



**The Characterization of GTP Cyclohydrolase I and
6-Pyruvoyl Tetrahydropterin Synthase Enzymes as
Potential Anti-Malarial Drug Targets**

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

DOCTOR OF PHILOSOPHY

IN BIOINFORMATICS

Rhodes University, South Africa

Department of Biochemistry and Microbiology

Faculty of Science

by

Afrah Yousif Huseein Khairallah

ORCID ID: <https://orcid.org/0000-0003-4838-7032>

March 2022

Abstract

Malaria remains a public health problem and a high burden of disease, especially in developing countries. The unicellular protozoan malaria parasite of the genus *Plasmodium* infects about a quarter of a billion people annually, with an estimated 409 000 death cases. The majority of malaria cases occurred in Africa; hence, the region is regarded as endemic for malaria. Global efforts to eradicate the disease led to a decrease in morbidity and mortality rates. However, an enormous burden of malaria infection remains, and it cannot go unnoticed. Countries with limited resources are more affected by the disease, mainly on its public health and socio-economic development, due to many factors besides malaria itself, such as lack of access to adequate, affordable treatments and preventative regimes.

Furthermore, the current antimalarial drugs are losing their efficacy because of parasite drug resistance. The emerged drug resistance has reduced the drug efficacy in clearing the parasite from the host system, causing prolonged illness and a higher risk of death. Therefore, the emerged antimalarial drug resistance has hindered the global efforts for malaria control and elimination and established an urgent need for new treatment strategies.

When the resistance against classical antimalarial drugs emerged, the class of antifolate antimalarial medicines became the most common alternative. The antifolate antimalarial drugs target the malaria parasite *de novo* folate biosynthesis pathway by limiting folate derivatives, which are essential for the parasite cell growth and survival. Yet again, the malaria parasite developed resistance against the available antifolate drugs, rendering the drugs ineffective in many cases. Given the previous success in targeting the malaria parasite *de novo* folate biosynthesis pathway, alternative enzymes within this pathway stand as good targets and can be explored to develop new antifolate drugs with novel mechanisms of action.

The primary focus of this thesis is to contribute to the existing and growing knowledge of antimalarial drug discovery. The study aims to characterise the malaria parasite *de novo* folate synthesis pathway

enzymes guanosine-5'-triphosphate (GTP) cyclohydrolase I (GCH1) and 6-pyruvoyl tetrahydropterin synthase (PTPS) as alternative drug targets for malaria treatment by using computational approaches. Further, discover new allosteric drug targeting sites within the two enzymes' 3D structures for future drug design and discovery. Sequence and structural analysis were carried out to characterise and pinpoint the two enzymes' unique sequence and structure-based features. From the analyses, key sequence and structure differences were identified between the malaria parasite enzymes relative to their human homolog; the identified sites can aid significantly in designing and developing new antimalarial antifolate drugs with good selectivity toward the parasites' enzymes.

GCH1 and PTPS contain a catalytically essential metal ion in their active site; therefore, force field parameters were needed to study their active sites accurately during all-atom molecular dynamic simulations (MD). The force field parameters were derived through quantum mechanics potential energy surface scans of the metals bonded terms and evaluated via all-atom MD simulations.

Proteins structural dynamics is imperative for many biological processes; thus, it is essential to consider the structural dynamics of proteins whilst understanding their function. In this regard, the normal mode analysis (NMA) approach based on the elastic network model (ENM) was employed to study the intrinsic dynamics and conformational changes of GCH1 and PTPS enzymes. The NMA disclosed essential structural information about the protein's intrinsic dynamics and mechanism of allosteric modulation of their binding properties, further highlighting regions that govern their conformational changes. The analysis also disclosed hotspot residues that are crucial for the proteins' fold stability and function. The NMA was further combined with sequence motif results and showed that conserved residues of GCH1 and PTPS were located within the identified key structural sites modulating the proteins' conformational rearrangement. The characterized structural features and hotspot residues were regarded as potential allosteric sites of important value for the design and development of allosteric drugs.

Both GCH1 and PTPS enzymes have never been targeted before and can provide an excellent opportunity to overcome the antimalarial antifolate drug resistance problem. The data presented in

this thesis contribute to the understanding of the sequence, structure, and global dynamics of both GCH1 and PTPS, further disclose potential allosteric drug targeting sites and unique structural features of both enzymes that can establish a solid starting point for drug design and development of new antimalarial drugs of a novel mechanism of actions. Lastly, the reported force field parameters will be of value for MD simulations for future *in-silico* drug discovery studies involving the two enzymes and other enzymes with the same Zn^{2+} binding motifs and coordination environments. The impact of this research can facilitate the discovery of new effective antimalarial medicines with novel mechanisms of action.

