11	CBA	58.1140	68.8300	61.4130	c3	1	HEM	-0.041797
12	CGA	57.7700	69.0520	59.9490	c	1	HEM	0.541097
13	O1A	58.6220	69.5690	59.1970	o	1	HEM	-0.620901
14	O2A	56.6400	68.7090	59.5600	o	1	HEM	-0.413672
15	C1B	59.7360	67.6250	67.7480	cc	1	HEM	-0.099086
16	C2B	60.5620	67.8750	68.9500	cc	1	HEM	0.060434
17	C3B	59.8020	67.5880	70.0560	cd	1	HEM	-0.027679
18	C4B	58.5100	67.1300	69.5160	cd	1	HEM	-0.054395
19	CMB	61.9990	68.4120	68.9750	c3	1	HEM	-0.068779
20	CAB	60.1620	67.7100	71.3790	cf	1	HEM	-0.095735
21	CBB	61.4930	67.3490	72.0660	c2	1	HEM	-0.533903
22	C1C	56.3060	66.0990	69.7970	cc	1	HEM	-0.019937
23	C2C	55.2140	65.4930	70.6100	cc	1	HEM	0.051407
24	C3C	54.2770	65.0860	69.6910	cd	1	HEM	-0.018292
25	C4C	54.7650	65.4670	68.4040	cd	1	HEM	-0.059841
26	CMC	55.1110	65.3790	72.1150	c3	1	HEM	-0.139834
27	CAC	53.0820	64.4560	69.8290	cf	1	HEM	-0.148847
28	CBC	52.5290	63.7950	71.1010	c2	1	HEM	-0.560753
29	C1D	54.3020	65.4620	65.9720	cc	1	HEM	-0.053396
30	C2D	53.4620	65.0580	64.8450	cc	1	HEM	0.040491
31	C3D	54.1720	65.4390	63.7570	cd	1	HEM	-0.066318
32	C4D	55.4860	65.9830	64.2460	cd	1	HEM	-0.055737
33	CMD	52.1540	64.2910	64.9280	c3	1	HEM	-0.101229
34	CAD	53.8050	65.2490	62.2930	c3	1	HEM	-0.019091
35	CBD	53.1580	66.4690	61.6420	c3	1	HEM	-0.006871
36	CGD	54.1670	67.5250	61.2390	c	1	HEM	0.611618
37	O1D	54.0410	68.6920	61.6790	o	1	HEM	-0.659831
38	O2D	55.0900	67.1730	60.4790	o	1	HEM	-0.516704
39	NA	58.1540	66.9310	65.2830	Y3	1	HEM	-0.094695
40	NB	58.4810	67.1140	68.0970	Y4	1	HEM	-0.011203
41	NC	55.9850	66.0330	68.4530	Y5	1	HEM	0.040970
42	ND	55.5520	65.9930	65.6200	Y6	1	HEM	-0.059847
43	HMA1	61.5890	67.8460	62.6730	hc	1	HEM	0.040745
44	HMA2	61.9880	68.0370	64.1940	hc	1	HEM	0.040745
45	HMA3	61.2440	69.2010	63.4190	hc	1	HEM	0.040745
46	HMB1	62.1430	68.9140	69.7920	hc	1	HEM	0.030254

47	HMB2	62.1410	68.9920	68.2110	hc	1	HEM	0.030254
48	HMB3	62.6230	67.6700	68.9380	hc	1	HEM	0.030254
49	HMC1	54.1770	65.3950	72.3780	hc	1	HEM	0.051120
50	HMC2	55.5760	66.1230	72.5290	hc	1	HEM	0.051120
51	HMC3	55.5140	64.5460	72.4050	hc	1	HEM	0.051120
52	HMD1	51.5830	64.5440	64.1860	hc	1	HEM	0.037984
53	HMD2	51.7090	64.4990	65.7640	hc	1	HEM	0.037984
54	HMD3	52.3340	63.3390	64.8870	hc	1	HEM	0.037984
55	HBB1	61.6010	67.4930	73.0190	ha	1	HEM	0.197664
56	HBB2	62.2220	66.9780	71.5450	ha	1	HEM	0.197664
57	HBC1	51.6570	63.3700	71.0870	ha	1	HEM	0.187462
58	HBC2	53.0520	63.8020	71.9180	ha	1	HEM	0.187462
59	HBA1	58.5800	69.6100	61.7520	hc	1	HEM	0.046864
60	HBA2	57.2930	68.7460	61.9230	hc	1	HEM	0.046864
61	HAA1	58.6210	66.8410	61.1520	hc	1	HEM	0.044499
62	HAA2	59.8760	67.7490	61.3290	hc	1	HEM	0.044499
63	HBD1	52.5180	66.8580	62.2580	hc	1	HEM	-0.005904
64	HBD2	52.6610	66.1870	60.8580	hc	1	HEM	-0.005904
65	HAD1	53.1980	64.4960	62.2180	hc	1	HEM	0.022799
66	HAD2	54.6060	65.0180	61.7970	hc	1	HEM	0.022799
67	HHA	56.4760	66.2540	62.5430	ha	1	HEM	0.159129
68	HHB	61.0020	68.1360	66.2970	ha	1	HEM	0.148531
69	HHC	57.5700	66.7530	71.1840	ha	1	HEM	0.110019
70	HHH	53.1290	64.9300	67.4170	ha	1	HEM	0.111705
71	HAB	59.4790	68.0760	71.9630	ha	1	HEM	0.129768
72	HAC	52.5180	64.4190	69.0410	ha	1	HEM	0.165550

@<TRIPOS>BOND

1	1	5	1
2	1	32	1
3	1	67	1
4	2	8	1
5	2	15	1
6	2	68	1
7	3	18	1
8	3	22	1
9	3	69	1
10	4	25	1
11	4	29	1
12	4	70	1
13	5	6	1
14	5	39	1
15	6	7	1
16	6	10	1
17	7	8	1
18	7	9	1
19	8	39	1
20	9	43	1
21	9	44	1
22	9	45	1
23	10	11	1
24	10	61	1
25	10	62	1
26	11	12	1
27	11	59	1
28	11	60	1
29	12	13	1
30	12	14	1
31	15	16	1
32	15	40	1
33	16	17	1
34	16	19	1
35	17	18	1

36	17	20	1
37	18	40	1
38	19	46	1
39	19	47	1
40	19	48	1
41	20	21	1
42	20	71	1
43	21	55	1
44	21	56	1
45	22	23	1
46	22	41	1
47	23	24	1
48	23	26	1
49	24	25	1
50	24	27	1
51	25	41	1
52	26	49	1
53	26	50	1
54	26	51	1
55	27	28	1
56	27	72	1
57	28	57	1
58	28	58	1
59	29	30	1
60	29	42	1
61	30	31	1
62	30	33	1
63	31	32	1
64	31	34	1
65	32	42	1
66	33	52	1
67	33	53	1
68	33	54	1
69	34	35	1
70	34	65	1
71	34	66	1
72	35	36	1
73	35	63	1
74	35	64	1
75	36	37	1
76	36	38	1

@<TRIPOS>SUBSTRUCTURE

1 HEM

1 TEMP

0 ***** 0 ROOT

(b) HEME proximal histidine ligand coordinates and atomic charges

His78.mol2

@<TRIPOS>MOLECULE

HD3

17 17 1 0 0

SMALL

RESP Charge

@<TRIPOS>ATOM

1	N	55.2010	74.0280	65.4910	N	1	HD3	-0.415700
2	H	54.4690	73.8790	65.0650	H	1	HD3	0.235973
3	CA	55.9850	72.8320	65.7720	CX	1	HD3	0.018800
4	HA	56.2340	72.8430	66.7090	H1	1	HD3	0.056284
5	CB	55.1420	71.5730	65.5280	CT	1	HD3	0.082654
6	HB2	54.2330	71.7390	65.8230	HC	1	HD3	0.035925
7	HB3	55.1030	71.3980	64.5750	HC	1	HD3	0.035925
8	CG	55.6730	70.3480	66.2320	CC	1	HD3	-0.172662
9	ND1	55.4930	70.0840	67.5890	NA	1	HD3	-0.192872
10	HD1	55.1160	70.5950	68.1690	H	1	HD3	0.329990
11	CE1	56.0260	68.8700	67.8090	CR	1	HD3	-0.034138
12	HE1	56.0320	68.4340	68.6300	H5	1	HD3	0.083349
13	NE2	56.5490	68.3760	66.6750	Y1	1	HD3	-0.014437
14	CD2	56.3410	69.2980	65.6670	CV	1	HD3	-0.005823
15	HD2	56.6040	69.2190	64.7780	H4	1	HD3	0.159459
16	C	57.2920	72.8160	64.9870	C	1	HD3	0.597300
17	O	58.3470	72.7770	65.6090	O	1	HD3	-0.567900

@<TRIPOS>BOND

1	1	2	1
2	1	3	1
3	3	4	1
4	3	5	1
5	3	16	1
6	5	6	1
7	5	7	1
8	5	8	1
9	8	9	1
10	8	14	1
11	9	10	1
12	9	11	1
13	11	12	1
14	11	13	1
15	13	14	1
16	14	15	1
17	16	17	1

@<TRIPOS>SUBSTRUCTURE

1 HD3 1 TEMP

0 ***** 0 ROOT

(c) HEME distal histidine ligand coordinates and atomic charges

His173.mol2

@<TRIPOS>MOLECULE

HD4

17 17 1 0 0

SMALL

RESP Charge

@<TRIPOS>ATOM

1	N	58.4080	59.0740	68.6040	N	1	HD4	-0.415700
2	H	58.1220	58.7250	67.8720	H	1	HD4	0.228349
3	CA	58.3640	60.5340	68.5970	CX	1	HD4	0.018800
4	HA	59.2630	60.8110	68.8350	H1	1	HD4	0.074433
5	CB	58.0100	61.1030	67.2150	CT	1	HD4	0.048962
6	HB2	58.6540	60.7720	66.5700	HC	1	HD4	0.034786
7	HB3	57.1420	60.7590	66.9520	HC	1	HD4	0.034786
8	CG	57.9760	62.6190	67.1310	CC	1	HD4	-0.116106
9	ND1	59.0670	63.4650	67.3120	NA	1	HD4	-0.223384
10	HD1	59.8780	63.2330	67.4780	H	1	HD4	0.336786
11	CE1	58.6010	64.7120	67.1760	CR	1	HD4	-0.009107
12	HE1	59.1130	65.4830	67.2680	H5	1	HD4	0.136437
13	NE2	57.2840	64.6910	66.8870	Y2	1	HD4	0.028113
14	CD2	56.8650	63.3740	66.8620	CV	1	HD4	-0.124850
15	HD2	56.0040	63.0640	66.6960	H4	1	HD4	0.152567
16	C	57.3820	61.0980	69.6230	C	1	HD4	0.597300
17	O	57.6140	62.1550	70.2020	O	1	HD4	-0.567900

@<TRIPOS>BOND

1	1	2	1
2	1	3	1
3	3	4	1
4	3	5	1
5	3	16	1
6	5	6	1
7	5	7	1
8	5	8	1
9	8	9	1
10	8	14	1
11	9	10	1
12	9	11	1
13	11	12	1
14	11	13	1
15	13	14	1
16	14	15	1
17	16	17	1

@<TRIPOS>SUBSTRUCTURE

1 HD4 1 TEMP

0 ***** 0 ROOT

(d) [2FE-2S] cluster proximal histidine ligand coordinates and atomic charges

His320.mol2

@<TRIPOS>MOLECULE

HE2

17 17 1 0 0

SMALL

RESP Charge

@<TRIPOS>ATOM

1	N	76.6730	60.4870	51.0020	N	1	HE2	-0.415700
2	H	76.4680	59.6910	50.7480	H	1	HE2	0.263227
3	CA	78.0730	60.6710	51.2760	CX	1	HE2	-0.058100
4	HA	78.0620	61.5640	51.6540	H1	1	HE2	0.078755
5	CB	78.6600	59.8310	52.3950	CT	1	HE2	-0.005376
6	HB2	78.7920	58.9110	52.1170	HC	1	HE2	0.059976
7	HB3	79.5240	60.1690	52.6770	HC	1	HE2	0.059976
8	CG	77.6500	59.9330	53.4880	CC	1	HE2	0.130370
9	ND1	76.5060	59.1620	53.4310	Z4	1	HE2	-0.382996
10	CE1	75.7900	59.4580	54.5470	CR	1	HE2	-0.021464
11	HE1	74.9690	59.0730	54.7530	H5	1	HE2	0.155613
12	NE2	76.3940	60.3570	55.3070	NA	1	HE2	-0.148993
13	HE2	76.1180	60.6840	56.0530	H	1	HE2	0.327309
14	CD2	77.5690	60.6560	54.6360	CW	1	HE2	-0.266746
15	HD2	78.2090	61.2630	54.9290	H4	1	HE2	0.185024
16	C	78.9640	60.7970	50.0760	C	1	HE2	0.597300
17	O	80.1690	60.5950	50.2030	O	1	HE2	-0.567900

@<TRIPOS>BOND

1	1	2	1
2	1	3	1
3	3	4	1
4	3	5	1
5	3	16	1
6	5	6	1
7	5	7	1
8	5	8	1
9	8	9	1
10	8	14	1
11	9	10	1
12	10	11	1
13	10	12	1
14	12	13	1
15	12	14	1
16	14	15	1
17	16	17	1

@<TRIPOS>SUBSTRUCTURE

1 HE2 1 TEMP

0 ***** 0 ROOT

(e) [2FE-2S] cluster distal histidine ligand coordinates and atomic charges

His301.mol2

@<TRIPOS>MOLECULE

HE1

17 17 1 0 0

SMALL

RESP Charge

@<TRIPOS>ATOM

1	N	76.9840	52.1080	51.4450	N	1	HE1	-0.415700
2	H	76.9200	52.5150	50.5930	H	1	HE1	0.325555
3	CA	77.2400	52.6660	52.4610	CX	1	HE1	-0.058100
4	HA	78.0560	52.1390	52.9350	H1	1	HE1	0.038670
5	CB	77.5370	54.0500	51.9540	CT	1	HE1	0.080741
6	HB2	78.3760	54.0220	51.2350	HC	1	HE1	-0.027235
7	HB3	76.6270	54.3380	51.4260	HC	1	HE1	-0.027235
8	CG	77.7700	55.1340	52.9230	CC	1	HE1	0.082017
9	ND1	77.1630	55.4180	54.1410	Z2	1	HE1	-0.454443
10	CE1	77.7440	56.4410	54.6630	CR	1	HE1	0.089071
11	HE1	77.5230	56.9120	55.6000	H5	1	HE1	0.115415
12	NE2	78.5580	56.9320	53.6630	NA	1	HE1	-0.082129
13	HE2	79.3430	57.6740	53.8690	H	1	HE1	0.212982
14	CD2	78.5950	56.0730	52.6800	CW	1	HE1	-0.079996
15	HD2	79.1690	56.1450	51.8220	H4	1	HE1	0.190476
16	C	76.0740	52.6280	53.3580	C	1	HE1	0.597300
17	O	75.7280	51.6550	53.9980	O	1	HE1	-0.567900

@<TRIPOS>BOND

1	1	2	1
2	1	3	1
3	3	4	1
4	3	5	1
5	3	16	1
6	5	6	1
7	5	7	1
8	5	8	1
9	8	9	1
10	8	14	1
11	9	10	1
12	10	11	1
13	10	12	1
14	12	13	1
15	12	14	1
16	14	15	1
17	16	17	1

@<TRIPOS>SUBSTRUCTURE

1 HE1 1 TEMP

0 ***** 0 ROOT

(f) [2FE-2S] cluster proximal cysteine ligand coordinates and atomic charges

Cys317.mol2

@<TRIPOS>MOLECULE

CX1

10 9 1 0 0

SMALL

RESP Charge

@<TRIPOS>ATOM

1	N	73.2370	51.9060	47.9180	N	1	CX1	-0.415700
2	H	72.7060	52.5760	48.0110	H	1	CX1	0.267345
3	CA	74.6370	52.1740	48.1270	CX	1	CX1	0.042900
4	HA	75.0900	51.7480	47.3830	H1	1	CX1	0.028004
5	CB	75.1250	53.6340	47.9420	2C	1	CX1	0.147741
6	HB2	76.0690	53.6020	47.7190	H1	1	CX1	-0.056234
7	HB3	74.6640	54.0020	47.1720	H1	1	CX1	-0.056234
8	SG	74.9250	54.8020	49.3090	Y7	1	CX1	-0.276273
9	C	75.0760	51.5540	49.4170	C	1	CX1	0.597300
10	O	74.4540	51.7120	50.4670	O	1	CX1	-0.567900

@<TRIPOS>BOND

1	1	2	1
2	1	3	1
3	3	4	1
4	3	5	1
5	3	9	1
6	5	6	1
7	5	7	1
8	5	8	1
9	9	10	1

@<TRIPOS>SUBSTRUCTURE

1	CX1	1	TEMP	0	****	****	0	ROOT

(g) [2FE-2S] cluster distal cysteine ligand coordinates and atomic charges

Cys299.mol2

@<TRIPOS>MOLECULE

CX2

10 9 1 0 0

SMALL

RESP Charge

@<TRIPOS>ATOM

1 N	76.1250	58.5360	44.7370	N	1 CX2	-0.415700
2 H	76.9360	58.3270	44.9330	H	1 CX2	0.247913
3 CA	75.2200	58.5520	45.8390	CX	1 CX2	0.042900
4 HA	74.2890	58.5170	45.5690	H1	1 CX2	0.028474
5 CB	75.4370	57.3240	46.6870	2C	1 CX2	-0.023435
6 HB2	75.4190	56.5340	46.1240	H1	1 CX2	0.098355
7 HB3	76.3170	57.3620	47.0940	H1	1 CX2	0.098355
8 SG	74.1830	57.1740	47.9760	Y8	1 CX2	-0.454266
9 C	75.4600	59.8410	46.5610	C	1 CX2	0.597300
10 O	76.5310	60.0950	47.1120	O	1 CX2	-0.567900

@<TRIPOS>BOND

1	1	2	1
2	1	3	1
3	3	4	1
4	3	5	1
5	3	9	1
6	5	6	1
7	5	7	1
8	5	8	1
9	9	10	1

@<TRIPOS>SUBSTRUCTURE

1 CX2	1 TEMP	0 ****	****	0 ROOT

(h) An AMBER parameter (*.frcmod) file

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

MASS

M1	55.85		Fe ion
M2	55.85		Fe ion
M3	55.85		Fe ion
Y1	14.01	0.530	sp2 N in 5 memb.ring w/LP (HIS,ADE,GUA)
Y2	14.01	0.530	sp2 N in 5 memb.ring w/LP (HIS,ADE,GUA)
Y7	32.06	2.900	S in disulfide linkage,pol:JPC,102,2399,98
Z2	14.01	0.530	sp2 N in 5 memb.ring w/LP (HIS,ADE,GUA)
Y8	32.06	2.900	S in disulfide linkage,pol:JPC,102,2399,98
Z4	14.01	0.530	sp2 N in 5 memb.ring w/LP (HIS,ADE,GUA)
Y3	14.01	0.530	Sp2 N in non-pure aromatic systems, identical
Y4	14.01	0.530	Sp2 N in non-pure aromatic systems, identical
Y5	14.01	0.530	Sp2 N in non-pure aromatic systems, identical
Y6	14.01	0.530	Sp2 N in non-pure aromatic systems, identical
Z1	32.060	2.900	same as s
Z3	32.060	2.900	same as s

BOND

M2-Z1	139.7	2.1207	Created by Seminario method using MCPB.py
M2-Z3	125.7	2.1429	Created by Seminario method using MCPB.py
M3-Z1	52.8	2.3568	Created by Seminario method using MCPB.py
M3-Z3	76.5	2.2768	Created by Seminario method using MCPB.py
Y1-M1	42.7	2.0205	Created by Seminario method using MCPB.py
Y2-M1	43.0	2.0146	Created by Seminario method using MCPB.py
Y3-M1	55.9	2.0073	Created by Seminario method using MCPB.py
Y4-M1	44.2	2.0251	Created by Seminario method using MCPB.py
Y5-M1	46.2	2.0276	Created by Seminario method using MCPB.py
Y6-M1	64.2	2.0069	Created by Seminario method using MCPB.py
Y7-M2	131.0	2.1410	Created by Seminario method using MCPB.py
Y8-M2	118.6	2.1544	Created by Seminario method using MCPB.py
Z2-M3	32.8	2.2053	Created by Seminario method using MCPB.py
Z4-M3	74.0	1.9660	Created by Seminario method using MCPB.py
2C-Y7	227.0	1.8100	
2C-Y8	227.0	1.8100	
CC-Z2	410.0	1.394	JCC,7,(1986),230; HIS
CC-Z4	410.0	1.394	JCC,7,(1986),230; HIS
CR-Y1	488.0	1.335	JCC,7,(1986),230; HIS
CR-Y2	488.0	1.335	JCC,7,(1986),230; HIS
Y1-CV	410.0	1.394	JCC,7,(1986),230; HIS
Y2-CV	410.0	1.394	JCC,7,(1986),230; HIS
Z2-CR	488.0	1.335	JCC,7,(1986),230; HIS
Z4-CR	488.0	1.335	JCC,7,(1986),230; HIS
cc-Y3	525.4	1.3172	SOURCE3_SOURCE5 4612 0.0083
cc-Y4	525.4	1.3172	SOURCE3_SOURCE5 4612 0.0083
cc-Y5	525.4	1.3172	SOURCE3_SOURCE5 4612 0.0083
cc-Y6	525.4	1.3172	SOURCE3_SOURCE5 4612 0.0083
cd-Y3	441.1	1.3694	SOURCE1_SOURCE5 2269 0.0086
cd-Y4	441.1	1.3694	SOURCE1_SOURCE5 2269 0.0086
cd-Y5	441.1	1.3694	SOURCE1_SOURCE5 2269 0.0086
cd-Y6	441.1	1.3694	SOURCE1_SOURCE5 2269 0.0086

ANGLE

2C-SG1-FE2	113.72	105.40	Created by Seminario method using MCPB.py
2C-SG2-FE2	114.44	108.29	Created by Seminario method using MCPB.py

CC-ND1-FE3	27.74	75.99	Created by Seminario method using MCPB.py
CC-NDI-FE3	111.09	131.96	Created by Seminario method using MCPB.py
CR-NE2-FE1	104.90	125.70	Created by Seminario method using MCPB.py
CR-NE2-FE1	116.43	125.74	Created by Seminario method using MCPB.py
FE1-NE2-CV	100.14	127.65	Created by Seminario method using MCPB.py
FE1-NE2-CV	116.63	127.90	Created by Seminario method using MCPB.py
FE2-S1-FE3	47.73	82.71	Created by Seminario method using MCPB.py
FE2-S2-FE3	59.04	84.16	Created by Seminario method using MCPB.py
FE3-ND1-CR	0.00	64.27	Created by Seminario method using MCPB.py
FE3-ND1-CR	113.94	120.57	Created by Seminario method using MCPB.py
NE2-FE1-NE2	168.21	178.29	Created by Seminario method using MCPB.py
NE2-FE1-NA	134.59	91.43	Created by Seminario method using MCPB.py
NE2-FE1-NB	139.06	91.23	Created by Seminario method using MCPB.py
NE2-FE1-NC	127.35	91.04	Created by Seminario method using MCPB.py
NE2-FE1-ND	134.33	87.49	Created by Seminario method using MCPB.py
NE2-FE1-NA	118.97	88.52	Created by Seminario method using MCPB.py
NE2-FE1-NB	130.18	90.49	Created by Seminario method using MCPB.py
NE2-FE1-NC	116.87	89.01	Created by Seminario method using MCPB.py
NE2-FE1-ND	126.00	90.80	Created by Seminario method using MCPB.py
NA-FE1-NB	148.86	90.13	Created by Seminario method using MCPB.py
NA-FE1-NC	98.29	177.53	Created by Seminario method using MCPB.py
NA-FE1-ND	141.95	90.05	Created by Seminario method using MCPB.py
NB-FE1-NC	132.38	89.77	Created by Seminario method using MCPB.py
NB-FE1-ND	120.06	178.71	Created by Seminario method using MCPB.py
NC-FE1-ND	130.63	90.11	Created by Seminario method using MCPB.py
SG1-FE2-SG2	71.89	112.81	Created by Seminario method using MCPB.py
SG1-FE2-S1	62.87	118.90	Created by Seminario method using MCPB.py
SG1-FE2-S2	49.14	102.83	Created by Seminario method using MCPB.py
SG2-FE2-S1	57.34	110.40	Created by Seminario method using MCPB.py
SG2-FE2-S2	61.01	109.01	Created by Seminario method using MCPB.py
ND1-FE3-S1	9.62	147.65	Created by Seminario method using MCPB.py
ND1-FE3-S2	40.97	97.83	Created by Seminario method using MCPB.py
ND1-FE3-ND1	23.70	114.21	Created by Seminario method using MCPB.py
S2-FE2-S1	46.25	101.55	Created by Seminario method using MCPB.py
S2-FE3-S1	93.46	90.92	Created by Seminario method using MCPB.py
ND1-FE3-S1	64.05	95.81	Created by Seminario method using MCPB.py
ND1-FE3-S2	55.17	95.03	Created by Seminario method using MCPB.py
cc-NA-FE1	144.34	127.03	Created by Seminario method using MCPB.py
cc-NB-FE1	132.43	126.78	Created by Seminario method using MCPB.py
cc-NC-FE1	139.16	127.11	Created by Seminario method using MCPB.py
cc-ND-FE1	154.01	126.94	Created by Seminario method using MCPB.py
cd-NA-FE1	138.95	127.38	Created by Seminario method using MCPB.py
cd-NB-FE1	133.72	127.36	Created by Seminario method using MCPB.py
cd-NC-FE1	138.87	126.95	Created by Seminario method using MCPB.py
cd-ND-FE1	149.99	126.74	Created by Seminario method using MCPB.py
CC-CV-Y1	70.0	120.00	AA his
CC-CV-Y2	70.0	120.00	AA his
CC-Z2-CR	70.0	117.00	AA his
CC-Z4-CR	70.0	117.00	AA his
CR-Y1-CV	70.0	117.00	AA his
CR-Y2-CV	70.0	117.00	AA his
CT-CC-Z2	70.0	120.00	AA his
CT-CC-Z4	70.0	120.00	AA his
CW-CC-Z2	70.0	120.00	AA his
CW-CC-Z4	70.0	120.00	AA his
CX-2C-SG1	50.0	114.70	
CX-2C-Y8	50.0	114.70	
NA-CR-Y1	70.0	120.00	AA his
NA-CR-Y2	70.0	120.00	AA his
Y1-CR-H5	50.0	120.00	AA his
Y1-CV-H4	50.0	120.00	AA his
Y2-CR-H5	50.0	120.00	AA his

Y2-CV-H4	50.0	120.00	AA his			
Y3-cc-cc	71.57	112.56		SOURCE3	141	4.2871
Y4-cc-cc	71.57	112.56		SOURCE3	141	4.2871
Y5-cc-cc	71.57	112.56		SOURCE3	141	4.2871
Y6-cc-cc	71.57	112.56		SOURCE3	141	4.2871
SG1-2C-H1	50.0	109.50				
Y8-2C-H1	50.0	109.50				
Z2-CR-H5	50.0	120.00	AA his			
Z2-CR-NA	70.0	120.00	AA his			
Z4-CR-H5	50.0	120.00	AA his			
Z4-CR-NA	70.0	120.00	AA his			
cc-Y3-cd	71.76	105.49		CORR_SOURCE5	1810	1.9032
cc-Y4-cd	71.76	105.49		CORR_SOURCE5	1810	1.9032
cc-Y5-cd	71.76	105.49		CORR_SOURCE5	1810	1.9032
cc-Y6-cd	71.76	105.49		CORR_SOURCE5	1810	1.9032
cd-cd-Y3	67.63	121.98		CORR_SOURCE5	141	1.9633
cd-cd-Y4	67.63	121.98		CORR_SOURCE5	141	1.9633
cd-cd-Y5	67.63	121.98		CORR_SOURCE5	141	1.9633
cd-cd-Y6	67.63	121.98		CORR_SOURCE5	141	1.9633
ce-cc-Y3	68.07	121.70		CORR_SOURCE5	58	1.4179
ce-cc-Y4	68.07	121.70		CORR_SOURCE5	58	1.4179
ce-cc-Y5	68.07	121.70		CORR_SOURCE5	58	1.4179
ce-cc-Y6	68.07	121.70		CORR_SOURCE5	58	1.4179
ce-cd-Y3	68.67	123.98		SOURCE4_SOURCE5	10	2.4097
ce-cd-Y4	68.67	123.98		SOURCE4_SOURCE5	10	2.4097
ce-cd-Y5	68.67	123.98		SOURCE4_SOURCE5	10	2.4097
ce-cd-Y6	68.67	123.98		SOURCE4_SOURCE5	10	2.4097

DIHE

X -CC-Z2-X	2	4.8	180.0	2.0	JCC, 7, (1986), 230
X -CC-Z4-X	2	4.8	180.0	2.0	JCC, 7, (1986), 230
X -CR-Y1-X	2	10.0	180.0	2.0	JCC, 7, (1986), 230
X -CR-Y2-X	2	10.0	180.0	2.0	JCC, 7, (1986), 230
X -CV-Y1-X	2	4.8	180.0	2.0	JCC, 7, (1986), 230
X -CV-Y2-X	2	4.8	180.0	2.0	JCC, 7, (1986), 230
X -Z2-CR-X	2	10.0	180.0	2.0	JCC, 7, (1986), 230
X -Z4-CR-X	2	10.0	180.0	2.0	JCC, 7, (1986), 230
X -cc-Y3-X	2	9.5	180.0	2.0	statistic value from parm94
X -cc-Y4-X	2	9.5	180.0	2.0	statistic value from parm94
X -cc-Y5-X	2	9.5	180.0	2.0	statistic value from parm94
X -cc-Y6-X	2	9.5	180.0	2.0	statistic value from parm94
X -cd-Y3-X	2	9.5	180.0	2.0	statistic value from parm94
X -cd-Y4-X	2	9.5	180.0	2.0	statistic value from parm94
X -cd-Y5-X	2	9.5	180.0	2.0	statistic value from parm94
X -cd-Y6-X	2	9.5	180.0	2.0	statistic value from parm94
2C-SG1-M2-Y8	3	0.00	0.00	3.0	Treat as zero by MCPB.py
2C-SG1-M2-Z1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
2C-SG1-M2-Z3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
2C-Y8-M2-Z1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
2C-Y8-M2-Z3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
C -CX-2C-SG1	1	0.278	0.0	-4.0	C-C
C -CX-2C-SG1	1	0.323	0.0	-3.0	
C -CX-2C-SG1	1	0.394	180.0	-2.0	
C -CX-2C-SG1	1	0.602	0.0	1.0	
C -CX-2C-Y8	1	0.278	0.0	-4.0	C-C
C -CX-2C-Y8	1	0.323	0.0	-3.0	
C -CX-2C-Y8	1	0.394	180.0	-2.0	
C -CX-2C-Y8	1	0.602	0.0	1.0	
CC-CV-Y1-M1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CC-CV-Y2-M1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CC-Z2-M3-Z1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CC-Z2-M3-Z3	3	0.00	0.00	3.0	Treat as zero by MCPB.py

CC-Z2-M3-Z4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CC-Z4-M3-Z1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CC-Z4-M3-Z3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CR-Y1-M1-Y2	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CR-Y1-M1-Y3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CR-Y1-M1-Y4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CR-Y1-M1-Y5	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CR-Y1-M1-Y6	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CR-Y2-M1-Y3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CR-Y2-M1-Y4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CR-Y2-M1-Y5	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CR-Y2-M1-Y6	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CT-CC-Z2-M3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CT-CC-Z4-M3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CW-CC-Z2-M3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CW-CC-Z4-M3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CX-2C-Y7-M2	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CX-2C-Y8-M2	3	0.00	0.00	3.0	Treat as zero by MCPB.py
CX-CT-CC-Z2	1	0.047	180.0	-4.0	
CX-CT-CC-Z2	1	0.74	0.0	-3.0	
CX-CT-CC-Z2	1	0.204	0.0	-2.0	
CX-CT-CC-Z2	1	0.69	0.0	1.0	
CX-CT-CC-Z4	1	0.047	180.0	-4.0	
CX-CT-CC-Z4	1	0.74	0.0	-3.0	
CX-CT-CC-Z4	1	0.204	0.0	-2.0	
CX-CT-CC-Z4	1	0.69	0.0	1.0	
M1-Y1-CR-H5	3	0.00	0.00	3.0	Treat as zero by MCPB.py
M1-Y1-CV-H4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
M1-Y2-CR-H5	3	0.00	0.00	3.0	Treat as zero by MCPB.py
M1-Y2-CV-H4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
M1-Y3-cc-cc	3	0.00	0.00	3.0	Treat as zero by MCPB.py
M1-Y4-cc-cc	3	0.00	0.00	3.0	Treat as zero by MCPB.py
M1-Y5-cc-cc	3	0.00	0.00	3.0	Treat as zero by MCPB.py
M1-Y6-cc-cc	3	0.00	0.00	3.0	Treat as zero by MCPB.py
M2-Y7-2C-H1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
M2-Y8-2C-H1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
M2-Z1-M3-Z3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
M2-Z3-M3-Z1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
M3-Z2-CR-H5	3	0.00	0.00	3.0	Treat as zero by MCPB.py
M3-Z2-CR-NA	3	0.00	0.00	3.0	Treat as zero by MCPB.py
M3-Z3-M2-Z1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
M3-Z4-CR-H5	3	0.00	0.00	3.0	Treat as zero by MCPB.py
M3-Z4-CR-NA	3	0.00	0.00	3.0	Treat as zero by MCPB.py
N -CX-2C-Y7	1	0.064	0.0	-4.0	
N -CX-2C-Y7	1	0.323	0.0	-3.0	
N -CX-2C-Y7	1	0.021	180.0	-2.0	
N -CX-2C-Y7	1	0.469	0.0	1.0	
N -CX-2C-Y8	1	0.064	0.0	-4.0	
N -CX-2C-Y8	1	0.323	0.0	-3.0	
N -CX-2C-Y8	1	0.021	180.0	-2.0	
N -CX-2C-Y8	1	0.469	0.0	1.0	
NA-CR-Y1-M1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
NA-CR-Y2-M1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y1-M1-Y2-CR	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y1-M1-Y2-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y1-M1-Y3-cc	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y1-M1-Y3-cd	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y1-M1-Y4-cc	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y1-M1-Y4-cd	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y1-M1-Y5-cc	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y1-M1-Y5-cd	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y1-M1-Y6-cc	3	0.00	0.00	3.0	Treat as zero by MCPB.py

Y1-M1-Y6-cd	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y2-M1-Y1-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y2-M1-Y3-cc	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y2-M1-Y3-cd	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y2-M1-Y4-cc	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y2-M1-Y4-cd	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y2-M1-Y5-cc	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y2-M1-Y5-cd	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y2-M1-Y6-cc	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y2-M1-Y6-cd	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y3-M1-Y1-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y3-M1-Y2-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y4-M1-Y1-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y4-M1-Y2-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y4-cc-ce-cd	1	1.0	180.0	2.0	same as X -ce-ce-X
Y5-M1-Y1-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y5-M1-Y2-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y5-cc-ce-cd	1	1.0	180.0	2.0	same as X -ce-ce-X
Y6-M1-Y1-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y6-M1-Y2-CV	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y6-cc-ce-cd	1	1.0	180.0	2.0	same as X -ce-ce-X
Y6-cd-ce-cc	1	1.0	180.0	2.0	same as X -ce-ce-X
Y7-M2-Y8-2C	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y7-M2-Z1-M3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y7-M2-Z3-M3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y8-M2-Z1-M3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Y8-M2-Z3-M3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Z1-M3-Z2-CR	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Z1-M3-Z4-CR	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Z2-M3-Z1-M2	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Z2-M3-Z3-M2	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Z2-M3-Z4-CC	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Z2-M3-Z4-CR	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Z3-M2-Z1-M3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Z3-M3-Z2-CR	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Z3-M3-Z4-CR	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Z4-M3-Z1-M2	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Z4-M3-Z2-CR	3	0.00	0.00	3.0	Treat as zero by MCPB.py
Z4-M3-Z3-M2	3	0.00	0.00	3.0	Treat as zero by MCPB.py
cc-Y3-M1-Y4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
cc-Y3-M1-Y5	3	0.00	0.00	3.0	Treat as zero by MCPB.py
cc-Y3-M1-Y6	3	0.00	0.00	3.0	Treat as zero by MCPB.py
cc-Y4-M1-Y3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
cc-Y4-M1-Y5	3	0.00	0.00	3.0	Treat as zero by MCPB.py
cc-Y4-M1-Y6	3	0.00	0.00	3.0	Treat as zero by MCPB.py
cc-Y5-M1-Y3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
cc-Y5-M1-Y4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
cc-Y5-M1-Y6	3	0.00	0.00	3.0	Treat as zero by MCPB.py
cc-Y6-M1-Y3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
cc-Y6-M1-Y4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
cc-Y6-M1-Y5	3	0.00	0.00	3.0	Treat as zero by MCPB.py
cc-ce-cd-Y3	1	1.0	180.0	2.0	same as X -ce-ce-X
cc-ce-cd-Y4	1	1.0	180.0	2.0	same as X -ce-ce-X
cc-ce-cd-Y5	1	1.0	180.0	2.0	same as X -ce-ce-X
cd-Y3-M1-Y4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
cd-Y3-M1-Y5	3	0.00	0.00	3.0	Treat as zero by MCPB.py
cd-Y3-M1-Y6	3	0.00	0.00	3.0	Treat as zero by MCPB.py
cd-Y4-M1-Y3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
cd-Y4-M1-Y5	3	0.00	0.00	3.0	Treat as zero by MCPB.py
cd-Y4-M1-Y6	3	0.00	0.00	3.0	Treat as zero by MCPB.py
cd-Y5-M1-Y3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
cd-Y5-M1-Y4	3	0.00	0.00	3.0	Treat as zero by MCPB.py

cd-Y5-M1-Y6	3	0.00	0.00	3.0	Treat as zero by MCPB.py
cd-Y6-M1-Y3	3	0.00	0.00	3.0	Treat as zero by MCPB.py
cd-Y6-M1-Y4	3	0.00	0.00	3.0	Treat as zero by MCPB.py
cd-Y6-M1-Y5	3	0.00	0.00	3.0	Treat as zero by MCPB.py
cd-cd-Y3-M1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
cd-cd-Y4-M1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
cd-cd-Y5-M1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
cd-cd-Y6-M1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
cd-ce-cc-Y3	1	1.0	180.0	2.0	same as X -ce-ce-X
ce-cc-Y3-M1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
ce-cc-Y4-M1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
ce-cc-Y5-M1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
ce-cc-Y6-M1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
ce-cd-Y3-M1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
ce-cd-Y4-M1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
ce-cd-Y5-M1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
ce-cd-Y6-M1	3	0.00	0.00	3.0	Treat as zero by MCPB.py
ha-ce-cc-Y3	1	1.0	180.0	2.0	same as X -ce-ce-X
ha-ce-cc-Y4	1	1.0	180.0	2.0	same as X -ce-ce-X
ha-ce-cc-Y5	1	1.0	180.0	2.0	same as X -ce-ce-X
ha-ce-cc-Y6	1	1.0	180.0	2.0	same as X -ce-ce-X
ha-ce-cd-Y3	1	1.0	180.0	2.0	same as X -ce-ce-X
ha-ce-cd-Y4	1	1.0	180.0	2.0	same as X -ce-ce-X
ha-ce-cd-Y5	1	1.0	180.0	2.0	same as X -ce-ce-X
ha-ce-cd-Y6	1	1.0	180.0	2.0	same as X -ce-ce-X

IMPR

CT-CW-CC-Z2	1.1	180.	2.	
Y3-cd-cd-ce	1.1	180.0	2.0	Using default value
Y4-cd-cd-ce	1.1	180.0	2.0	Using default value
CT-CW-CC-Z4	1.1	180.	2.	
Y6-cc-cc-ce	1.1	180.0	2.0	Using default value
Y6-cd-cd-ce	1.1	180.0	2.0	Using default value
Y4-cc-cc-ce	1.1	180.0	2.0	Using default value
Y5-cd-cd-ce	1.1	180.0	2.0	Using default value
Y5-cc-cc-ce	1.1	180.0	2.0	Using default value
Y3-cc-cc-ce	1.1	180.0	2.0	Using default value

NONB

M1	1.4090	0.0172100000	IOD set for Fe2+ ion from Li et al. 2013
M2	1.4090	0.0172100000	IOD set for Fe2+ ion from Li et al. 2013
M3	1.4090	0.0172100000	IOD set for Fe2+ ion from Li et al. 2013
Y1	1.8240	0.1700	OPLS
Y2	1.8240	0.1700	OPLS
Y7	2.0000	0.2500	W. Cornell CH3SH and CH3SCH3 FEP's
Z2	1.8240	0.1700	OPLS
Y8	2.0000	0.2500	W. Cornell CH3SH and CH3SCH3 FEP's
Z4	1.8240	0.1700	OPLS
Y3	1.8240	0.1700	OPLS
Y4	1.8240	0.1700	OPLS
Y5	1.8240	0.1700	OPLS
Y6	1.8240	0.1700	OPLS
Z1	2.0000	0.2500	same as s
Z3	2.0000	0.2500	same as s

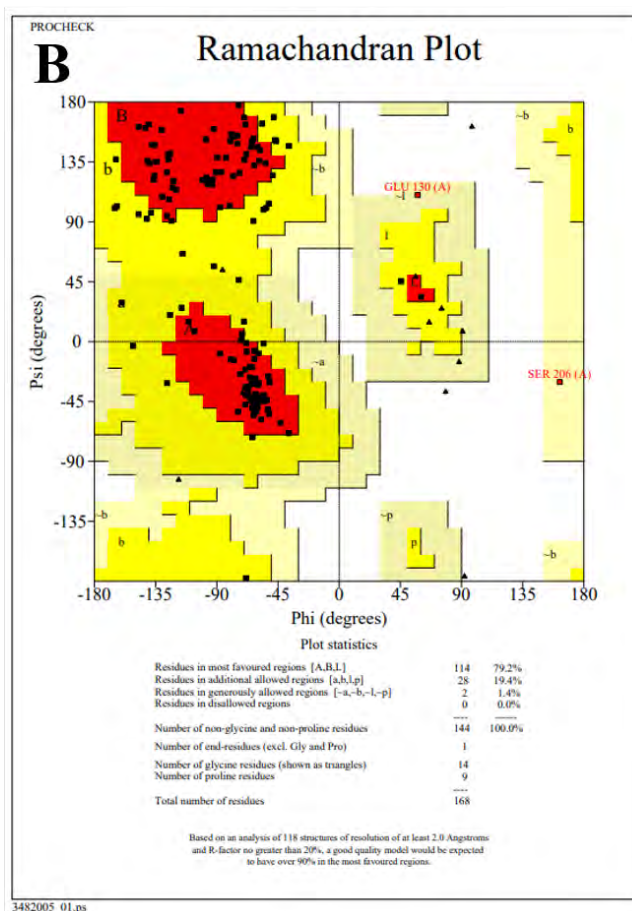
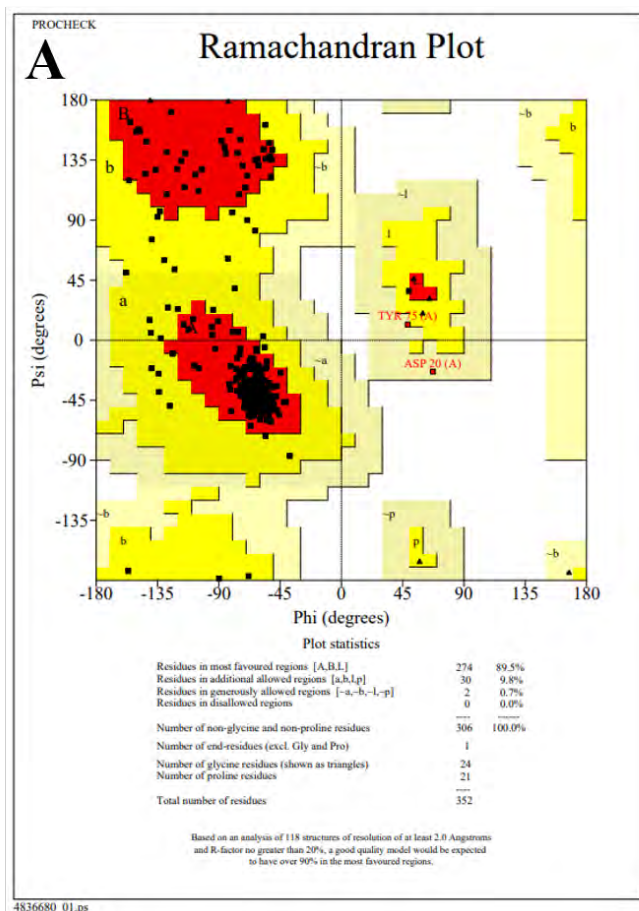


Figure S5.1: HsCytb-ISP model validation in form of Ramachandran plots. (A) and (B) representing HsCytb and ISP respectively using RAMPAGE. Most of the residues are observed to be in the most favored and allowed regions. Overall, the model quality was good.