Declaration

I, Afrah Yousif Huseein Khairallah, declare that this thesis is my own original work. The thesis is being submitted for the degree of Doctor of Philosophy at Rhodes University and has not previously been submitted before for any degree or examination at any other institution.



Signature

10/10/2021

Date

Acknowledgement

I would like to extend my heartfelt thanks and deep appreciation to my research supervisor Prof. Özlem Tastan Bishop, who has constantly encouraged me to pursue my research with utmost enthusiasm. Her continuous guidance, constructive feedback, and support have always pushed me to do my research with perseverance. I extend my gratitude to my co-supervisor, Dr Vuyani Moses, for his guidance, valuable assistance, and encouragement. I express my sincere appreciation to all my colleagues from the research unit in bioinformatics (RUBi) for their support, constructive discussions, and creating a homely environment to rejoice. It was indeed an excellent learning opportunity in this lab. I thank my parents and all family members, friends, and well-wishers for their constant support throughout my study and their sacrifices and blessings that have always encouraged me to take my further steps.

Acknowledgement of funding

This work was supported by the **Organization for Women in Science for the Developing World (OWSD) and the Swedish International Development Cooperation Agency (SIDA)**. Fund Reservation No. 3240287275. Thank you for granting me the opportunity to pursue my postgraduate studies and thrive for excellence.

Dedication

~This thesis is dedicated to my beloved Father ~

Table of Contents

Abstract	ii
Declaration	v
Acknowledgement	vi
Dedication	vii
Table of Contents	viii
List of Figures	xiv
List of Tables	xxiii
Research Outputs	xxv
PRIMARY PUBLICATIONS.....	XXV
SECONDARY PUBLICATIONS.....	XXV
CONFERENCE PRESENTATIONS.....	XXV
AWARDS.....	XXVI
Thesis overview	xxvii
Chapter 1	1
CHAPTER OVERVIEW.....	1
1.1 INTRODUCTION.....	2
1.1.1 <i>Malaria</i>	2
1.1.2 <i>The malaria parasite life cycle</i>	3
1.1.3 <i>The symptoms</i>	5
1.1.4 <i>Classical antimalarial drugs and their mechanism of action</i>	5
Aryl aminoalcohol compounds.....	6
Artemisinins compounds.....	6
Antifolate compounds.....	7
Atovaquone.....	7
1.1.5 <i>The current state of the fight against malaria</i>	8
1.1.6 <i>Malaria vaccines</i>	8
1.1.7 <i>The de novo folate synthesis pathway</i>	9
1.1.8 <i>Antifolate drug resistance</i>	12

1.1.9	<i>Alternative metabolic targets: the search for new antifolates</i>	13
1.1.10	<i>Enzymes of the tunnelling fold superfamily as alternative antifolate drug targets</i>	14
	Guanosine-5'-triphosphate cyclohydrolase	14
	Guanosine-5'-triphosphate cyclohydrolase 3D structure	15
	GCH1 feedback regulatory protein	16
	GCH1 as a potential drug target	17
	6-pyruvoyl tetrahydropterin synthase as a potential drug target	19
1.2	KNOWLEDGE GAP	20
1.3	STUDY RATIONALE	21
1.4	RESEARCH AIM AND OBJECTIVES	21
	Aim	21
	Main objectives	21
Chapter 2		23
CHAPTER OVERVIEW		23
2.1	INTRODUCTION	24
2.1.1	<i>Database similarity search tool</i>	24
2.1.2	<i>Sequence alignment strategies and algorithms</i>	25
2.1.3	<i>Alignment scoring schemes</i>	26
2.1.4	<i>Multiple Sequence Alignment</i>	27
	Progressive approach	27
	Iterative approach	28
	Block-based approaches	28
2.1.5	<i>Phylogenetic analysis</i>	29
2.1.6	<i>Motif discovery</i>	30
2.2	METHODS	31
2.2.1	<i>Sequence retrieval</i>	31
2.2.2	<i>Multiple sequence alignment</i>	31
2.2.3	<i>Motif discovery</i>	31
2.2.4	<i>Phylogenetic analysis</i>	32
2.3	RESULTS AND DISCUSSION	33
2.3.1	<i>Sequence retrieval</i>	33
2.3.2	<i>Multiple sequence alignment</i>	34
	GCH1 MSA	35
	PTPS MSA	38
2.3.3	<i>Pair-wise sequence identity</i>	39
2.3.4	<i>Phylogenetic analysis</i>	40