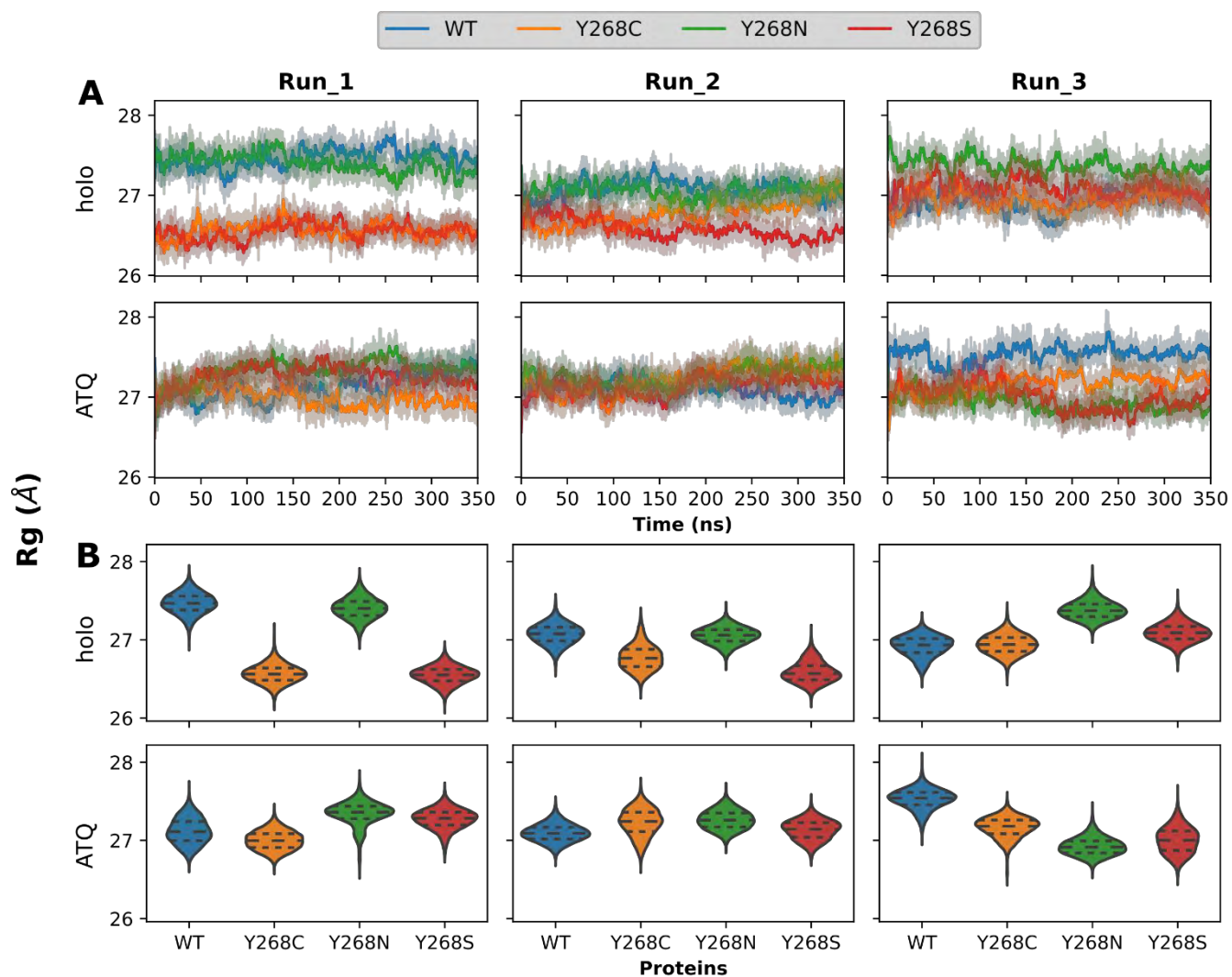


Figure S5.2: R_g comparison of WT and mutant both in the presence and absence of ATQ during the 350 ns MD simulation. **(A)** Line graph representation. **(B)** Violin plot presentation. Reproduced from [141].

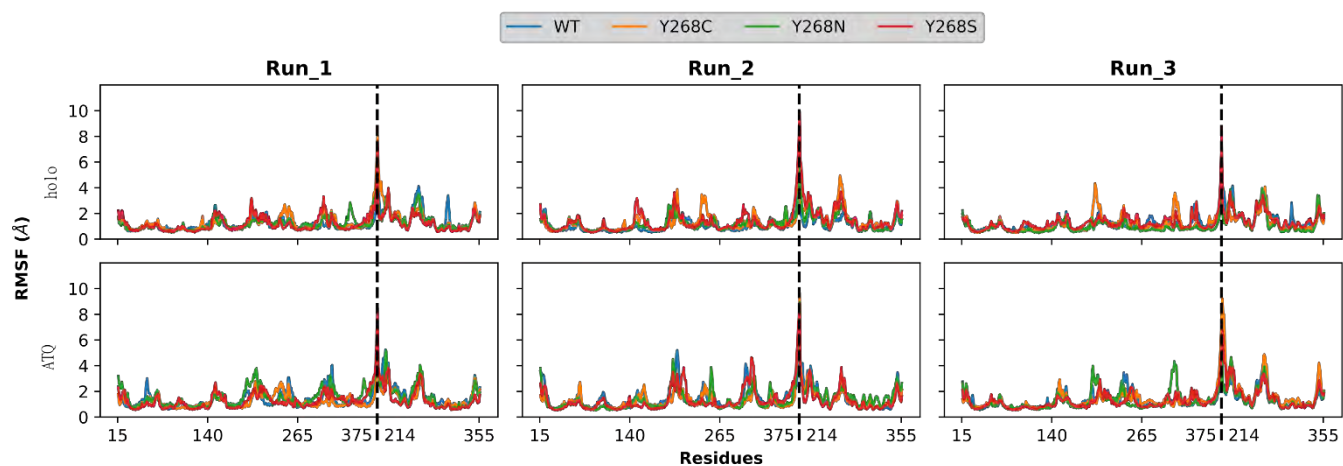


Figure S5.3: Line graph comparison of RMSF as observed in the WT and mutant PfCytb-ISP protein systems during 350 ns MD simulation in the presence and absence of ATQ. A total of three replicate runs were calculated and displayed. High values show regions of high flexibility. Reproduced from [141].

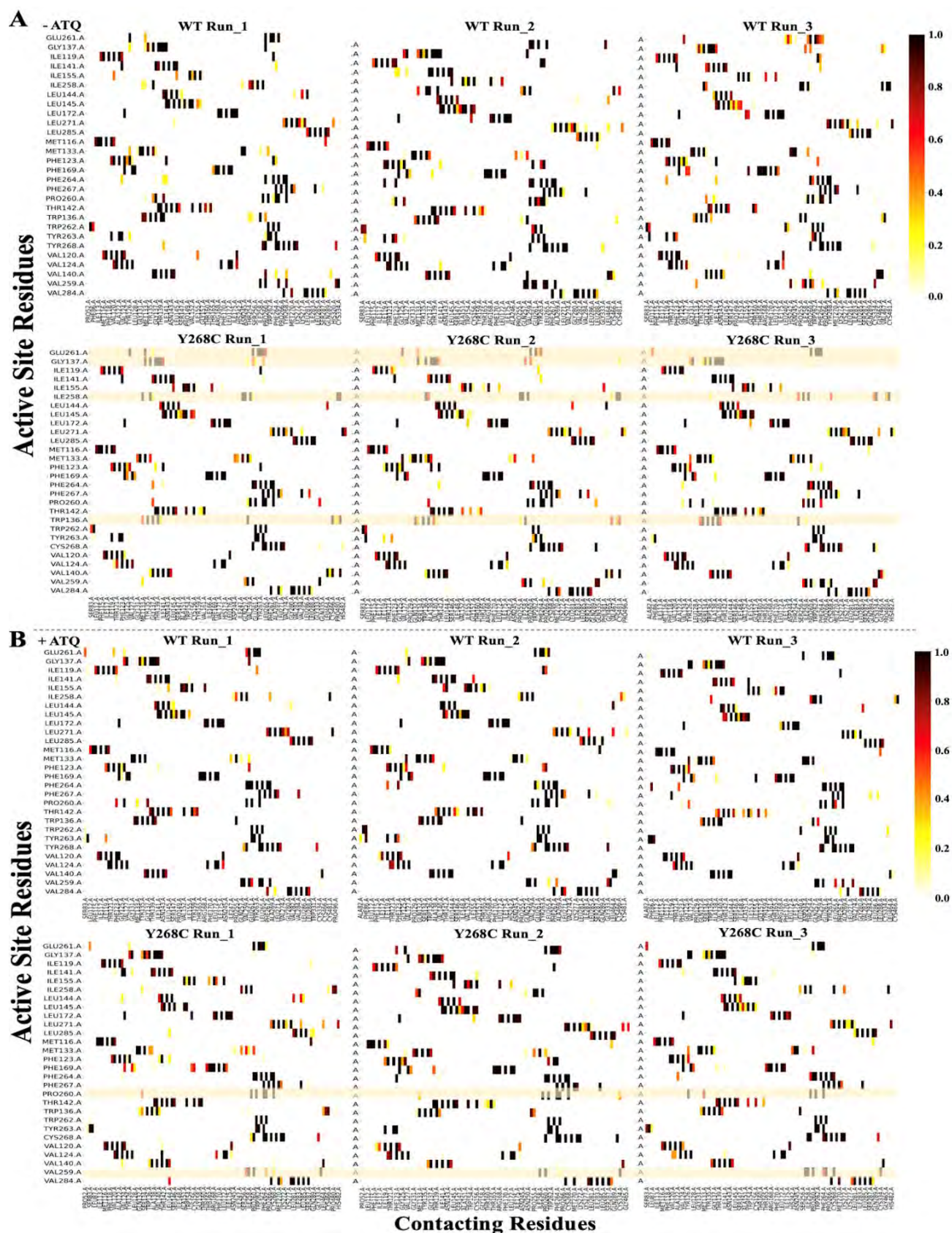


Figure S5.4: Residue contact heatmap between active site residues ($n = 28$) and other residue contacts. Comparison of WT and mutant system Y268C was performed in both Holo and ATQ-bound proteins in their respective triplicate runs in the last 100 ns of MD simulation time. Regions observed to reduce or lose residue contact are shaded in light golden colour. Reproduced from [141].

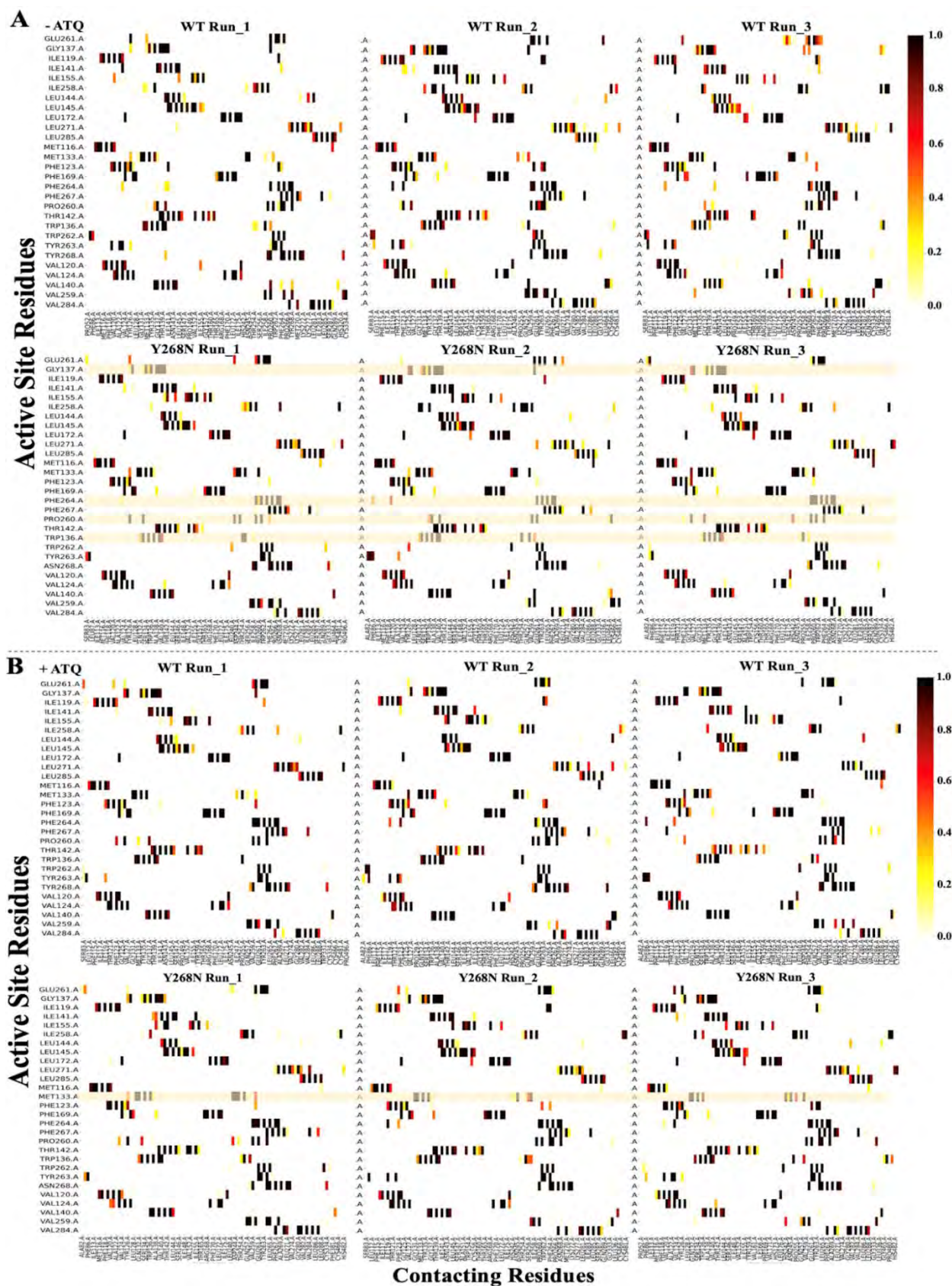


Figure S5.5: Residue contact heatmap between active site residues ($n = 28$) and other residue contacts. Comparison of the WT and mutant system Y268N was performed in both Holo and ATQ-bound proteins in their respective triplicate runs in the last 100 ns of MD simulation time. Regions observed to reduce or lose residue contact are shaded in light golden colour. Reproduced from [141].

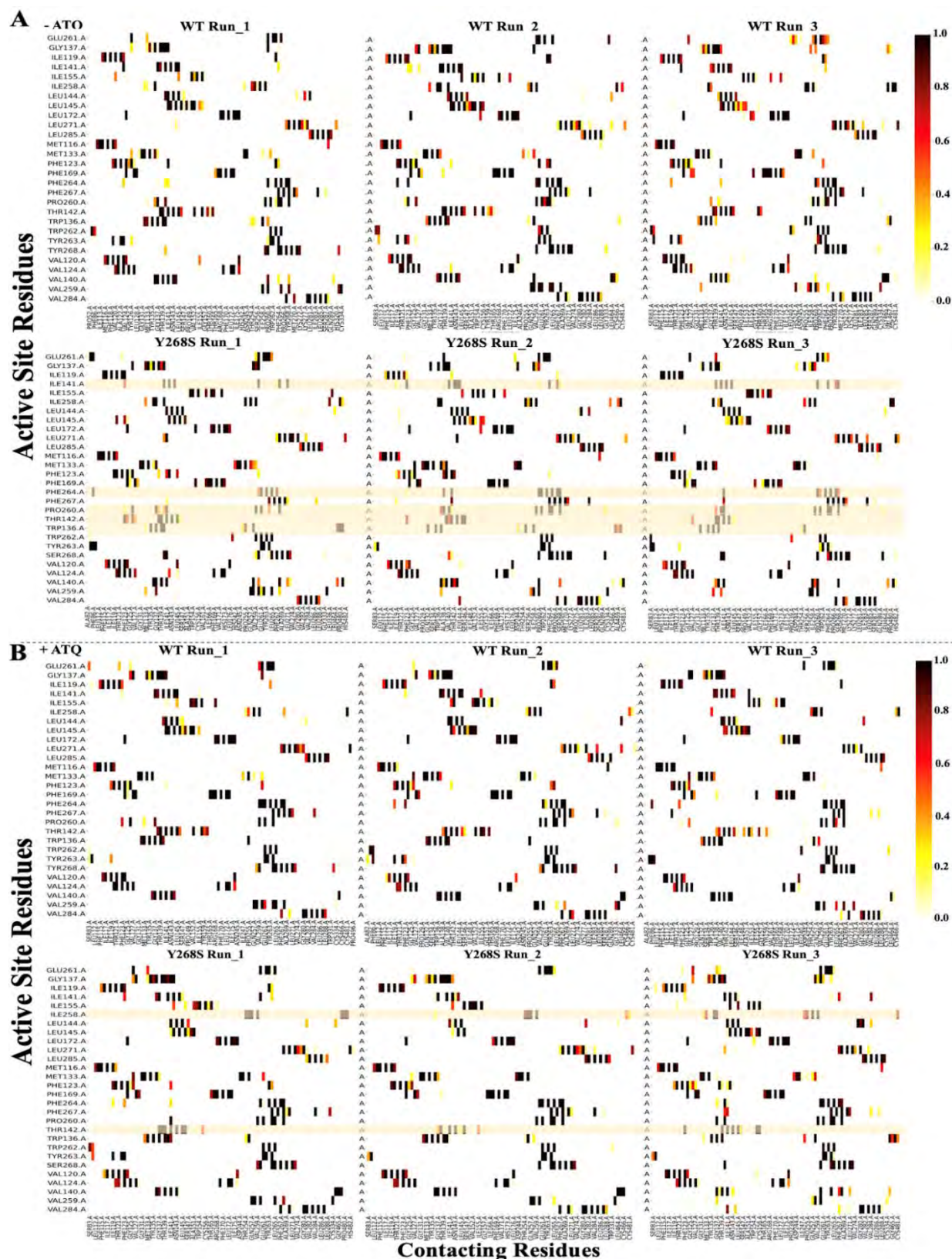


Figure S5.6: Residue contact heat map between active site residues ($n = 28$) and other residue contacts. Comparison of the WT and mutant system Y268S was performed in both holo and ATQ-bound proteins in their respective triplicate runs in the last 100 ns of MD simulation time. Regions observed to reduce or lose residue contact are shaded in light golden colour. Reproduced from [141].

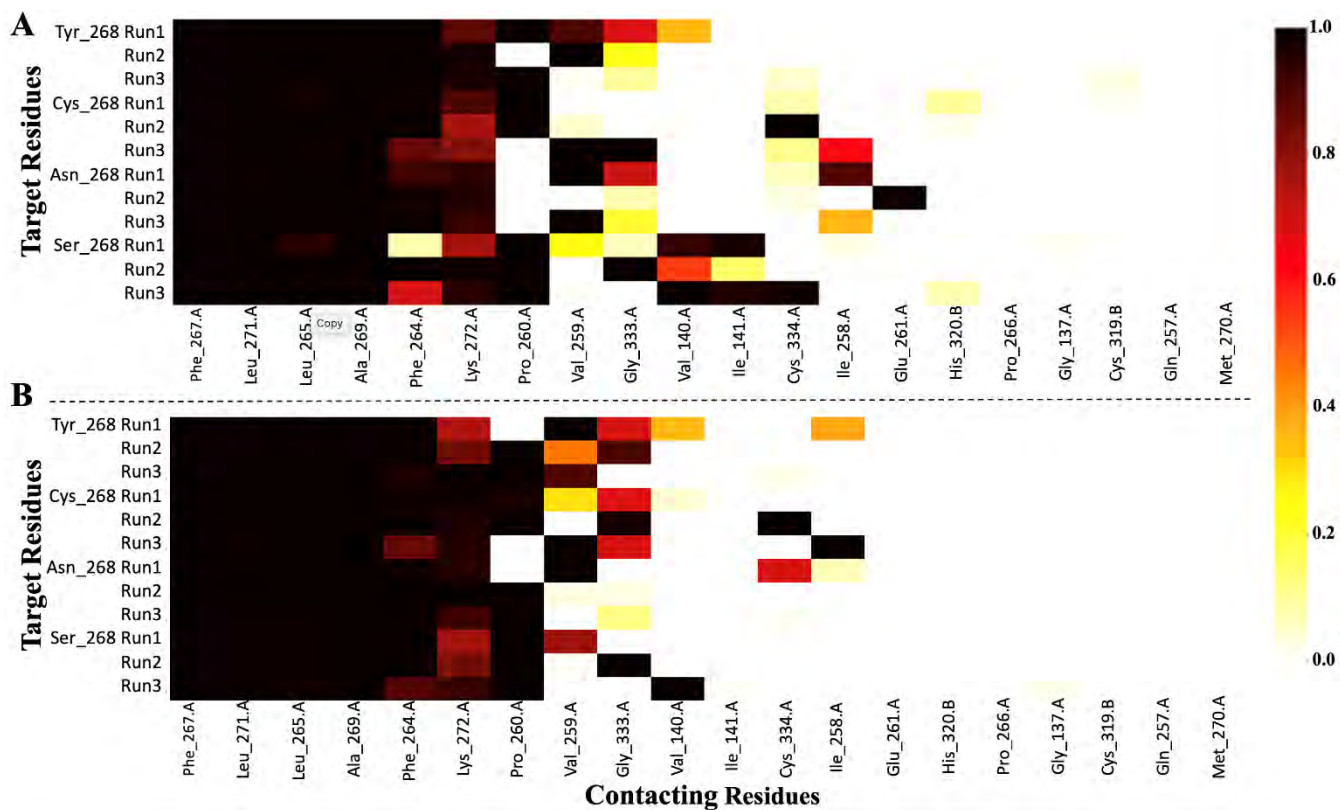


Figure S5.7: Residue contact heat map of Y268C, Y268N, Y268S mutant systems with reference to wild type yielded during the last 100 ns of MD simulations for all systems in the triplicate runs. (A) PfCytb-ISP holo protein (B) PfCytb-ISP ATQ-bound protein. Individual residue contact maps for WT and all mutant systems in triplicate runs are displayed in Figure S5.4 – 5.6. Residues ordered by the proportion of contacts across simulation time. Residue names ending with A or B represent chain A and B, respectively. Reproduced from [141].

Table S5.1: Protein residues having high normalized *Betweenness Centrality* (*BC*) values in holo WT and three mutant systems across the triplicate runs; + sign indicate significant high values. Active site residues are shown in bold. The * and # suffixes indicate common residues across 3 runs and 3 runs, respectively. (A threshold value of ± 2 standard deviations (SD) of the *BC* values were used for each system. Reproduced from [141].

<i>Betweenness Centrality</i> (<i>BC</i>); Significant Residues			
Protein Systems	RUN_1	RUN_2	RUN_3
WT	PfCytb Chain: Ser_83, Ile_93, Ala_122, Phe_123* , Ile_141 , Asn_143, Trp_262# , Leu_265, Tyr_268 , Lys_272, Leu_290, Leu_291 ISP Chain: Gly_297, Ile_298, His_301* , Leu_302, His_320# , Ser_322, Ile_330	PfCytb Chain: Val_85, Thr_89, Met_102, Phe_115, Phe_118, Ile_119 , Ala_122, Val_124 , Ile_258 , Val_259* , Trp_262# , Phe_264* , Leu_265, Lys_272, Leu_291 ISP Chain: Gly_297, Ile_298, Gly_303, His_320# , Ser_322, Ile_330	PfCytb Chain: Ser_83, Val_85, Phe_86, Thr_89, Leu_94, Phe_118, Ala_122, Phe_123* , Thr_139, Ile_141 , Val_259* , Trp_262# , Phe_264* , Leu_265, Leu_271 , Lys_272, Lys_277, Leu_290, Leu_294 ISP Chain: Gly_297, Ile_298, His_301* , Val_305, His_320# , Ser_322, Ile_330
	PfCytb Chain: Thr_80, Val_85, Phe_86, Thr_89, Leu_112, Phe_118, Ala_122, Phe_123 , Ser_134, Thr_139, Ile_141 , Thr_160, Val_161, Glu_261 , Trp_262 , Leu_265, Cys_268* , Leu_271# , Lys_272 ISP Chain: Lys_252, Cys_304, His_320 , Ser_322	PfCytb Chain: Val_85, Phe_86, Thr_89, Ile_117, Phe_118, Val_120, Ala_122, Phe_123 , Ala_138, Thr_139, Val_140 , Ile_141 , Tyr_263* , Phe_264* , Met_270, Leu_271# , Lys_272, Leu_281, Leu_290, Ile_320 ISP Chain: Gly_297, Ile_298, Gly_303, Cys_304, His_320 , Ser_322, Ile_330	PfCytb Chain: Val_85, Phe_86, Thr_89, Leu_112, Phe_118, Thr_121, Val_124 , Thr_160, Thr_165, Ile_258 , Pro_260 , Glu_261 , Tyr_263* , Phe_264* , Cys_268* , Leu_271# , Lys_272, Lys_277, Ile_349 ISP Chain: Lys_252, His_301, Leu_302, His_320 , Ser_322
Y268C	PfCytb Chain: Val_85, Phe_86, Thr_89, Ile_93, Val_124 , Phe_169, Ser_241, His_242, Thr_254, Tyr_263# , Phe_264# , Asn_268#, Lys_272, Lys_277, Leu_288, Leu_291 ISP Chain: Gly_297, Ile_298, His_301* , Leu_302, Gly_303, His_320# , Ser_322, Ile_330	PfCytb Chain: Val_85, Phe_86, Thr_89, Leu_112, Ile_117, Phe_118, Ile_119 , Gly_137 , Thr_160, Leu_172, Leu_176, His_242, Thr_254, Pro_260* , Glu_261* , Tyr_263# , Phe_264# , Asn_268#, Lys_272, ISP Chain: Cys_299*, Gly_303, His_320# , Ser_322, Ile_330, Pro_334	PfCytb Chain: Ser_83, Val_85, Phe_86, Leu_112, Ile_117, Met_133 , Thr_139, Leu_172, Leu_176, Pro_260* , Glu_261* , Tyr_263# , Phe_264# , Asn_268#, Leu_271 , Lys_272, Lys_277, Leu_288, Leu_291 ISP Chain: Gly_297, Ile_298, Cys_299* , His_301* , Gly_303, His_320# , Ser_322, Pro_334
	PfCytb Chain: Ile_35, Val_85, Phe_86, Thr_89, Trp_108, Ile_117, Val_120 , Phe_123 , Val_140 , Ile_141 , Thr_142, Leu_145, Thr_160, Val_259 , Pro_260 , Glu_261 , Leu_265, Leu_271 ISP Chain: Lys_252, Ile_298, Cys_304, Pro_318, His_320# , Ser_322, His_323	PfCytb Chain: Leu_31, Val_85, Phe_86, Thr_89, Leu_112, Ile_117, Phe_118, Thr_121, Ala_122, Phe_123 , Val_124 , Val_127, Trp_136 , Gly_137 , Thr_160, Ser_241, Tyr_263* , Phe_264* , Lys_272, Ile_320 ISP Chain: Gly_297, Ile_298, Val_305, Ala_307, His_320# , Ser_322, Ile_330	PfCytb Chain: Ser_83, Phe_86, Thr_89, Arg_95, Phe_118, Thr_121, Ala_138, Thr_139, Val_140 , Ile_141 , Phe_169 , Tyr_263* , Phe_264* , Ser_268 , Lys_272, Ile_349 ISP Chain: Ile_298, Cys_299 , Cys_304, Val_305, His_320# , Gly_321, Ser_322, Ile_330, Pro_334
Y268S	PfCytb Chain: Ile_35, Val_85, Phe_86, Thr_89, Trp_108, Ile_117, Val_120 , Phe_123 , Val_140 , Ile_141 , Thr_142, Leu_145, Thr_160, Val_259 , Pro_260 , Glu_261 , Leu_265, Leu_271 ISP Chain: Lys_252, Ile_298, Cys_304, Pro_318, His_320# , Ser_322, His_323	PfCytb Chain: Leu_31, Val_85, Phe_86, Thr_89, Leu_112, Ile_117, Phe_118, Thr_121, Ala_122, Phe_123 , Val_124 , Val_127, Trp_136 , Gly_137 , Thr_160, Ser_241, Tyr_263* , Phe_264* , Lys_272, Ile_320 ISP Chain: Gly_297, Ile_298, Val_305, Ala_307, His_320# , Ser_322, Ile_330	PfCytb Chain: Ser_83, Phe_86, Thr_89, Arg_95, Phe_118, Thr_121, Ala_138, Thr_139, Val_140 , Ile_141 , Phe_169 , Tyr_263* , Phe_264* , Ser_268 , Lys_272, Ile_349 ISP Chain: Ile_298, Cys_299 , Cys_304, Val_305, His_320# , Gly_321, Ser_322, Ile_330, Pro_334

Table S5.2: Residues having high *Betweenness Centrality* (BC) values in ATQ-bound protein systems (WT, Y268C, Y268N and Y268S) to assess effect of mutation; + significant positive values while – stands for significant negative values. Active site residues are shown in bold, common residues in all systems (Bold#) and common residues in at least two runs (Bold*). A threshold value of +/- 2 SD of the BC values was used for each system. Reproduced from [141].

<i>Betweenness Centrality</i> (BC); Significant Residues				
Mutant systems		RUN_1	RUN_2	RUN_3
WT	+	PfCytb Chain: Ser_83, Val_85, Phe_86, Thr_89, Leu_112, Ile_117, Phe_118, Ala_122, Tyr_126 , Val_127, Gly_137, Thr_139, Asn_250, Val_259*, Pro_260 , Glu_261 , Tyr_263 , Phe_264* , Leu_271 , Lys_277, Gly_280, Leu_291 ISP Chain: Gly_297, Ile_298, Cys_299* , His_301* , Gly_303, His_320* , Ile_330, Ala_335	PfCytb Chain: Ser_83, Val_85, Thr_89, Ala_122, Phe_123 , Val_127, Thr_139, Thr_160, Thr_165, Phe_169* , Trp_262 , Leu_265, Ala_269, Met_270, Leu_271 , Lys_272, Cys_324 ISP Chain: Lys_252, Ile_296, Gly_297, Ile_298, Leu_302, His_320* , Ser_322	PfCytb Chain: Val_85, Leu_88, Arg_95, Thr_121, Val_124 , Val_127, Phe_169* , Ile_237, Ser_241, His_242, Thr_254, Ile_258 , Val_259* , Phe_264* , Leu_265, Tyr_268 , Val_274, Lys_277, Gly_280, Leu_281, Val_284, Ser_287, Leu_290, Leu_291 ISP Chain: Gly_297, Cys_299* , His_301* , Gly_303, Val_305, Ile_330, Pro_334, Ala_335
Y268C	+	PfCytb Chain: Ser_83, Val_85, Thr_89, Tyr_99, Leu_112, Phe_118, Ala_122, Phe_123* , Val_124* , Thr_139, Ile_141 , Ser_162, Tyr_263* , Phe_264# , Leu_271* , Lys_272, Lys_277, Val_284 ISP Chain: Leu_226, Ile_298, His_301# , Leu_302, Gly_303, His_320# , Ser_322, Ile_330	PfCytb Chain: Ala_82, Val_85, Arg_95, Phe_118, Thr_121, Ala_122, Phe_123* , Val_124* , Val_127, Ala_138, Thr_160, Thr_165, Arg_168, Phe_169 , Leu_172 , Leu_176, Phe_264# , Leu_271* , Lys_272, Lys_277, Leu_281, Ile_320 ISP Chain: Lys_252, His_301# , Leu_302, Gly_303, His_320# , Ser_322, Ile_330	PfCytb Chain: Ser_83, Thr_89, Leu_112, Phe_118, Thr_121, Ala_122, Tyr_126, Val_127, Met_133 , Gly_137 , Thr_139, Pro_255, Pro_260 , Tyr_263* , Phe_264# , Cys_268 , Lys_272, Lys_277, Leu_288 ISP Chain: Gly_297, Ile_298, His_301# , Gly_303, Val_305, His_320# , Ser_322, Ile_330
Y268N	+	PfCytb Chain: Ser_83, Val_85, Phe_86, Thr_89, Leu_112, Val_127, Met_133* , Tyr_252, Thr_254, Ile_258 , Val_259 , Pro_260 , Tyr_263* , Phe_264* , Asn_268* , Lys_272, Lys_277 ISP Chain: Gly_297, His_301* , Leu_302, Val_305, His_320# , Ser_322, Ile_330	PfCytb Chain: Val_85, Phe_86, Thr_89, Leu_112, Phe_118, Ile_119 , Ala_122, Ser_134, Asn_143, Val_249, Tyr_263* , Phe_264* , Asn_268* , Ala_269, Lys_272, Lys_277 ISP Chain: Ile_298, His_301* , Leu_302, Gly_303, His_320# , Ser_322, Ile_330	PfCytb Chain: Leu_112, Ala_122, Phe_123 , Val_124 , Tyr_126, Val_127, Met_133* , Ser_134, Trp_136 , Thr_139, Ile_237, Ser_241, His_242, Phe_264, Met_270, Leu_271 , Lys_272, Leu_281 ISP Chain: Gly_297, Ile_298, Gly_303, His_320# , Ser_322, Ile_330, Pro_334, Ala_335
Y268S	+	PfCytb Chain: Ser_83, Val_85, Phe_86, Thr_89, Ile_93, Leu_97, Leu_112, Ala_122, Phe_123* , Val_140* , Ile_141 , Ile_258 , Trp_262 , Phe_264# , Leu_265, Ser_268* , Lys_272, Leu_288 ISP Chain: Gly_297, Ile_298, His_301* , Gly_303, Val_305, His_320# , Ser_322	PfCytb Chain: Val_85, Phe_86, Thr_89, Arg_95, Phe_118, Thr_121, Val_124 , Val_127, Gly_137* , Val_140* , Thr_160, Val_161, Thr_165, Phe_169 , Tyr_263* , Phe_264# , Ser_268* , Leu_271 , Lys_272, Lys_277, Leu_281, Gln_289, Leu_290, Phe_292 ISP Chain: Lys_252, His_301* , Leu_302, His_320# , Ser_322, Ile_330	PfCytb Chain: Ala_82, Val_85, Phe_86, Thr_89, Leu_112, Ala_122, Phe_123* , Trp_136 , Gly_137* , Thr_142, Asn_143, Thr_160, Thr_165, Tyr_263* , Phe_264# , Lys_272, Leu_290, Leu_294 ISP Chain: Gly_297, Ile_298, Leu_302, Val_305, His_320# , Pro_336

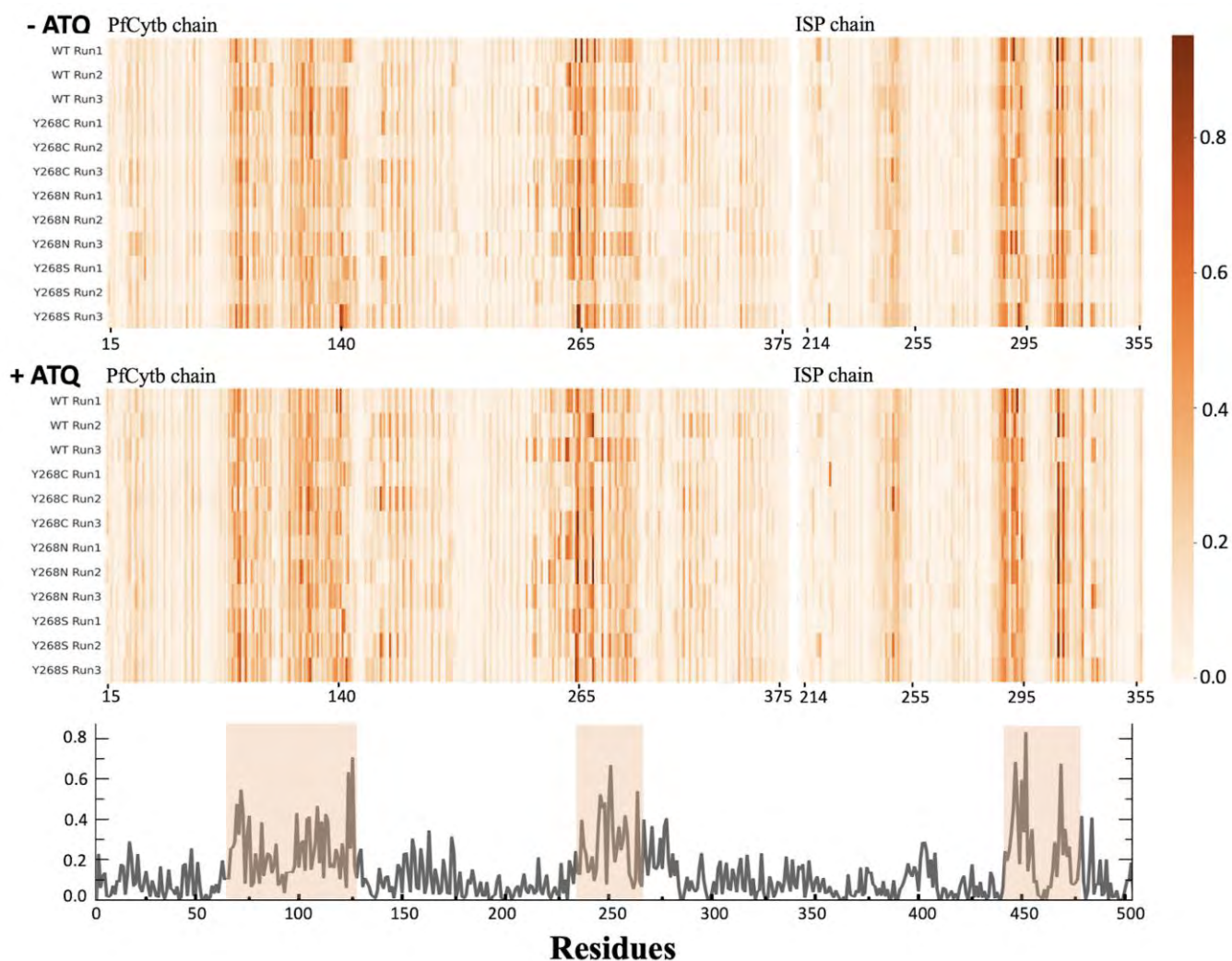


Figure S5.8: Heat map comparison of *BC* values in PfCytb-ISP holo and ATQ-bound proteins. High *BC* values indicate regions of high residue connectivity. Reproduced from [141].

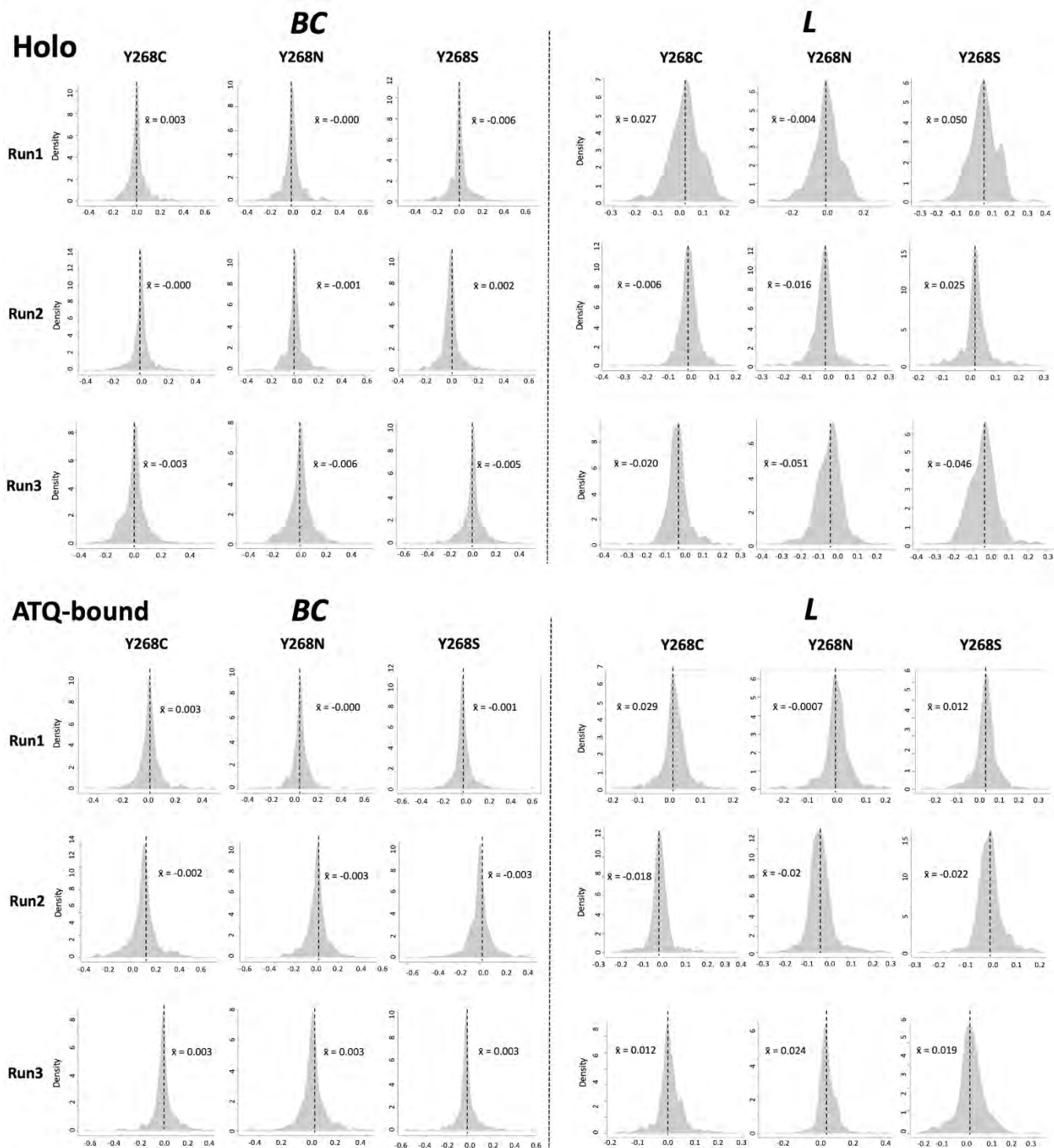


Figure S5.9: Density plot for delta *BC* and *L* values (N=503) for mutant systems (Y268C, Y268N, and Y268S) in three replicate runs. The plots show that (A) *BC* and (B) *L* data was normally distributed where the mean is displayed by a dotted line. Reproduced from [141].

Table S5.3: Table of residues having significant changes in *Betweenness Centrality* (ΔBC) in three holo mutant systems to assess effect of mutation; + significant positive values while – stands for significant negative values. Active site residues are shown in bold, common residues in all systems (Bold#) and common residues in at least two runs (Bold*). A threshold value of +/- 2 SD of the ΔBC values was used for each system. Reproduced from [141].

<i>Betweenness Centrality</i> (ΔBC); (WT– Mutant); Significant Residues				
Mutant systems		RUN_1	RUN_2	RUN_3
Y268C	+	PfCytb chain: Ser_83, Ile_93, Asn_143, Trp_262# , Leu_265, Cys_268 , Lys_272, Val_274, Ile_283, Leu_288 ISP chain: Trp_221, Ile_222, Ile_298, Leu_302, His_320, Ser_322, Ile_330	PfCytb chain: Trp_26, Tyr_101, Met_102, Phe_115, Ile_119* , Val_124 , Thr_160, Gln_257, Ile_258 , Val_259 , Trp_262# , Leu_265, Leu_291 ISP chain: Val_227	PfCytb chain: Leu_25, Leu_94, Ile_119* , Ala_122, Phe_123 , Trp_136 , Ile_141 , Ile_237, Ser_241, Trp_262# , Leu_265, Thr_301, Trp_331, Val_305
	-	PfCytb chain: Phe_86, Met_102, Phe_118, Ala_122, Phe_123* , Thr_160, Val_161, His_187, Gln_232, Glu_261	PfCytb chain: Leu_94, Val_120 , Phe_123* , Ala_138, Thr_139, Val_140 , Ile_141 , Ser_241, Val_249, Tyr_263, Gln_289, Leu_290, Cys_324, Leu_357 ISP chain: Ser_223, Cys_304, Pro_334	PfCytb chain: Tyr_99, Leu_112, Ile_155 , Thr_160, Thr_165, Arg_168, Leu_240, Asp_244, Tyr_252, Gln_257, Phe_264 , Lys_252, Leu_302
Y268N	+	PfCytb chain: Ser_83, Ala_122, Phe_123* , Val_140* , Ile_141* , Asn_143, Leu_144 , Trp_262# , Leu_265, Asn_268 , Lys_272, Ile_309 ISP chain: Trp_221, Ile_222, Leu_302	PfCytb chain: Trp_26, Phe_118, Ala_122, Thr_197, Ile_247, Thr_251, Gln_257, Ile_258 , Val_259 , Trp_262# , Lys_277, Leu_281, Leu_291 ISP chain: Ile_222, Ile_298, Cys_304, Pro_318	PfCytb chain: Phe_118, Ala_122, Phe_123* , Trp_136 , Val_140* , Ile_141* , Thr_160, Ser_162, Ile_237, Trp_262# , Leu_265, Leu_294, Val_305
	-	PfCytb chain: Phe_86, Val_124 , Thr_160, Phe_169, Ile_237, Ser_241, His_242, Asn_250, Tyr_263, Phe_264# , Lys_277, Leu_357	PfCytb chain: Ser_241, His_242, Val_253, Thr_254, Pro_260* , Tyr_263, Phe_264# , Asn_268 , Ile_342, Pro_334, Pro_336	PfCytb chain: Phe_30, His_92, Leu_112, Ile_155, Leu_172, Leu_176, Leu_215, Tyr_252, Pro_260* , Phe_264# , Val_274, Leu_288, Gly_303, Pro_334, Pro_336
Y268S	+	PfCytb chain: Thr_80, Ser_83, Asn_143, Tyr_252, Trp_262# , Leu_265, Ser_268 , Lys_272, Leu_281, Leu_288, Leu_290, Leu_291 ISP chain: Ile_222, Leu_302, His_320*, Ile_330	PfCytb chain: Tyr_101, Met_102, Phe_115, Ile_119* , Thr_197, Thr_251, Gln_257, Ile_258 , Val_259* , Glu_261 , Trp_262# , Phe_264 , Lys_277, Leu_291, Leu_294, Cys_304	PfCytb chain: Leu_25, Ala_82, Ile_119* , Ala_122, Phe_123 , Gly_131, Trp_136 , Ile_237, Asn_250, Val_259* , Trp_262# , Leu_265 ISP chain: Ile_222, His_320*
	-	PfCytb chain: Ile_35, Val_85, Phe_86, Thr_89, Met_102, Trp_108, Val_120 , Leu_145 , Ser_146, Tyr_159, Thr_160, Pro_260 , Glu_261 , Leu_271 , Cys_304, Pro_318	PfCytb chain: Tyr_16, Val_127, Trp_136 , Asn_143, Ser_146, Ile_188, Ser_241, His_242, Tyr_263* , Cys_324, Cys_319*	PfCytb chain: Ser_83, Thr_121, Ala_138, Thr_139, Ile_155 , Tyr_263* , Phe_264 , Phe_267 , Ser_268 , Val_274 ISP chain: Leu_226, Cys_304, Cys_319*, Gly_321, Pro_334

Table S5.4: Table of residues having significant changes in *Betweenness Centrality* (ΔBC) in three ATQ-bound mutant systems to assess effect of mutation; + significant positive values while – stands for significant negative values. Active site residues are shown in bold, common residues in all systems (Bold#) and common residues in at least two runs (Bold*). A threshold value of +/- 2 SD of the ΔBC values was used for each system. Reproduced from [141].