2.3.5	<i>Motif analysis</i>	42
	GCHI motif analysis.....	43
	PTPS motif analysis.....	45
2.4	CONCLUSIONS	47
Chapter 3		48
CHAPTER OVERVIEW		49
3.1	INTRODUCTION.....	50
3.1.1	<i>Metalloproteins</i>	50
3.1.2	<i>Methods of molecular energy calculation</i>	51
	Quantum Mechanics	51
	Molecular Mechanics.....	53
3.1.3	<i>Force fields</i>	55
3.1.4	<i>Force field parametrization of metal ions</i>	56
3.2	METHODS	59
3.2.1	<i>Structures retrieval</i>	59
3.2.2	<i>Homology modelling of the P. falciparum GCHI</i>	59
3.2.3	<i>Subsets selection</i>	60
3.2.4	<i>Subsets geometry optimization</i>	60
3.2.5	<i>The restrained electrostatic potential charge evaluation</i>	61
3.2.6	<i>Force field parameter determination</i>	61
3.2.7	<i>Force field parameters evaluation</i>	62
3.2.8	<i>MD simulation protocol</i>	62
3.2.9	<i>Post-MD trajectory analysis</i>	63
3.3	RESULTS AND DISCUSSION	65
3.3.1	<i>Subsets selection</i>	65
3.3.2	<i>Subsets geometry optimization</i>	67
3.3.3	<i>The restrained electrostatic potential charge evaluation</i>	68
3.3.4	<i>Force field parameter determination</i>	69
3.3.5	<i>Force field parameters evaluation</i>	73
3.3.6	<i>Metal ions RMSD</i>	75
3.3.7	<i>GCHI and PTPS reaction mechanisms</i>	79
3.3.8	<i>Validation of the force field parameters in the presence of an inhibitor compound</i>	81
3.3.9	<i>Validation of the force field parameters on the P. falciparum GCHI homology model structure</i>	
	83	
3.4	CONCLUSIONS	85

Chapter 4	86
CHAPTER OVERVIEW	87
4.1 INTRODUCTION.....	88
4.1.1 <i>Protein conformational dynamics and function</i>	88
4.1.2 <i>Protein conformational dynamics and allostery</i>	88
4.1.3 <i>Approaches for studying proteins conformational dynamics</i>	89
4.1.4 <i>Computational approaches</i>	90
MD simulations.....	90
Normal modes analysis	91
Anisotropic network model and Gaussian network model	93
4.2 METHODS	94
4.2.2 <i>Multiple sequence alignment</i>	94
4.2.3 <i>Calculation of the normal modes and atomic fluctuation</i>	95
4.2.4 <i>Visualization of the normal modes</i>	99
4.2.5 <i>Residue cross-correlation analysis</i>	99
4.2.6 <i>Deformation energy analysis</i>	99
4.2.7 <i>Mechanical stiffness analysis</i>	100
4.2.8 <i>Modes overlap analysis between the inhibitory and stimulatory GCHI-GFRP complexes</i> ...	100
4.2.9 <i>Root mean square inner product analysis</i>	101
4.3 RESULTS AND DISCUSSION	102
4.3.1 <i>NMA of the P. falciparum GCHI homology model</i>	102
4.3.2 <i>Visual characterization of the P. falciparum GCHI low-frequency non-degenerate normal modes</i>	104
4.3.3 <i>Multiple sequence alignment and NMA</i>	106
4.3.4 <i>Collective MSF analysis of the P. falciparum GCHI normal modes</i>	107
4.3.5 <i>P. falciparum GCHI hinge residues</i>	110
4.3.6 <i>Deformation energy analysis of the P. falciparum GCHI</i>	110
4.3.6 <i>Residue cross-correlation of the P. falciparum GCHI</i>	111
4.3.7 <i>NMA of R. norvegicus stand-alone GCHI structure</i>	113
4.3.8 <i>The linear combination of R. norvegicus GCHI degenerate modes</i>	114
4.3.9 <i>Visual characterization of the R. norvegicus GCHI low-frequency non-degenerate normal modes</i>	114
4.3.10 <i>MSF analysis of the R. norvegicus GCHI individual non-degenerate slow modes</i>	116
4.3.7 <i>Collective MSF analysis of the R. norvegicus GCHI lowest and highest-frequency normal modes</i>	117
4.3.8 <i>Deformation energy analysis of the R. norvegicus stand-alone GCHI structure</i>	118