<i>Betweenness Centrality</i> (ΔBC); (WT– Mutant); Significant Residues				
Mutant systems		RUN_1	RUN_2	RUN_3
Y268C	+	PfCytb Chain: Phe_86, Arg_95, Val_127, Gly_137 , Thr_139, Pro_164, Arg_168, His_187 , Val_259* , Pro_260 , Glu_261 , Phe_264 , Lys_277, Gly_280, Leu_291 ISP Chain: Ile_298, Cys_299 , His_301 , Gly_303, Cys_319 , Ala_335	PfCytb Chain: Leu_25, Ser_83, Phe_86, Tyr_90, Thr_139, Ile_141 , Cys_156, Pro_164, Leu_191, His_242, Asp_244, Trp_262 , Leu_265, Ala_269, Met_270, Lys_272, Cys_324, Tyr_363 ISP Chain: Trp_221, Ile_298	PfCytb Chain: Ile_155 , Thr_160, Phe_169 , Ile_237, Ser_241, His_242, Thr_254, Ile_258 , Val_259* , Leu_265, Thr_273, Val_274, Gly_280, Leu_281, Ser_287, Pro_334, Ala_335
	-	PfCytb Chain: Tyr_99, Met_116 , Phe_123 , Trp_136 , Ile_141 , Leu_144 , Ser_162, Ser_196, Ser_241, Ser_287 ISP Chain: Leu_226, Lys_252, Leu_302, Ser_322	PfCytb Chain: Asn_27, Leu_31, Arg_95, Met_102, Phe_118, Ala_122, Thr_160, Leu_172 , Ala_246, Phe_264* , Lys_277, Leu_281, Leu_285 , Gln_289, Ile_320 ISP Chain: Phe_219, Gly_251, Lys_252, His_301	PfCytb Chain: Ser_83, Thr_89, Ala_122, Phe_264* , Leu_271, Lys_272, Tyr_327, Pro_337 ISP Chain: His_320 , Gly_321, Ser_322
Y268N	+	PfCytb Chain: Tyr_16, Ala_122, Gly_137 , Thr_139, Ile_247, Leu_271, Pro_337, Ile_298, Gly_303, Ala_335	PfCytb Chain: Ser_83, Thr_139, Thr_160, Ser_162, Pro_164, Thr_165, Trp_262 , Leu_265, Met_270, Leu_271, Cys_324, Gly_333 ISP Chain: Trp_221	PfCytb Chain: Tyr_28, Val_85, Leu_88, Thr_160, Phe_169 , Asn_245, Ile_258 , Val_259 , Val_274, Pro_275, Ser_276, Lys_277, Gly_280, Ser_287, Leu_294, Ile_309 ISP Chain: His_301 , Val_305
	-	PfCytb Chain: Leu_94, Met_133 , Ser_241, His_242, Tyr_252, Thr_254, Ile_258 , Lys_272 ISP Chain: Val_305, His_320* , Ser_322	PfCytb Chain: Phe_86, His_92 , Phe_118, Gly_131, His_192* , Asn_245, Val_249, Tyr_252, Tyr_263 , Phe_264 , Tyr_268 , Pro_278 ISP Chain: His_301 , Gly_303	PfCytb Chain: Ala_122, Trp_136 , Ser_147, Leu_172 , His_187 , His_192* , Leu_271 , Lys_272 ISP Chain: Cys_324, His_320* , Ser_322, Pro_334, Pro_336
Y268S	+	PfCytb Chain: Arg_95, Gly_137 , Thr_139, Ile_247, Pro_260 , Phe_264 , Lys_277, Leu_291 ISP Chain: Ile_298, His_301* , Gly_303, Cys_319 , Ala_335	PfCytb Chain: Leu_25, Ala_122, Phe_123 , Thr_139, Pro_164, Ile_188, Trp_262 , Leu_265, Ala_269, Lys_272, Leu_288, Cys_324, Ala_328 ISP Chain: Ile_298, Leu_302	PfCytb Chain: Val_124 , Val_127, Phe_169 , Ile_237, Ser_241, His_242, Thr_254, Pro_255, Ile_258 , Val_259 , Leu_265, Val_274, Lys_277, Gly_280, Leu_291 ISP Chain: His_301*
	-	PfCytb Chain: Thr_89, Ile_93, Leu_97, Phe_123* , Met_133 , Val_140 , Ile_141 , Thr_160, Ser_241, Trp_262 , Leu_265, Cys_304, Ser_322	PfCytb Chain: Phe_86, Arg_95, Leu_97, Met_102, Gly_137 , Val_140 , Leu_144 , Val_161, Ser_241, Tyr_263 , Phe_264 , Lys_277, Leu_281, Leu_285, Gln_289, Leu_290, Phe_292 ISP Chain: His_301	PfCytb Chain: Phe_86, Thr_89, Ala_122, Phe_123* , Trp_136 , Thr_142 , Asn_143, Gly_195, Tyr_263* , Lys_272 ISP Chain: Tyr_327, Pro_337, is_320 , Pro_336

Table S5.5: List of a total of 38 inhibitors (MMV and SANCDB) docked into the Qo active site; 29 selective compounds unique to Plasmodium and not binding in human active site, Out of these, 21 compounds excluded from analysis as they bound in the heterodimer interface and were unstable in during MD simulations (WT system)

Compound name (MMV)	Binding energies	Compound name (SANCDB)	Binding energies
1. MMV007638	-9.5	1. SANC00409	-7.8; Excluded (Interface)
2. MMV019551	-8.9	2. SANC00373	-9.7
3. MMV659004	-9.9	3. SANC00459	-9.3
4. MMV676444	-8.8	4. SANC00548	-8.9
5. MMV676162	-9.8; Excluded (MD Unstable)	5. SANC00725	-8.3
6. MMV676528	-10.9; Excluded (MD Unstable)	6. SANC00518	-11.4; Excluded (MD Unstable)
7. MMV688798	-9.3; Excluded (MD Unstable)	7. SANC00531	-8.6; Excluded (MD Unstable)
8. MMV676409	-9.8; Excluded (Interface)	8. SANC00636	-8.1; Excluded (MD Unstable)
9. MMV010545	-9.5; Excluded (MD Unstable)	9. SANC00570	-8.0; Excluded (MD Unstable)
10. MMV023233	-9.2; Excluded (MD Unstable)	10. SANC00573	-9.0; Excluded (MD Unstable)
11. MMV019189	-9.7; Excluded (Interface)	11. SANC00576	-8.5; Excluded (MD Unstable)
12. MMV023949	-9.0; Excluded (Interface)	12. SANC00563	-9.0; Excluded (MD Unstable)
13. MMV687189	-8.5; Excluded (MD Unstable)	13. SANC00713	-8.9; Excluded (MD Unstable)
14. MMV676008	-10.2; Excluded (Interface)	14. SANC00208	-8.7; Excluded (MD Unstable)
15. MMV024443	-9.8; Excluded (MD Unstable)	15. SANC00638	-8.2; Excluded (MD Unstable)
16. MMV688372	-11.8; Excluded (Interface)	16. SANC00649	-8.2; Excluded (MD Unstable)
17. MMV676477	-9.0; Excluded (Interface)	17. SANC00704	-8.2; Excluded (MD Unstable)
18. MMV687765	-9.9; Excluded (Interface)	18. SANC00329	-8.1; Excluded (MD Unstable)
19. MMV688179	-11.5; Excluded (Interface)	19. SANC00209	-9.4; Excluded (MD Unstable)

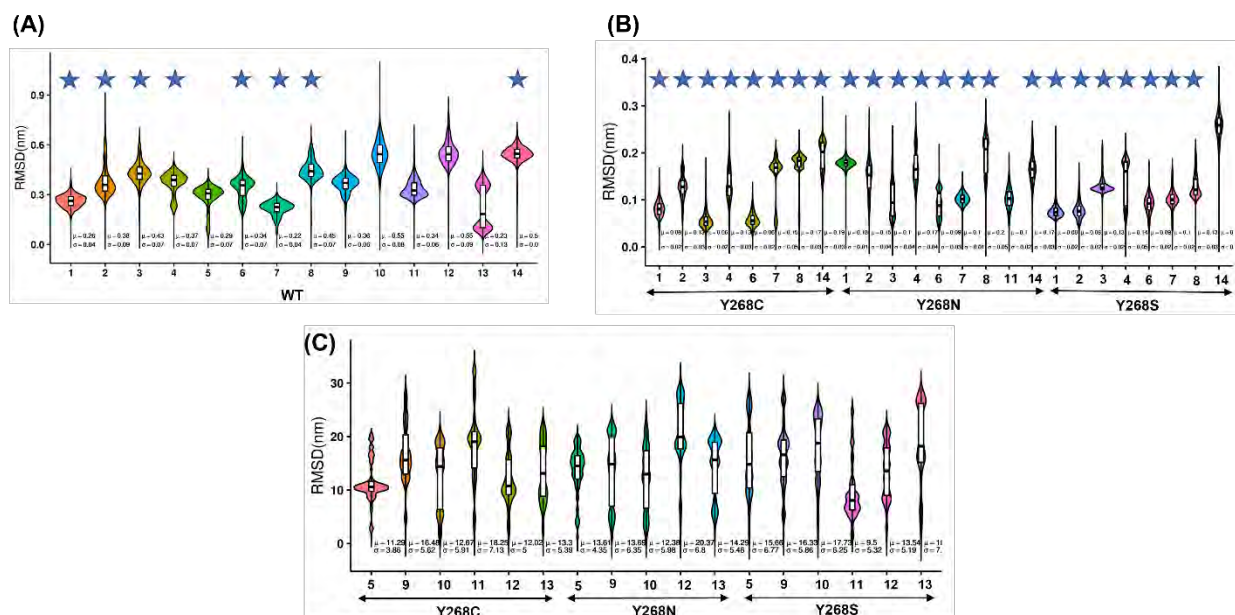


Figure S5.11: A violin plot representation of ligand RMSD for 14 hit compounds (A) All the ligands in WT showing normal RMSDs (B) The ligands binding to the different mutants with normal RMSDs (C) Ligands binding to the mutants with high RMSDs, regarded as unstable. A total of 8 ligands (1-4 (MMV007638, MMV019551, MMV659004, MMV676444), 6-8 (SANC00373, SANC00459, SANC00548) and 14 (SANC00725) were stable in WT and mutant systems (Y268C, Y268N and Y268S); represented in blue stars.

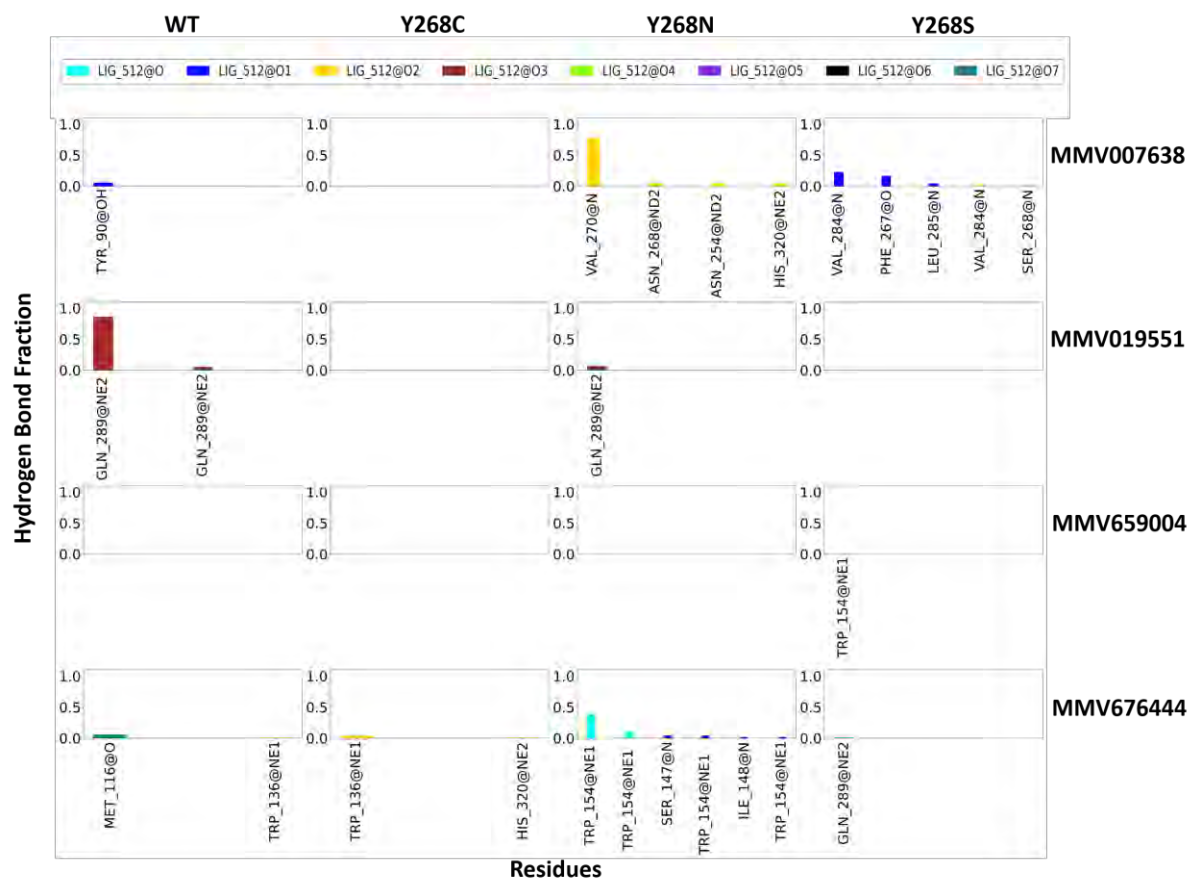


Figure S5.12: Analysis of average number of intermolecular hydrogen bonds formed between the MMV potential hits and PfCytb-ISP protein in four protein systems (WT, Y268C, Y268N and Y268S) during 350 ns molecular dynamics simulations. LIG512 represent colour code in each of the 4 ligand systems. Consider a table rather than this figure for MMV compounds.

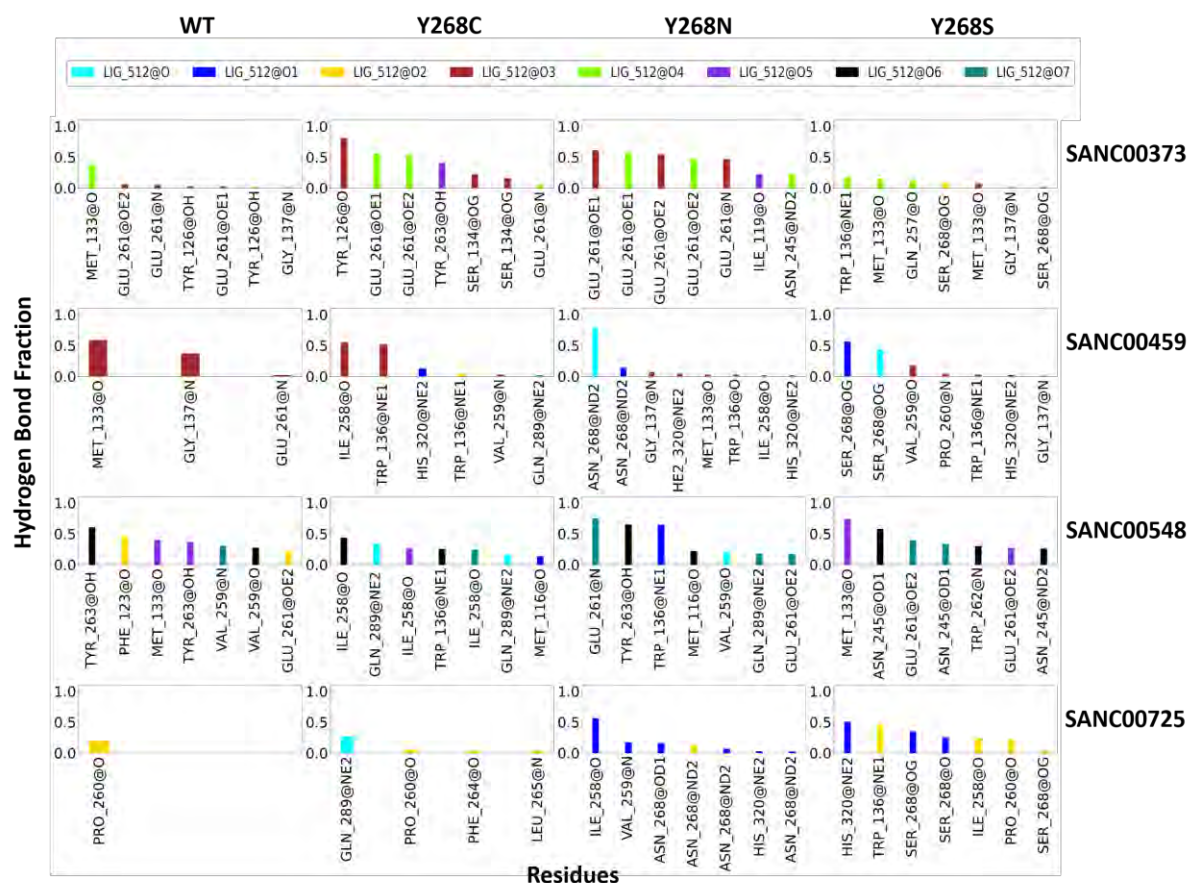


Figure S5.13: Analysis of average number of intermolecular hydrogen bonds formed between the SANCDB potential hits and PfCytb-ISP protein in four protein systems (WT, Y268C, Y268N and Y268S) during 350 ns molecular dynamics simulations. LIG512 represent colour code in each of the 4 ligand systems.

Table S5.6: Hydrogen bond (Hbond) and van der Waals (vdW) molecular interactions formed in eight protein-ligand complexes (Lig1; MMV007638, Lig2; MMV019551, Lig3; MMV659004, Lig4; MMV676444, Lig5; SANC00373, Lig6; SANC00459, Lig7; SANC00548, Lig8; SANC00725) in WT, 268C, Y268N and Y268S protein systems 350 ns MD simulation (25 ns timestep).

			DRUG COMPOUNDS							
Systems	Time step (ns)		Lig1	Lig2	Lig3	Lig4	Lig5	Lig6	Lig7	Lig8
WT	0	VdW Hbond	14	6 1 (Gln289)	11 1 (Leu144)	13	10 1 (Glu261)	14 1 (Met133)	12 1 (Glu261)	9 (Leu281)
	25	VdW Hbond	16	6 1 (Gln289)	14	10 1 (Met116)	12 1 (Met133)	14 1 (Met133)	12 5 (Val259, Ser134, Tyr263, Phe137)	13
	50	VdW Hbond	14	8 1 (Gln289)	8	11	14 1 (Met133)	10 1 (Met133)	13 3 (Tyr263, Phe137, Val259)	9
	75	VdW Hbond	13	5 1 (Gln289)	11	14	8 1 (Met133)	14	11 4 (Glu261, Val259, Phe137, Tyr263)	10
	100	VdW Hbond	16 1 (Tyr263)	10 1 (Gln289)	13	15	7 1 (Met133)	15 1 (Met133)	12 4 (Glu261, Tyr263, Met133, Met133)	14 1 (Pro260)
	125	VdW Hbond	16	8 1 (Gln289)	11	11	8	15 1 (Met133)	9 5 (Tyr263, Glu261, Met133, Val259, Met116)	14
	150	VdW Hbond	13	9 1 (Gln289)	10	13	10 1 (Met133)	16 1 (Met133)	13 3 (Met133, Tyr263, Val259)	9
	175	VdW Hbond	16 1 (Phe264)	6	7	12	8 3 (Met133, Trp136, Val259)	13 1 (Met133)	13 4 (Met133, Gly137, Val259, Met116)	11
	200	VdW Hbond	16	1 1 (Gln289)	9 1 (Leu144)	11	9 1 (Met133)	13 1 (Met133)	12 4 (Glu261, Tyr263, Met133, Val259)	11
	225	VdW Hbond	17	5 1 (Gln275)	10	9	8	13 1 (Met133)	10 3 (Tyr263, Met133, Val259)	9
	250	VdW Hbond	16	4 1 (Gln275)	9	10	9	13 1 (Met133)	7 4 (Tyr263, Met133, Val259, Met116)	10
	275	VdW Hbond	15	7 1 (Met102)	9	12	10	11 1 (Met133)	8 4 (Glu261, Tyr263, Met133, Val259)	10
	300	VdW Hbond	13	6 1 (Gln275)	10	10	8 1 (Val245)	13 1 (Met133)	8 5 (Glu261, Tyr263, Met133, Met116, Val259)	13

	325	VdW Hbond	17	8 1 (Gln275)	9	13	10	9 1 (Met133)	11 4 (Glu261, Tyr263, Met133, Val259)	14 1 (Pro246)
	350	VdW Hbond	15	5 1 (Gln275)	9	13	8 1 (Met133)	13 1 (Met133)	10 5 (Glu261, Tyr263, Met133, Met116, Val259)	15
Y268C	0	VdW Hbond	15	8	14	16	12 2 (Phe137, Glu261)	12 2 (Gln289, Cys268)	11 2 (Gln289, Tyr263)	8 1 (Pro260)
	25	VdW Hbond	18	9	12	13	12 1 (Tyr263)	15 1 (Cys268)	10 3 (Gln289, Ile258, Trp136)	9 1 (Phe264)
	50	VdW Hbond	19	6	12	13	11 4 (Glu261, Ser134, Tyr126, Tyr263)	13 1 (Cys268)	10 1 (Trp136)	10
	75	VdW Hbond	11	6	10	10 1 (Cys268)	12 2 (Ser134, Glu261)	15 1 (Trp136)	9 3 (Gln289, Ile258, Trp136)	3 2 (Cys268, Phe264)
	100	VdW Hbond	14	6	9	13	11 3 (Tyr263, Glu261, Ser134)	19 (Trp136, Ile258)	8 2 (Gln289, Ile258)	8
	125	VdW Hbond	10	4	10	13 1 (Trp136)	11 4 (Tyr263, Glu261, Tyr126, Ser134)	12 3 (Trp136, Ile258, Cys268)	8 2 (Gln289, Ile258)	10 1 (Phe264)
	150	VdW Hbond	12	6	10	14	14 3 (Ser134, Tyr126, Tyr263)	16 3 (Trp136, Ile258, Cys268)	9 (Gln289, Ile258)	7 1 (Cys268)
	175	VdW Hbond	15	8	8	13	14 2 (Tyr263, Tyr126)	16 2 (Trp136, Ile258)	14 1 (Gln289)	10 1 (Gln289)
	200	VdW Hbond	15	7	12	12	11 2 (Tyr126, Ser134)	16 3 (Trp136, Ile258, Cys268)	11 (Gln289, Met116)	12
	225	VdW Hbond	12	4	10	10	11 1 (Ser134)	15 3 (Trp136, Ile258, Cys268)	11 1 (Ile258)	10
	250	VdW Hbond	11 1 (Val284)	7	10	13	15 2 (Tyr126, Ser134)	14 2 (Trp136, Ile258)	11 1 (Ile258)	11
	275	VdW Hbond	10 1 (Val284)	6	10	12 1 (Trp136)	11 1 (Tyr126)	17 3 (Trp136, Ile258, Cys268)	13 1 (Ile258)	10
	300	VdW Hbond	12 1 (Val284)	8	10	14	11	14 2 (Trp136, Ile258)	7 2 (Ile258, Val259)	9
	325	VdW Hbond	12 2 (Val284, Tyr327)	5	12	14	14 2 (Val259, Ser134)	12 2 (Trp136, Ile258)	11 1 (Ile258)	10 1 (Cys268)
	350	VdW Hbond	12 1 (Tyr327)	10	12	14	11 3 (Val259, Asn245, Glu261)	15 3 (Trp136, Ile258, Cys268)	12 1 (Ile258)	8 1 (Gln289)

Y268N	0	VdW Hbond	19 1 (Asn268)	12	11	6 1 (Trp154)	14 1 (Glu261)	13 1 (<u>His320</u>)	12 3 (Glu261, Met133, Tyr263)	13 1 (Asn268)
	25	VdW Hbond	20 1 (Asn268)	10	14	7 1 (Trp154)	10 1 (Glu261)	17 2 (C304, Asn268)	14 2 (Val259, Phe137)	9 2 (Phe267, Asn268)
	50	VdW Hbond	18 1 (Asn268)	12	12	7 1 (Trp154)	10 2 (Glu261, Asn245)	19	11 2 (<u>Gln289</u> , Tyr263)	13(Asn268)
	75	VdW Hbond	16 1 (Asn268)	6	14	9 1 (Trp154)	11 3 (Glu261, <u>Asn245</u> , Ile119)	14 1 (Asn268)	13 4 (Trp136, Val259, Glu261, Tyr263)	15
	100	VdW Hbond	14 1 (Asn268)	4	13	6	7 3 (Glu261, <u>Asn245</u> , Ile119)	15 1 (Asn268)	11 3 (<u>Gln289</u> , Tyr263, Trp136)	11 1 (Asn268)
	125	VdW Hbond	15 1 (Asn268)	6	10	6	9 2 (Glu261, <u>Asn245</u>)	13 1 (Asn268)	13 2 (Tyr263, Trp136)	13 1 (Asn268)
	150	VdW Hbond	17 1 (Asn268)	9 1 (<u>Gln289</u>)	13	4	11 1 (Glu261)	15 1 (Asn268)	11, 2 (Tyr263, Asn268)	16 1 (Ile258)
	175	VdW Hbond	18 1 (Asn268)	7	13	7	9 1 (Ile119)	13 1 (Asn268)	9	13 1 (Ile258)
	200	VdW Hbond	22 1 (Asn268)	8	14	7	7 1 (Asn268)	15 1 (Asn268)	11(Met116)	15 1 (Ile258)
	225	VdW Hbond	17 1 (Asn268)	7	14	5	12	18 1 (Asn268)	8 3 (<u>Glu261</u> , Val259, Trp136)	13 1 (Ile258)
	250	VdW Hbond	16 1 (Asn268)	6	11	2	10 2 (Glu261, Ile119)	20 1 (Asn268)	13 2 (<u>Met116</u> , Trp136)	10 1 (Ile258)
	275	VdW Hbond	16 1 (Asn268)	5	11	6	9 2 (Glu261, Ile119)	15 1 (Asn268)	11 4 (Met116, <u>Glu261</u> , Gln289, Trp136)	8 1 (Ile258, Gly137)
	300	VdW Hbond	17 1 (Asn268)	5	12	3	8 2 (Glu261, Ile119)	14 1 (Asn268)	8 3 (Met116, <u>Glu261</u> , Trp136)	13 1 (Ile258)
	325	VdW Hbond	16 1 (Asn268)	4	11	2	11 2 (Glu261, Ile119)	15 1 (Asn268)	12 3 (Met116, <u>Glu261</u> , Trp136)	10 1 (Ile258)
	350	VdW Hbond	14 1 (Asn268)	6	14	5	10 1 (Glu261)	13 1 (Asn268)	15 4 (Met116, <u>Glu261</u> , Val259, Trp136)	10 1 (Ile258)
Y268S	0	VdW Hbond	15	12	13	12 1 (<u>Gln289</u>)	9	13 3 (Val259, Ser268, Gln289)	14 3 (W262, Glu261, Met133)	12 1 (Pro260)
	25	VdW Hbond	13 11(Phe267)	8	12	12	9 1 (Ser268)	15,	9 4 (W262, Glu261, Asn245, Met133)	14 1 (Pro260)
	50	VdW Hbond	11	10	18	10	10 3 (Met133, Ser268, Gln257)	14 1 (Ser268)	10 3 (Glu261, <u>Asn245</u> , Met133)	12 1 (Pro260)
	75	VdW Hbond	18	10	16	9	8 2 (Met133, <u>Gln257</u>)	19 2 (Trp136, Ser268)	15 3 (Tyr263, Glu261, <u>Asn245</u>)	14 2 (Trp136, Ser268)

	100	VdW Hbond	18	13	12	12	10	14 1 (Ser268)	13 1 (Met133)	13 1 (Ser268)
	125	VdW Hbond	14	11	12	9	9	18 2 (Ser268, Val259)	15	15
	150	VdW Hbond	15	11	17	12	9	18 1 (Val259)	12 3 (Leu285, Met133, <u>Asn245</u>)	12 (Trp136, Ser268)
	175	VdW Hbond	15	9	13	11 1 (Met116)	10	20 2 (Val259, Ser268)	15 2 (Leu285, Met133)	13 2 (Trp136, Ser268)
	200	VdW Hbond	14 2 (Val284, Phe267)	10	15	10	5 1 (Met133)	14	18 2 (Glu261, Phe264)	14 3 (Trp136, Ser268, Ile258)
	225	VdW Hbond	14 1 (Val284)	10	15	10	10 1 (Met133)	17 1 (Val259)	16 4 (Glu261, <u>Leu285</u> , Met133, Tyr263)	14 2 (His301, Ser268)
	250	VdW Hbond	14 2 (Val284, Leu285)	9	14	9	7	16 1 (Val259)	15 4 (Glu261, <u>Asn245</u> , Met133, Tyr263)	10 3 (Trp136, Ser268, Ile258)
	275	VdW Hbond	22 1 (Val284)	10	10	10	7	12 2 (Val259, Ser268)	16 3 (Glu261, <u>Asn245</u> , Tyr263)	7 3 (Trp136, Ser268, Ile258, His301)
	300	VdW Hbond	17	9	13	10 1 (Gln289)	11 1 (Met133)	11 2 (Val259, Ser268)	17 2 (Glu261, Tyr263)	12 1 (Ser268)
	325	VdW Hbond	18	7	13	7 1 (Gln289)	13 1 (Met133)	15 2 (Val259, Ser268, Trp136)	13 3 (Glu261, Met133, Tyr263)	9 3 (Trp136, Ser268, Ile258)
	350	VdW Hbond	17	9	14	7 1 (Gln289)	13 1 (Met133)	14 2 (Val259, Trp136)	17 3 (Glu261, Met133, Tyr263)	12 2 (Trp136, Ser268)

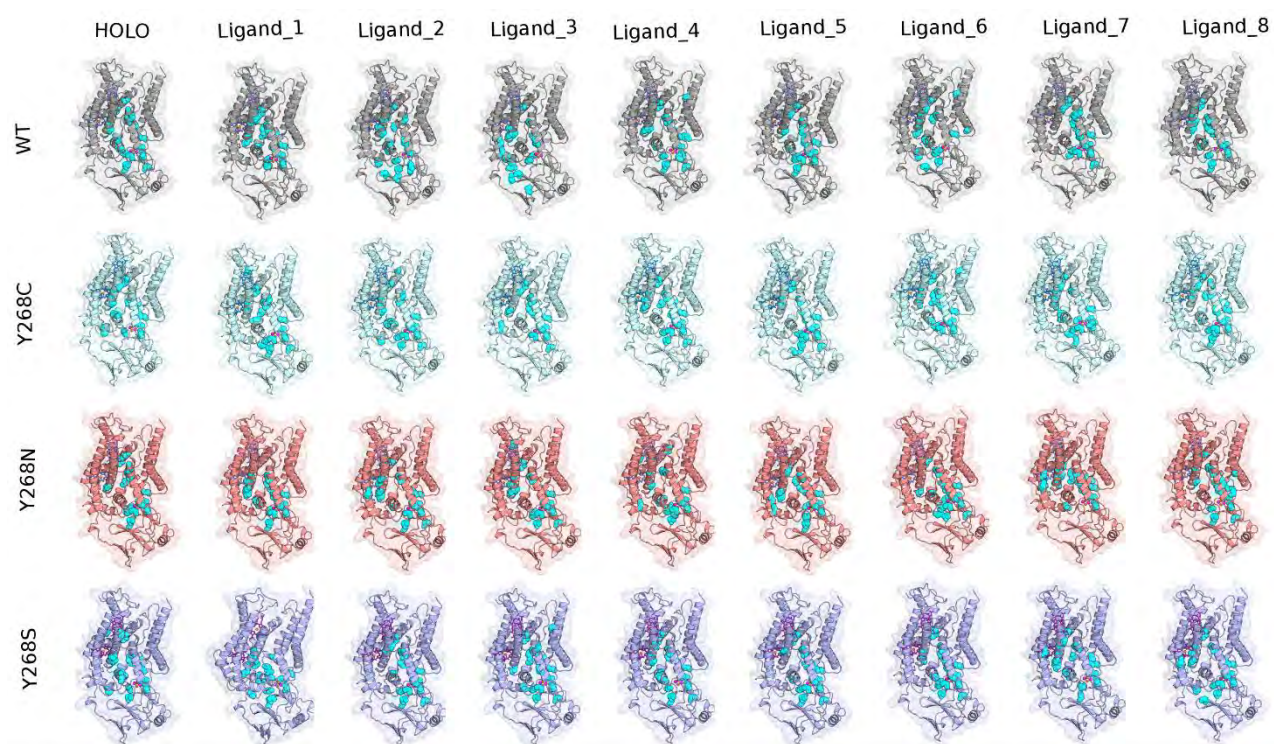


Figure S5.14: Structural mapping of top 4% BC residues onto PfCytb-ISP protein in WT and mutant models (Y268C, Y268N, and Y268S). Protein communication was assessed in a total of 8 ligand bound protein systems Ligand 1-8 representing ligand systems MMV007638, MMV019551, MMV659004, MMV676444, SANC00373, SANC00459, SANC00548, and SANC00725, respectively.

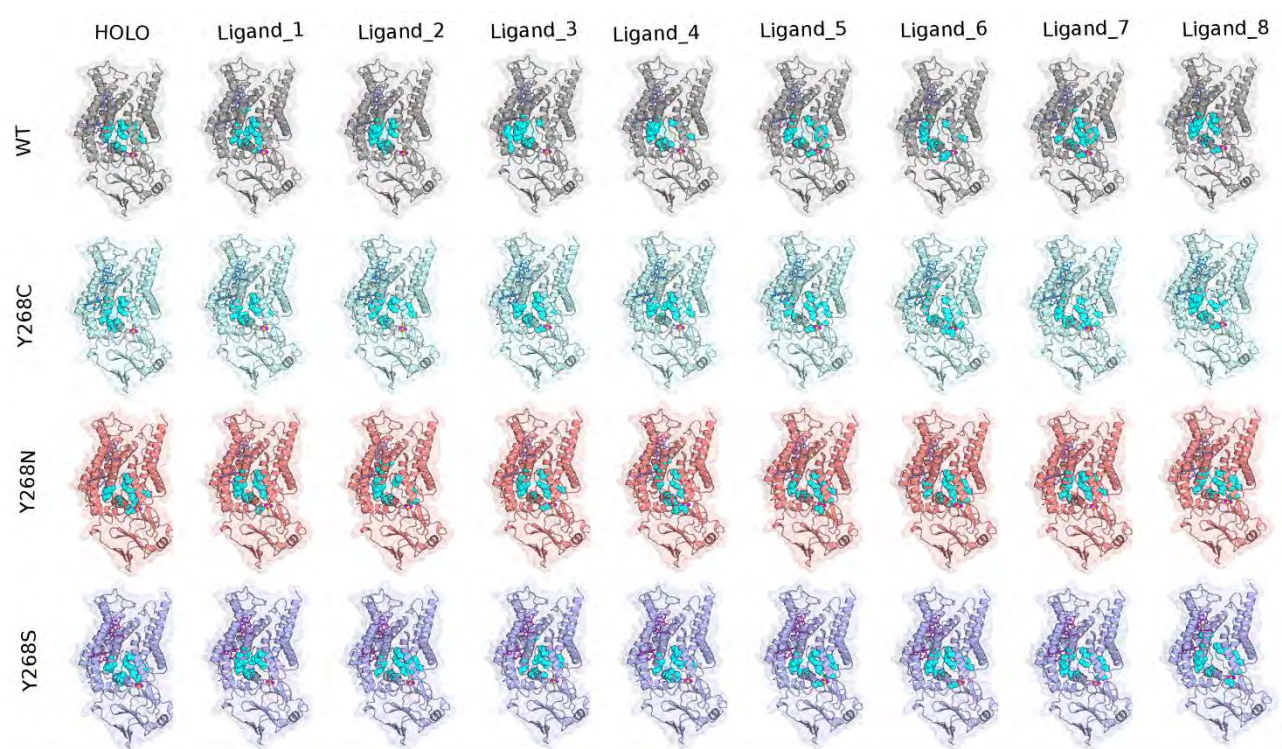


Figure S5.15: Structural mapping of top 4% CC residues onto PfCytb-ISP protein in WT and mutant models (Y268C, Y268N, and Y268S). Protein communication was assessed in a total of 8 ligand bound protein systems Ligand 1-8 representing ligand systems MMV007638, MMV019551, MMV659004, MMV676444, SANC00373, SANC00459, SANC00548, and SANC00725, respectively.

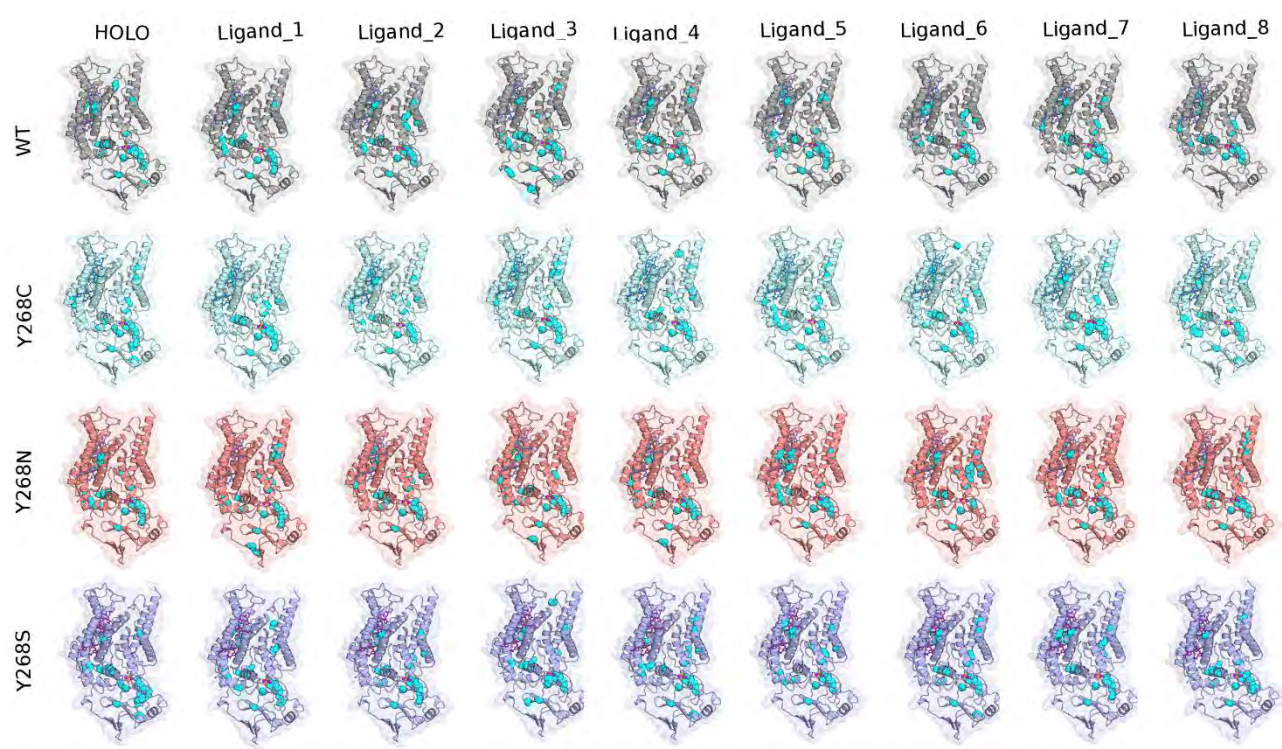


Figure S5.16: Structural mapping of top 4% DC residues onto PfCytb-ISP protein in WT and mutant models (Y268C, Y268N, and Y268S). Protein communication was assessed in a total of 8 ligand bound protein systems Ligand 1-8 representing ligand systems MMV007638, MMV019551, MMV659004, MMV676444, SANC00373, SANC00459, SANC00548, and SANC00725, respectively.

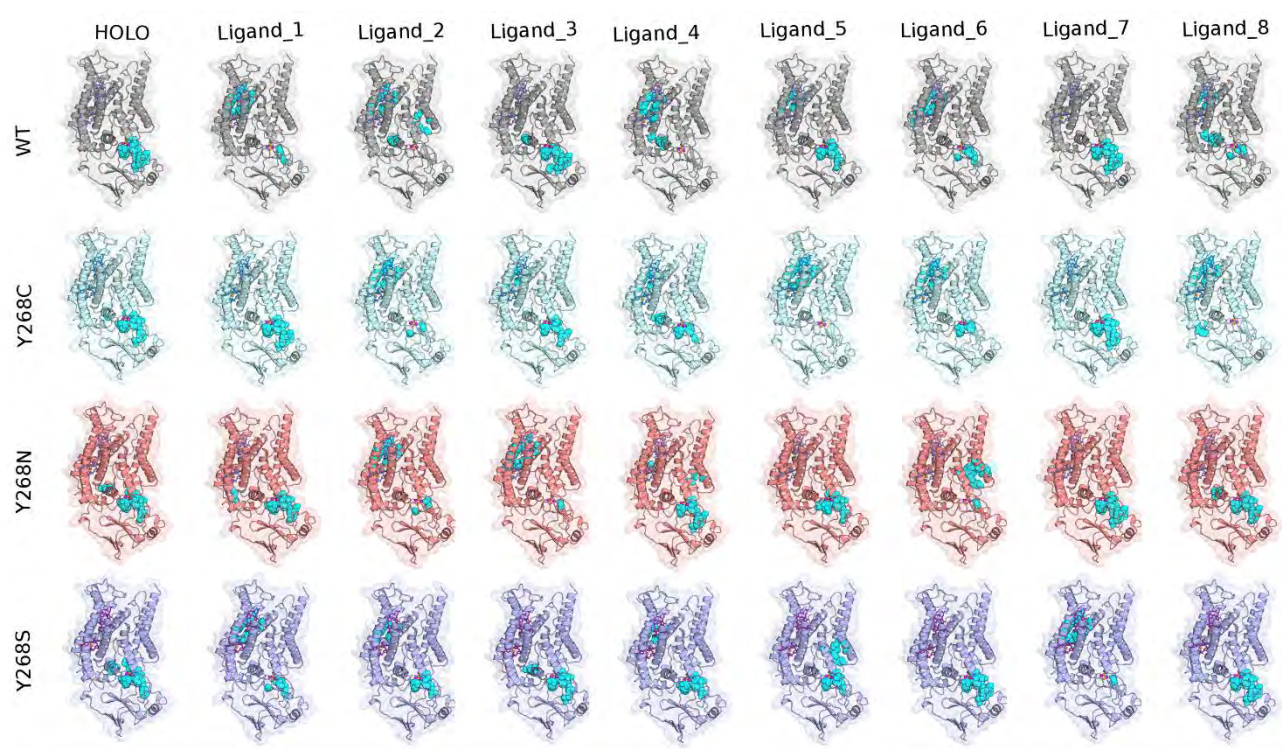


Figure S5.17: Structural mapping of top 4% *EC* residues onto PfCytb-ISP protein in WT and mutant models (Y268C, Y268N, and Y268S). Protein communication was assessed in a total of 8 ligand bound protein systems Ligand 1-8 representing ligand systems MMV007638, MMV019551, MMV659004, MMV676444, SANC00373, SANC00459, SANC00548, and SANC00725, respectively.