4.3.9	<i>Mechanical stiffness analysis</i>	120
4.3.10	<i>NMA of GCHI in complex with its feedback regulatory protein</i>	122
4.3.11	<i>Mode overlap analysis</i>	124
4.3.12	<i>Collective MSF analysis of the R. norvegicus GCHI-GFRP complexes</i>	126
4.3.13	<i>GCHI-GFRP residues cross-correlation analysis</i>	132
4.4	CONCLUSIONS.....	134
Chapter 5		136
CHAPTER OVERVIEW.....		137
5.1	INTRODUCTION.....	138
5.2	METHODS.....	140
5.2.1	<i>Structure and sequence data retrieval</i>	140
5.2.2	<i>Calculation of the normal modes and atomic fluctuation</i>	140
5.2.3	<i>Modes contribution</i>	140
5.2.4	<i>Visualization of the normal modes</i>	140
5.2.5	<i>Residue cross-correlation analysis</i>	141
5.2.6	<i>Deformation energy analysis</i>	141
5.2.7	<i>Mechanical stiffness analysis</i>	141
5.2.8	<i>Gaussian network model calculations</i>	141
5.3	RESULTS AND DISCUSSION.....	142
5.3.1	<i>NMA of the P. falciparum PTPS</i>	142
5.3.2	<i>Modes contribution</i>	142
5.3.3	<i>Visualisation of the P. falciparum PTPS low-frequency non-degenerate normal modes</i>	143
5.3.4	<i>MSF analysis of the individual eight low-frequency non-degenerate modes</i>	146
5.3.5	<i>Deformation energy analysis</i>	147
5.3.6	<i>Residue cross-correlation analysis</i>	148
5.3.7	<i>Collective MSF analysis of the lowest and highest-frequency normal modes</i>	149
5.3.8	<i>P. falciparum PTPS hinge residues</i>	153
5.3.9	<i>Mechanical stiffness analysis</i>	155
5.4	CONCLUSIONS.....	157
Concluding Remarks and Future Perspectives		158
Appendices		163
Appendix 1		164
Appendix 2		175

Appendix 3	182
References	183

List of Figures

Figure 1.1. Key events of the malaria parasite life cycle in the human host. First, the sporozoites are injected into the host system through the mosquito bite during its blood meal. The sporozoites then travel to the liver and mature into schizonts. The schizonts rupture and release daughter merozoites into the host bloodstream. The merozoites infect the erythrocyte, then enlarge to ring stage trophozoites and progress to mature trophozoites. The mature trophozoites develop into replicative schizont forms and release multiple invasive daughter merozoites, starting a new infection cycle. A portion of the merozoites differentiates into sexual erythrocytic stages (male and female gametocytes) that can be transferred back to the female Anopheles mosquito during a blood meal to start the sexual stage in the mosquito carrier. Inoculation of the sporozoites into a new human host repeats the malaria life cycle. (Adapted from (Molina-Franky *et al.*, 2020)). 4

Figure 1.2. Structure of the folate molecule. (Adapted from (Rosenberg, 2012))..... 10

Figure 1.3. Illustration of the malaria parasite *de novo* folate biosynthesis pathway. The pathway is constituted by six enzymes: GTP cyclohydrolase (GCH1); 6-pyruvoly tetrahydropterin synthase (PTPS); 6-hydroxymethyl dihydropterin pyrophosphokinase (PPPK); dihydropteroate synthase (DHPS); dihydrofolate synthase (DHFS); dihydrofolate reductase (DHFR). Para-aminobenzoic acid (pABA) enters as a substrate for DHPS. Enzymes inhibited by the antifolate drugs are indicated by a red cross, whereas the Inhibitors of folate synthesis are shown in red at the right of the pathway. The salvaged folate enters the pathway upstream of DHFR. (Adapted from (Heinberg and Kirkman, 2015))..... 11

Figure 1.4. Coloured by chain the 3D structure of the GFRP two pentameric units that bind to both sides of the GCH1 homo decameric structure feedback inhibition (-) of GCH1 activity by BH4 (indicated by the red arrow) and feed-forward stimulation (+) by phenylalanine (indicated by the blue arrow). GCH1-GFRP complex of the *R. norvegicus* (PDB ID: 1WPL) (B) Side view of the GCH1-GFRP complex structure and (C) Top view of the GCH1-GFRP complex structure. (Reproduced with permission form (Khairallah, Ross and Tastan Bishop, 2021)). 18

Figure 1.5. Coloured by chain (A) top and (B) side views of *P. falciparum* PTPS 3D structure (PDB: 1Y13). The homo-hexameric functional assembly unit is formed by the face to face association of two trimeric units. Each trimer comprises three monomeric units that assemble via tight hydrogen bonds of the protein N- and C-terminal β -strands. 20