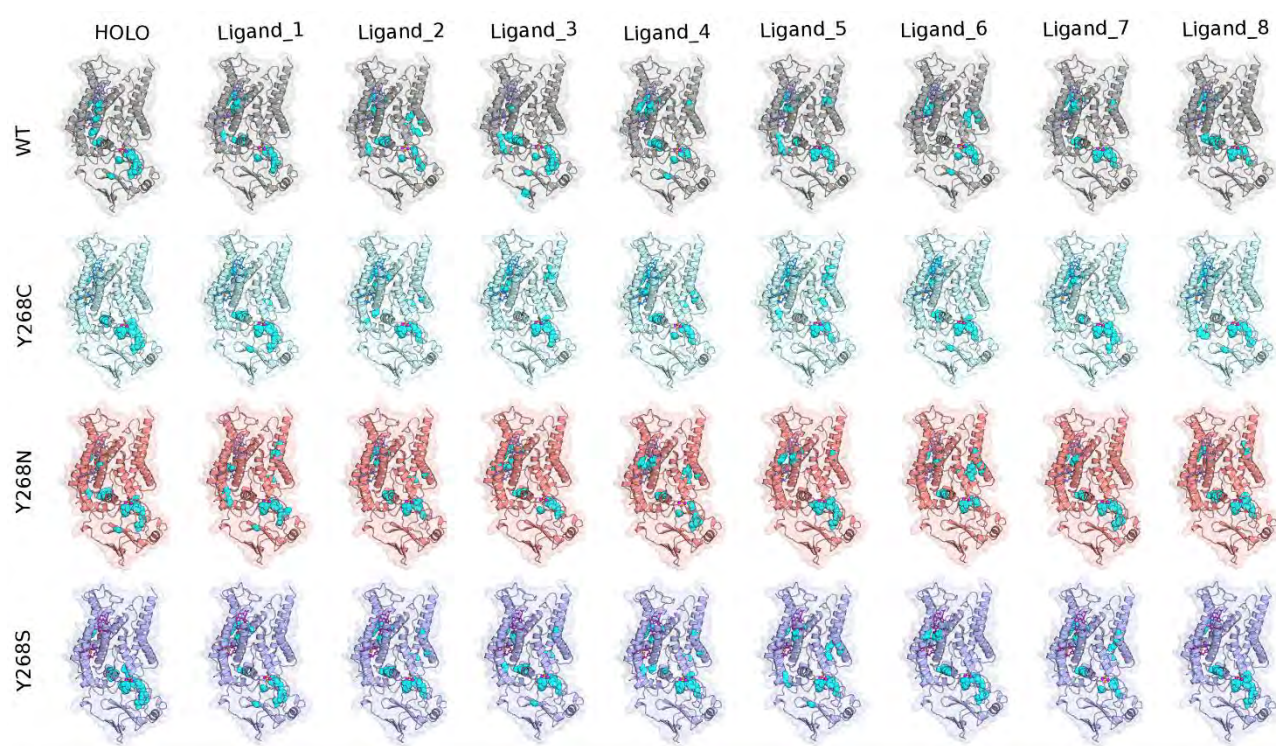


Figure S5.18: Structural mapping of top 4% *KC* residues onto PfCytb-ISP protein in WT and mutant models (Y268C, Y268N, and Y268S). Protein communication was assessed in a total of 8 ligand bound protein systems Ligand 1-8 representing ligand systems MMV007638, MMV019551, MMV659004, MMV676444, SANC00373, SANC00459, SANC00548, and SANC00725, respectively.

Table S5.7: Residues having significant ΔBC for eight protein-ligand systems showing conformational/structural changes ; + significant positive values while – stands for significant negative values. In bold are active site residues while in blue are metal coordinating residues. Calculations performed as follows: (Ligand-free – Ligand-bound).

		Change in <i>Betweenness Centrality</i> (ΔBC)			
		WT	Y268C	Y268N	Y268S
MMV007638	+	Ser83, Ile93, Leu97, Phe123 , Ile141 , Ser256, Trp262 , Leu265, Lys277, Ile309, Gly333, Ile298	Leu25, Thr80, Met102, Ile119, Phe123 , Val140 , Ile141 , Thr160, Val161, Trp262 , Leu265, Leu281, Lys252, Cys304	Met133 , Asn143, Cys156, Ile237, Ser241, His242, Asp244, Tyr252, Pro255, Gln257, Leu302, Gly303, Val305, Cys319, His320 , Ser322	Ile35, Phe86, Phe123 , Val140 , Ile141 , Thr142 , Ile155 , Glu261 , Leu265, Leu271 , Ala295, Cys304, Ala307, Pro318, Gly321
	-	Phe86, Phe118, Thr121, Val124, Val127, Met133 , Ile237, Ser241, His242, Asn250, Thr251, Tyr252, Thr254, Ile258 , Val259 , Tyr263 , Phe264	Ile22, Asn27, Tyr252, Thr254, Val259 , Pro260 , Tyr263 , Phe264 , Ile320, Gly303, His320	Thr80, Ala82, Val85, Phe118, Ala122, Val127, Thr139, Val161, Leu172, Ala246, Val259 , Ile320, Ile298, Arg331	Phe118, Tyr126, Asn143, Leu144, Cys156, Thr165, Phe169 , Ile237, Ser241, His242, Tyr252, Thr254, Ile258 , Val259 , Tyr263 , Leu281, Val282, Gly303, Cys319
MMV019551	+	Ser83, Ile93, Asn143, Pro255, Trp262 , Leu265, Tyr268 , Leu271, Lys272, Leu291, Ile309, Pro337, Gly297, Val305, His320 , Gly321, Ser322, Pro334	Leu25, Phe86, Phe123 , Met133 , Ile141 , Val161, Gln232, Trp262 , Leu265, Ile349, Lys252, Cys304, Pro336	Leu25, Ser83, His92 , Ile93, Ile237, Ser241, His242, Thr254, Pro255, Gln257, Pro260 , Leu288, Leu291, Leu294, Leu302, Val305, Cys319, Ser322	Ile35, Phe123 , Val140 , Thr142 , Tyr159, Thr160, Pro260 , Glu261 , Leu265, Phe267 , Leu271 , Cys304, Ala307, Pro318
	-	Phe118, Ala122, Thr139, Thr160, Leu172 , Ser276, Leu281, Ile320, Lys252, Gly303, Pro336, Ser337	Ala138, Asn143, Thr254, Phe264, Lys277, Leu290, Leu357, His301 , Leu302, Gly303, His320	Val85, Phe86, Thr89, Leu94, Asn98, Phe118, Ala122, Thr139, Thr160, Ile238, Ile320, Pro334, Ala335	Leu112, Phe118, Ile237, Ser241, Tyr252, Thr254, Tyr263 , Lys272, Lys277, Leu281, Leu285 , Leu288, Gln289, Phe292, His301 , Gly303
MMV659004	+	Ser83, Phe123 , Val140 , Ile141 , Asn143, Trp262, Leu265, Lys272, Pro337, Gly297, Ile298, His301 , Val305, His320 , Gly321, Ser322, Ile330	Leu25, Ala82, Ile119 , Ala122, Phe123 , Val140 , Ile141 , Thr160, Trp262 , Leu265, Leu281, Lys252, Cys304, Ala307	Ile93, Met116 , Phe229, Ser233, Ile237, Ser241, Ile258 , Val259 , Leu281, Leu285 , Leu291, Met308, Arg313, Leu357, Val361, Ser322, Tyr324	Val120 , Phe123 , Thr139, Ile141 , Thr142 , Ser146, Tyr159, Thr160, Pro260 , Glu261 , Cys304, Ala307
	-	Thr121, Val124 , Ile153, Cys156, Thr160, Ser162, Thr165, Phe169 , Val259 , Tyr263 , Pro275, Leu281, Arg250, Pro336, Ser337	Ile22, Asn27, Val124 , Pro149, Phe169 , Thr254, Val259 , Pro260 , Tyr263 , Phe264 , Gly280, Leu288, Gly303, His320 , Ser322	Ile22, Val85, Phe86, Thr89, Leu97, Asn98, Phe118, Ala122, Tyr126, Met133 , Gly221, Tyr263 , Phe264 , Ile309, Tyr363, His323	Thr89, Leu112, Val127, Gly137 , Leu144 , Ser241, His242, Tyr252, Tyr263, Ala269, Lys272, Leu281, Leu285 , Leu288, Ile320, Gly303, Val305, His320 , Ser322
MMV676444	+	Ser83, Ile93, Ala122, Phe123 , Ile141 , Asn143, Ser256, Trp262 , Leu265, Leu271, Ile309, His320 , Ser322	Ala82, Tyr101, Ile119 , Ala122, Phe123 , Val140 , Ile141 , Leu144 , Thr160, Thr197, Trp262 , Leu265, Leu271 , Leu281, Lys252, Cys304	Ser83, Ile93, Ser107, Ile237, Ser241, His242, Thr251, Tyr252, Pro255, Gln257, Glu261 , Leu271 , Leu281, Leu285 , Leu357, Cys319, Ser322	Ile35, Phe86, Val120 , Thr142 , Ile155 , Thr160, Ile258 , Pro260 , Leu265, Phe267 , Ala295, Ala307
	-	Phe86, Val124 , Thr139, Leu172, Gly236, Pro239, Asn245, Thr251, Tyr263 , Phe264 , Ile320, Ala335	Val124 , Leu172 , Ile237, Ser241, His242, Tyr252, Thr254, Ile258 , Val259 , Tyr263 , Phe264 , Lys277,	Ile35, Val85, Leu94, Asn98, Phe118, Val120 , Ala122, Phe123 , Thr139, Leu144 , Leu240, Val253, Phe326	Tyr16, Asn27, Tyr28, Leu31, Leu88, Gly137 , Asn143, Tyr252, Tyr263 , Lys272, Leu281, Ile320, Gly303, His320 , Ser322, Ile330

			Leu288, His301 , Leu302, Gly303		
SANC00373	+	Leu25, Ser83, Ile93, Val140 , Ile141 , Asn143, Pro199, Pro255, Ser256, Trp262, Leu265, Val274, Ile283, Ile309, Ile298, Val305	Leu25, Thr80, Ser83, Tyr101, Ile119 , Phe123 , Thr139, Ile141 , Thr160, Val161, Trp262 , Leu265, Lys252, Cys304, Pro336	His92 , Ile93, Tyr99, Val249, Asn250, Thr251, Tyr252, Thr254, Glu261 , Gln289, Leu357, His301 , Leu302, Gly303, Ser322, Tyr324	Phe86, Val140 , Thr142 , Thr160, Val225, Pro260 , Glu261 , Leu265, Leu271 , Ala295, Cys304, Ala307, Pro318
	-	Phe86, Val124 , Thr160, Phe169 , Leu172 , Leu176, Thr251, Tyr252, Tyr263 , Phe264 , Leu285	Ile22, Val124 , Val127, Leu240, Tyr252, Thr254, Val259 , Pro260 , Tyr263 , Phe264 , Pro278, Gly280, Val284, Leu288, Pro253, Leu302, Gly303	Val85, Met102, Leu112, Phe118, Thr121, Val127, Gly137 , Thr160, Val161, Thr165, Phe169 , Leu240, Pro255, Trp262 , Leu271 , Thr300, Arg331, Ala335	Tyr252, Thr254, Val259 , Tyr263 , Lys272, Leu281, Leu285, Leu290, Ile320, Leu293, Gly297, His301 , Gly303, Val305
SANC00459	+	Ser83, Ile93, Phe123 , Ile141 , Thr254, Pro255, Trp262 , Leu265, Ile309, Val305, His320 , Gly321, Ser322, Ile330	Ile119 , Phe123 , Met133 , Val140 , Ile141 , Leu144 , Arg168, Trp262 , Leu265, Cys304	Leu25, Ile93, Trp154, Ile155 , His192 , Thr197, Ile237, Ser241, His242, Thr254, Gln257, Leu271 , Arg313, Leu357, Val361, Val305, Cys319, Ser322, Tyr324, Asn326	Ser83, Thr142, Thr160, Pro260 , Glu261 , Leu265, Cys304
	-	Phe118, Val127, Thr139, Thr160, Ser241, Thr251, Tyr263 , Phe264 , Ile320, Leu302, Pro336	Phe118, Gly137 , Asn143, Thr165, Phe169 , Ile237, Ser241, His242, Tyr252, Thr254, Val259 , Tyr263 , Lys277, His301 , Leu302, Gly303, His320 , Ser322	Thr80, Val85, Thr139, His194, Pro199, Gly221, Ile258 , Phe264 , Leu281, Leu294, Ile320, Gly303, Gln308	Leu112, Ala122, Tyr126, Gly137 , His187 , Ser241, Tyr263 , Lys272, Lys277, Leu281, Leu285 , Gln289, Ile320, Ile298, His301 , Leu302, Gly303, His320 , Ser322
SANC00548	+	Leu25, Ser83, Ile93, Phe123 , Ile141 , Thr142 , Asn143, Phe229, Pro255, Trp262 , Leu265, Leu271 , Ile309, Ile298	Ile119 , Phe123 , Val140 , Ile141 , Thr160, Thr197, Ser241, Trp262 , Leu265, Leu281, Lys252, Cys304	Ile93, Met116 , Val124, Met133 , Cys156, Thr254, Gln257, Glu261 , Leu288, Leu291, Leu293, Leu357, Val361, Val305	Phe86, Val120 , Phe123 , Val140 , Ile141 , Thr142 , Pro260 , Glu261 , Phe264 , Leu265, Phe267 , Leu271 , Cys304, Ala307, Pro318
	-	Phe86, Met116 , Thr121, Val127, Thr165, Arg168, Leu240, Val259 , Tyr263 , Phe264 , Lys277, Leu281, Val284, Gln289, His301	Phe118, Val124, Tyr135, Pro199, Thr251, Tyr252, Pro260 , Lys272, Gly303, Gly321, Pro334	Ala82, Val85, Leu97, Tyr101, Met102, Ile119 , Ala122, Tyr126, Ser134, Thr160, Val161, Trp262 , Leu265, Tyr268 , Ala269, Ala295, Val317, Ile320, Leu360, Gly251	Phe118, Ala122, Val127, Ala138, Ile237, Ser241, His242, Tyr263, Lys272, Lys277, Leu288, Ile320, His320
SANC00725	+	Thr80, Ser83, Ile93, Ile141 , Asn143, Thr254, Pro255, Ser256, Trp262, Leu265, Lys272, Ile283, Ile309, His320	Thr80, Ser83, Tyr101, Ala122, Phe123 , Thr139, Ile141 , Leu145, Thr160, His187 , Trp262 , Leu265, Leu271, Lys252, Cys304, Ala307, Pro318	Asn23, Ser83, Leu84, Ile93, Met116 , Val124 , Ile237, His242, Thr254, Phe264 , Lys277, Leu288, Leu302, Val305, Ser322	Ile35, Phe123 , Val140 , Ile141 , Thr142 , Ser146, Val225, Pro260 , Glu261 , Leu265, Ala295, Cys304, Ala307
	-	Asn27, Phe118, Ile119 , Thr139, Val140 , Ser241, Val259 , Pro260 , Glu261 , Tyr263 , Phe264	Tyr90, Val124 , Pro149, Ser241, His242, Asp244, Thr251, Thr254, Pro260 , Tyr263 , Phe264 , Ala269, Leu288, Leu302	Val85, Leu97, Phe118, Ala122, Phe123 , Tyr126, Gly137 , Thr139, Leu172 , Leu240, Phe267 , Val274, Ile320, Ile298, Gly303, Pro334	Leu31, Leu112, Val124 , Val127, Ala138, Tyr263 , Phe264 , Leu271 , Lys272, Lys277, His301 , Leu302, His320 , Ser322

Table S5.8: Residues having significant ΔBC for eight protein-ligand systems showing conformational/structural changes ; + significant positive values while – stands for significant negative values. In bold are active site residues while in blue are metal coordinating residues. Calculations performed as follows: (WT Ligand-bound – Mutant Ligand-bound).

		Change in <i>Betweenness Centrality</i> (ΔBC)		
		Y268C	Y268N	Y268S
MMV007638	+	Phe86, Asn98, Phe118, Val124 , Val127, Met133 , Ile237, Asp244, Asn250, Thr254, Pro255, Ile258 , Tyr263 , Leu265, Lys272, Thr273, Leu294, Val305	Phe86, Thr89, Tyr126, Met133 , Leu144 , Ile237, Ser241, His242, Asp244, Val249, Tyr252, Thr254, Pro255, Tyr263 , Leu265, Lys272, Val305	Phe86, Leu112, Val124, Val127, Met133 , Arg168, Tyr252, Thr254, Pro255, Glu261 , Tyr263 , Phe264 , Leu265, Lys272, Pro275, Val305, His320 , Gly321
	-	Trp26, Ser83, Leu97, Trp136 , Pro260 , Leu271 , Leu290, Ile320, Leu357, Gly303	Leu25, Ile35, Ser83, Thr139, Leu172 , Ala246, Lys277, Leu281, Leu285 , Leu291, Leu293, Ile298, His301	Leu144, Thr160, Val161, Thr165, Phe169 , Ile258 , Val259 , Pro260 , Leu281, Val282, Leu336, Cys319
MMV019551	+	Ala82, Ala122, Phe123 , Thr139, Ile141 , Thr160, Val161, Leu172 , Leu176, Ser276, Lys252, Gly303, Pro336, Ser337	Ala82, Ala122, Phe123 , Val140 , Ile141 , Thr160, Ile237, Ser276, Gly280, Leu290, Leu294, Lys252, Gly303, Pro336, Ser337	Ala82, Tyr101, Ala122, Phe123 , Thr139, Val140 , Ile141 , Thr160, Val161, Ser276, Lys252, Gly303, Pro336, Ser337
	-	Ser83, Asn250, Leu271 , Lys272, Pro337, His320 , Ser322, Pro334	Ser83, Phe86, Val124 , Val127, Thr251, Tyr252, Tyr263 , Leu271 , Lys272, His320 , Pro334	Ser83, Phe86, Thr251, Tyr252, Tyr263 , Lys272, Pro337, Val305, His320 , Ser322
MMV659004	+	Tyr99, Leu144 , Cys156, Tyr159, Thr160, Ser162, Thr165, Phe169 , Leu215, Thr273, Val274, Pro275, Leu281, Arg250, Ile296, Pro336, Ser337, Pro343	Tyr99, Leu144 , Ile153, Thr160, Ser162, Thr165, Phe169 , Val259 , Val274, Pro275, Leu281, Arg250, Pro336, Ser337, Leu339	Tyr99, Val124 , Ile153, Cys156, Tyr159, Thr160, Ser162, Thr165, Phe169 , Thr254, Thr273, Val274, Pro275, Ile222, Trp249, Arg250, Lys252, Pro336, Ser337
	-	Ile237, Phe264 , Lys272, Gly297, His301 , His320 , Ser322	Met133 , Glu261 , Phe264 , Asn268 , Lys272, Ile309, Pro337, Gly297, Ile298, Leu302, Cys319, His320 , Gly321, Ser322, His323	Gly137 , Val140 , His242, Glu261 , Ala269, Lys272, Gly297, Ile298, Val305, His320 , Gly321, Ser322
MMV676444	+	Phe115, Ile141 , Leu144 , Ser147, Gly236, Pro239, Thr251, Leu281, Leu294, Ile298, Ala335	Phe86, Val124 , Ser147, Leu172 , Gln232, Gly236, Pro239, Thr251, Tyr252, Pro255, Phe264 , Leu281, Phe292, Ala335	Phe86, Thr89, Thr121, Val124 , Ile155 , Leu172 , Gln232, Gly236, Pro239, Val249, Thr251, Pro255, Ile258 , Lys272, Pro278, Val282, Leu286, Leu288, Leu294, His301 , Leu302, Gly303, Ala335
	-	Thr160, Ile237, Ser241, His242, Ile258 , Val259 , Glu261 , Leu265, Leu293, Ser322, Asn326	Ile35, Leu94, Phe118, Val120 , Phe123 , Thr139, Thr160, Leu240, Val253, Thr254, Ser256, Leu265, Phe326, Val246, Leu302, Ser322	Ser83, Arg95, Gly137 , Ile141 , Leu240, Pro260 , Glu261 , Leu271 , Cys304, Cys319, Ser322
SANC00373	+	Thr80, Ser83, Tyr101, Leu112, Phe123 , Tyr135, Thr139, Thr160, Ser162, Leu172 , Leu176, Asp244, Phe264 , Tyr268 , Leu271 , Leu288, Gly251, Lys252, His320 , Ser322	Phe86, Tyr101, Phe123 , Tyr126, Tyr135, Leu172 , Thr251, Tyr252, Tyr263 , Phe264 , His301 , Leu302, Gly303, Ser322	Phe86, Ser107, Leu112, Thr160, Val161, Arg168, Leu172 , Leu176, Tyr263 , Phe264 , Ser268 , Leu271 , Lys277, Leu288, Gly251, Leu302, His320 , Ser322

	-	Phe118, Ala122, Val127, Gln132, Trp136, Val140 , Ser147, Pro199, Ile237, Ser241, Asn250, Thr254, Val259, Pro260, Trp262 , Pro278, Leu300, Ile320, Pro253, Cys304, Val305, Cys319	Met133, Gly137 , Thr165, Phe169 , Ile237, Ser241, His242, Pro255, Val259, Pro260, Trp262 , Val284, Ser287, Arg313, Arg250, Pro253, Ala335	Leu94, Thr139, Leu144 , Asn250, Thr254, Val259, Glu261 , Arg313, Ile320, Cys324, Leu357, Leu293, Ile298, Gly303, Val305
SANC00459	+	Asn27, Val127, Met133 , Ser134, Ala138, Thr139, Leu144 , Arg168, His192, Asn245, Leu288, Phe219, Ile298, Leu302, Gly303, Pro336, Ser337	Asn27, Thr89, Leu97, Phe118, Ala122, Leu144 , Thr160, Val161, His192, Ser241, Leu271 , Ile298, Leu302, Pro336	Asn27, Leu97, Phe118, Val127, Gln132, Met133, Trp136 , Ala138, Asn143, Leu144 , Thr160, Val161, Ser241, Leu288, Phe219, Trp249, Ile298, Leu302, Pro336, Ser337
	-	Trp154, Thr165, Ile237, Thr254, Pro255, Ile258, Val259 , Lys277, His301, His320 , Ser322	Ser83, Leu84, Gly195, Pro255, Ile258, Val259, Pro260, Phe264 , Lys277, His301 , Gln308, Pro318, His320 , Ile330	Phe123, His187 , Pro239, Ile258, Val259 , Lys277, Thr301, Pro318, His320 , Ser322
SANC00548	+	Leu144 , Thr160, Ile188, Ser241, Thr254, Trp262 , Leu265, Lys277, Leu281, Val284, Leu288, Gln289, Phe219, Lys252, His301 , Leu302, Val305, His320	Tyr16, Met116, Trp136, Leu144 , Ile188, Leu281, Val284, Leu286, Leu288, Gln289, Leu290, Leu357, His301 , Val305, Ile330	Tyr16, Phe86, Met116, Leu144 , Ile188, Leu240, Thr254, Val259, Pro260, Glu261, Tyr263, Ser268 , Lys277, Leu281, Val284, Leu285 , Gln289, Leu290, His301 , Ile330
	-	Ser83, Tyr101, Phe118, Tyr126, Tyr135, Pro199, Leu200, Tyr252, Pro255, Leu271 , Gly303, Cys304, Gly321, Pro334	Leu25, Asn27, Tyr101, Ala122, Tyr126, Asn143, Ile153, Thr197, Gln232, Ile237, His242, Ile258, Leu271 , Ala295, Val246, His320	Asn27, Leu94, Tyr99, Phe118, Ala122, Thr139, Ile153, Ile237, His242, Ile320, Cys324, Cys334
SANC00725	+	Asn27, Phe118, Ile119 , Thr139, Val140 , Ser162, His192, Val259, Pro260, Glu261 , Leu357, Ile296, Ile298, Gly303	Asn27, Met102, Ile119, Val140 , Thr160, Val161, Ser162, Pro260, Glu261, Phe264	Phe118, Ile119 , Ala122, Phe123 , Thr139, Val140 , Ile237, Ser241, Val259, Pro260, Glu261 , Pro266, Met308, Ile298
	-	Leu25, Tyr90, Ile141, Pro149, Val152, Cys156, His194, Thr254, Glu340	Leu25, Ile35, Leu97, Tyr126, Asn245, Thr254, Ile320, His301 , Gly303	Leu25, Thr80, Val124 , Val127, Leu271 , Val282, His320

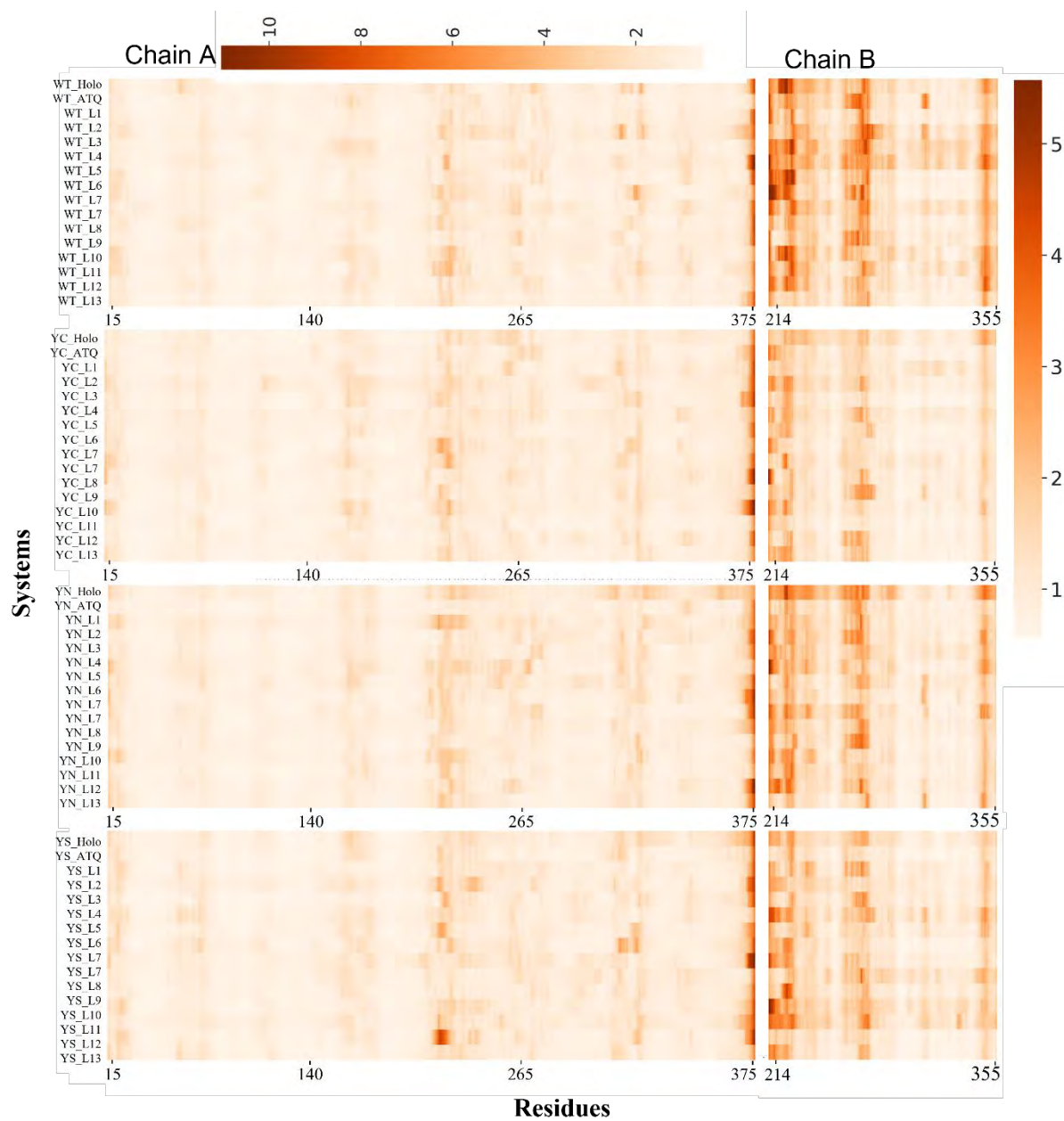


Figure S5.19: A heat map representation of RMSF in holo, ATQ and thirteen ligand systems during 350 ns MD simulations. White to dark orange represent the degree of protein flexibility going from more flexible to less flexible respectively.

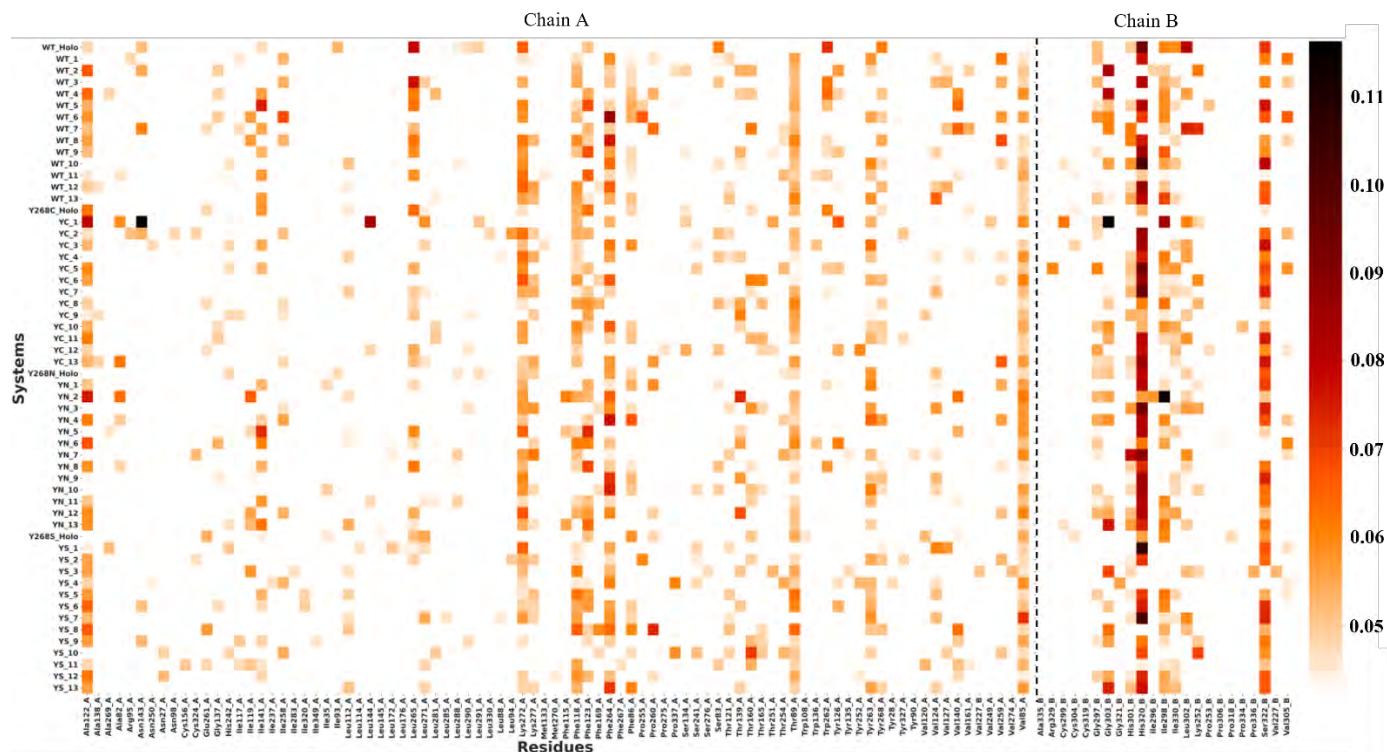


Figure S5.20: A heat map representation of top 4% *BC* values representing significant residues affecting protein communication across 13 ligand systems in their WT and mutant states. Ligand systems 1-13 representing MMV020982, MMV023860, MMV024101, MMV024937, MMV676050, MMV676599, MMV687700, SANC00257, SANC00264, SANC00347, SANC00559, SANC00587 and SANC00595 respectively. Active site residues and metal interacting residues are displayed in purple and green dots respectively.

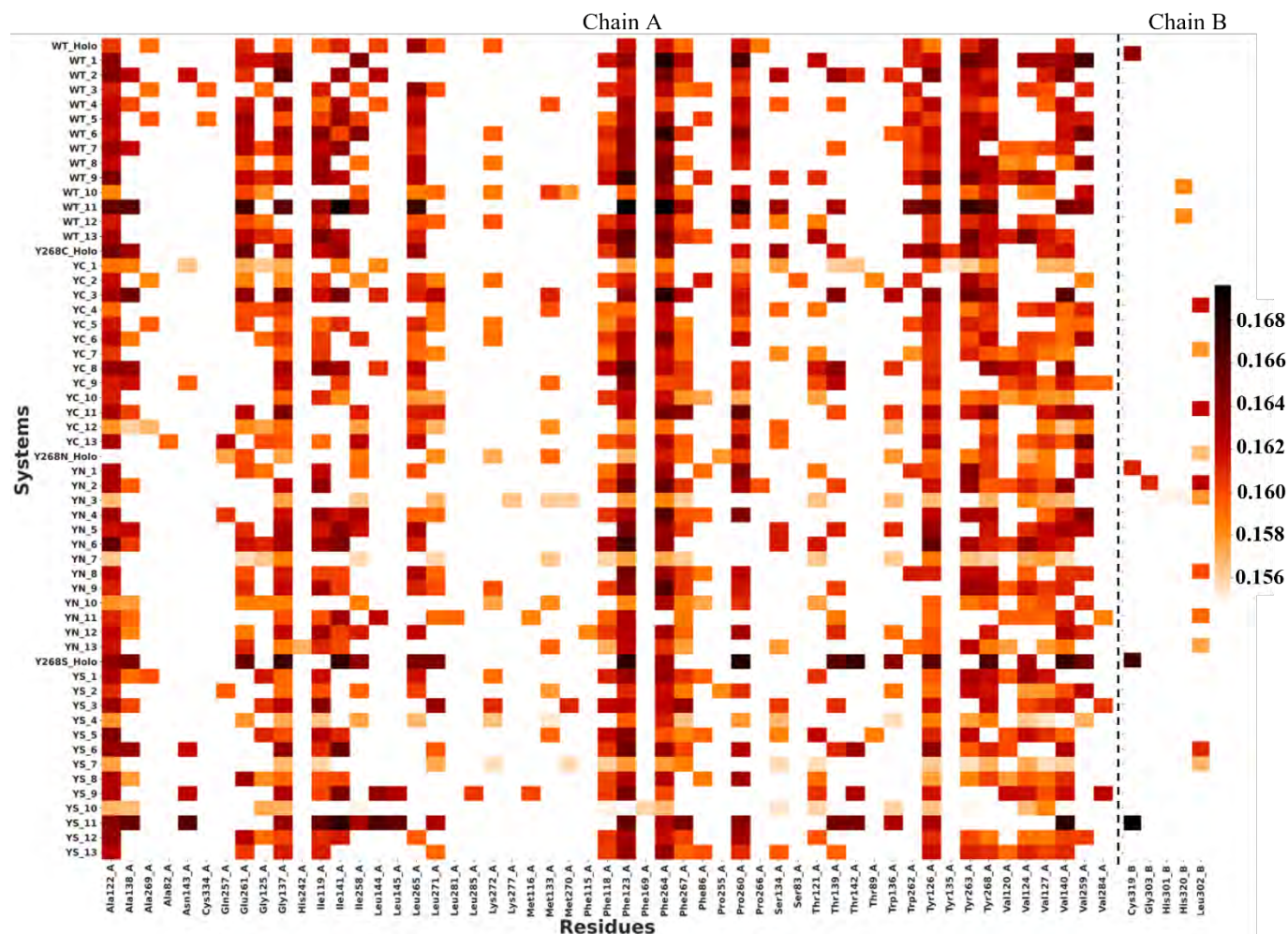


Figure S5.21: A heat map representation of top 4% *CC* values representing significant residues affecting protein communication across 13 ligand systems in their WT and mutant states. Ligand systems 1-13 representing MMV020982, MMV023860, MMV024101, MMV024937, MMV676050, MMV676599, MMV687700, SANC00257, SANC00264, SANC00347, SANC00559, SANC00587 and SANC00595 respectively. Active site residues and metal interacting residues are displayed in purple and green dots respectively.

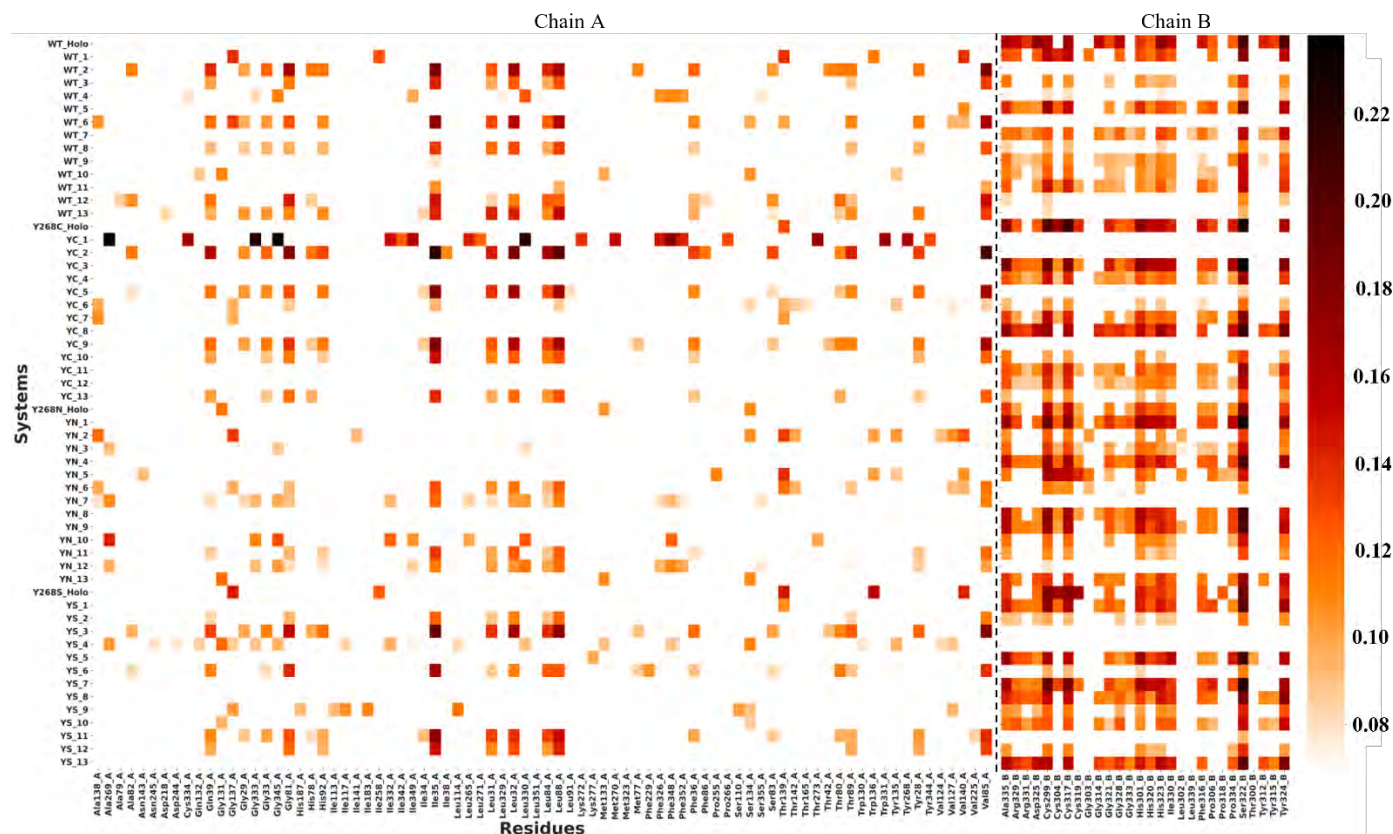


Figure S5.23: A heat map representation of top 4% *EC* values representing significant residues affecting protein communication across 13 ligand systems in their WT and mutant states. Ligand systems 1-13 representing MMV020982, MMV023860, MMV024101, MMV024937, MMV676050, MMV676599, MMV687700, SANC00257, SANC00264, SANC00347, SANC00559, SANC00587 and SANC00595 respectively. Active site residues and metal interacting residues are displayed in purple and green dots respectively.

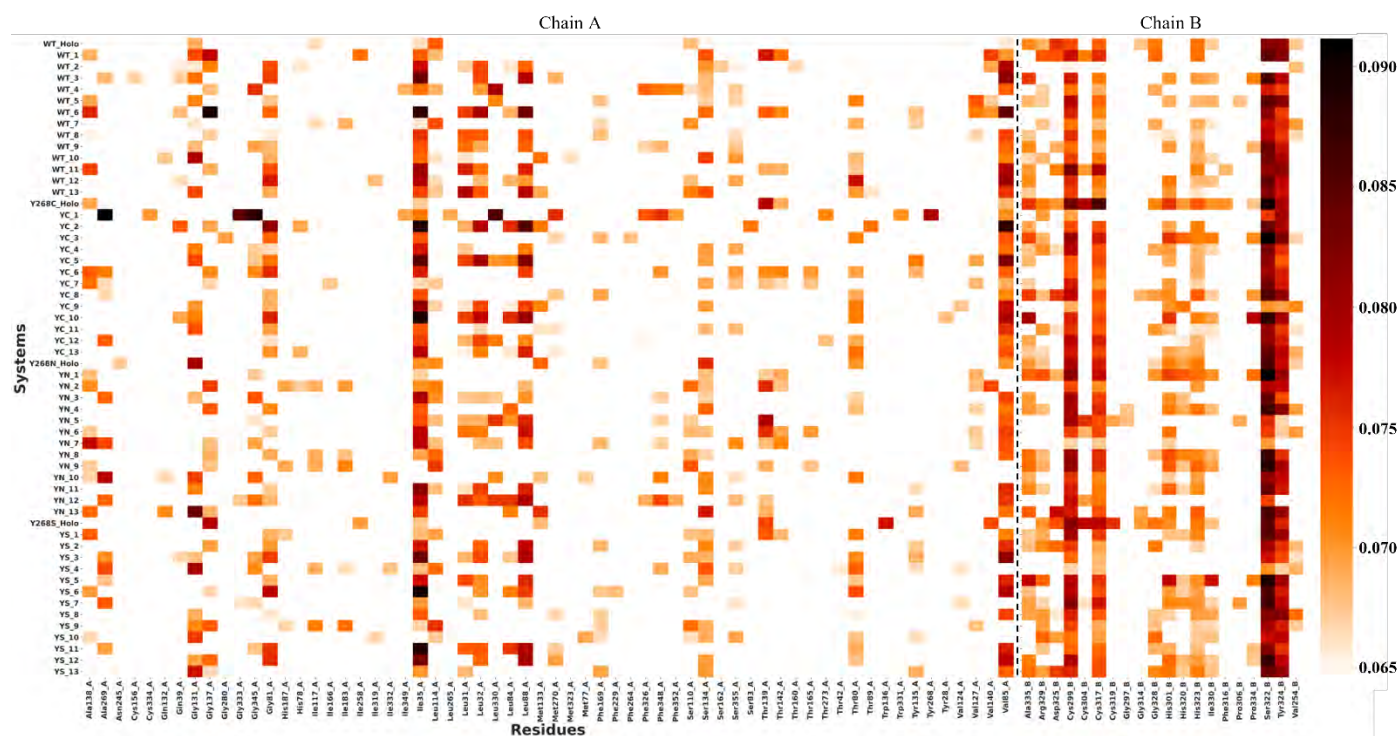


Table S5.9: Residue-residue interactions observed to be either reduced, lost, or gained/enhanced around the hubs identified in the 8 ligand systems across mutant (Y268C, Y268N and Y268S) systems as compared to WT. Residue interactions reduced/lost/gained in two mutant ligand systems (coloured in green).

Ligand systems	REDUCED/LOST			GAINED/ENHANCED		
	Y268C	Y268N	Y268S	Y268C	Y268N	Y268S
MMV007638	-	-	-	His301-Lys115	His301-Lys115	-
MMV019551	-	-	-	-	Tyr263-Ser83	-
MMV659004	-	-	Tyr263-Glu261	-	-	-
MMV676444	-	Pro260-Leu265	-	His301-Lys115	His301-Lys115	-
SANC00373	Phe264-Glu261	-	-	-	His320-Pro318	-
SANC00459	-	Thr80-Gly236, Pro260-Leu265	Thr80-Gly236, Thr80-Ile237	-	-	-
SANC00548	Phe264-Glu261	Thr80-Gly236	Thr80-Gly236, Tyr263-Glu261	-	His320-Pro318	-
SANC00725	-	-	Thr80-Ile237	-	Tyr263-Ser83	-

Table S5.10: Residue-residue interactions observed to be either reduced, lost, or gained/enhanced around the hubs identified in the 13 ligand systems across mutant (Y268C, Y268N and Y268S) systems as compared to WT. Residue interactions reduced/lost/gained across all the three-mutation ligand systems (coloured in blue) and in two mutant ligand systems (coloured in green).

Ligand systems	REDUCED/LOST			GAINED/ENHANCED		
	Y268C	Y268N	Y268S	Y268C	Y268N	Y268S
MMV020982	Tyr126-Pro260, Phe264-Val259, His320-Pro318	-	Tyr126-Pro260, Gly137-Ile258	-	-	Tyr263-Phe123
MMV023860	Gly137-Pro260, Tyr263-Glu261	-	-	-	-	Tyr263-Phe123
MMV024101	Tyr263-Phe123, Gly137-Ile141	Ala122-Tyr263, Phe264-Ile119, Tyr263-Ile119, Tyr263-Phe123	Ala122-Tyr263, Phe264-Ile119, Tyr263-Ile119, Tyr263-Phe123	-	-	-
MMV024937	Gly137-Pro260, Tyr263-Phe123, Ala122-Tyr263, Tyr263-Ile119	Gly137-Pro260, Tyr263-Phe123	Gly137-Pro260, Tyr263-Phe123	-	-	-
MMV676050		Ala122-Tyr263	Gly137-Ile258	-	-	-
MMV676599	Tyr126-Glu261, Gly137-Pro260, Tyr263-Glu261	Gly137-Pro260, Gly137-Ile1258	Phe264-Ile119, Tyr126-Pro260	-	-	-
MMV687700	Tyr126-Ala138, Gly137-Pro260	Ala122-Tyr263, Gly137-Pro260, Tyr263-Ile119	-	-	Tyr263-Ser83	-
SANC00257	Gly137-Pro260, Ala122-Tyr263, Tyr263-Ile119, Tyr126-Glu261, Gly137-Ile141, Gly137-Val140, His320-Pro318	-	Tyr126-Pro260, Ala122-Tyr263, Tyr263-Ile119	-	-	-
SANC00264	Ala122-Tyr263, Gly137-Ile141, Gly137-Val140, Tyr263-Ile119	-	Ala122-Tyr263, Tyr263-Ile119	-	-	-
SANC00347	Tyr126-Pro260, Phe264-Val259	-	-	-	-	-
SANC00559	Tyr126-Ala138	Ala122-Tyr263, Gly137-Ile258	Gly137-Pro260, Ala122-Tyr263	-	Tyr263-Ser83	-
SANC00587	Ala122-Tyr263, Gly137-Ile141	Phe264-Phe123, Ala122-Tyr263	-	-	-	-
SANC00595	Tyr126-Glu261, Ala122-Tyr263	Phe264-Ile119, Phe264-Phe123	-	-	-	-