Figure 2.1. MSA of the GCH1 retrieved sequences. The black stars illustrate the catalytic and metal-coordinating residues. The green circles show the functionally important active site neighbouring

residues. Identical/highly conserved residues are printed in white text and highlighted in red and similar residues are printed in black text and highlighted in yellow. The blue and red dashed boxes show Motif 5 and Motif 8 of the Plasmodium sequences, respectively..... 37

Figure 2.2. MSA of the PTPS retrieved sequences. The black stars illustrate the catalytic and metal-coordinating residues. The green circles illustrate the functionally important active site neighbouring residues. Identical/highly conserved residues are printed in white text and highlighted in red and similar residues are printed in black text and highlighted in yellow. The forest green, red, lime green and purple dashed boxes shows Motif 5,6,7 and 8 of the Plasmodium sequences, respectively. 39

Figure 2.3. (A) GCH1 pair-wise sequence identity heat map. (B) PTPS pair-wise sequence identity heat map. The heat map shows the pairwise sequence identity scores of the MSA as a colour-coded matrix; the identity score increases from blue to red /least to most conserved sequences. (C) Phylogenetic tree of the retrieved GCH1 sequences. The red dashed square highlights the Plasmodium GCH1 sequences. (D) Phylogenetic tree of the retrieved PTPS sequences. The red dashed square highlights the Plasmodium PTPS sequences 42

Figure 2.4. Motif analysis of the retrieved GCH1 sequences. (A) Motif numbering and web logos. The sequence logo contains a stack of letters at every position in the motif. The height of the letters demonstrates the probability (in bits) of the letter occurring at that position multiplied by the number of times that residue occurs within that site in each motif site in the total dataset (Bailey *et al.*, 2015). Residues are coloured based on their chemical properties: blue–hydrophobic residues (A, C, F, I, L, M, V and W); red–positively charged residues (K and R); green–polar, non-charged and non-aliphatic residues (N, Q, S and T); magenta–most acidic residues (D and E); light pink, orange, turquoise and yellow are for H, G, Y and P respectively. (B) MEME heat map summarizing motif information for GCH1 homologue sequences. The white regions show sequences lacking a motif, and the level of conservation increases from blue to red. (C) The identified unique motifs of the *P. falciparum* GCH1 mapped onto its 3D structures. The sequence logos of the two unique motifs are shown next to the 3D structure. 44

Figure 2.5. Motif analysis of the retrieved PTPS sequences. (A) Motif numbering and web logos. The sequence logo contains a stack of letters at every position in the motif. The height of the letters demonstrates the probability (in bits) of the letter occurring at that position multiplied by the number of times that residue occurs within that site in each motif site in the total dataset (Bailey *et al.*, 2015). Residues are coloured based on their chemical properties: blue–hydrophobic residues (A, C, F, I, L, M, V and W); red–positively charged residues (K and R); green–polar, non-charged and non-aliphatic residues (N, Q, S and T); magenta–most acidic residues (D and E); light pink, orange, turquoise and yellow are for H, G, Y and P respectively. (B) MEME heat map summarizing motif information for

GH1 homologue sequences. The white regions show sequences lacking a motif, and the level of conservation increases from blue to red. (C) The identified unique motifs of the *P. falciparum* PTPS mapped onto its 3D structures. The sequence logos of the two unique motifs are shown next to the 3D structure. 46

Figure 3.1. The potential energy of a molecule is described as the sum of its bonded and non-bonded terms. The total potential energy (E_{tot}) is decomposed as follows: Energy required to stretch/compress a covalent bond (E_{str}), the energy required to bend a bond from its equilibrium angle (E_{bend}), the energy required to rotate a planar group of atoms from its equilibrium dihedral angle (E_{tor}). Lastly, the energy contribution of interactions between unbound atoms, the latter constitutes Van der Waals interactions (E_{vdw}): the steric exclusion, this potential is attractive at long distance, but becomes repulsive to prevents atoms from overlapping with each other, lastly the electrostatic interactions of charges (E_{elec}). 54

Figure 3.2. A flowchart illustrating the steps followed to derive and evaluate the force field parameters of both GCH1 and PTPS Zn^{+2} ions..... 64

Figure 3.3. A diagram illustrating the overall X-ray crystal structures and the active sites. (A) GCH1 of the *T. thermophilus* (strain HB8 / ATCC 27634 / DSM 579) (PDB ID: 1WUR) (B) PTPS of the *P. falciparum* (PDB ID: 1Y13). The boxed zoom shows the active site metal ions as grey spheres, coordinating residues in the sticks, coordinating water molecules in red spheres, and the bonds in yellow dashes. (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020)). ... 66

Figure 3.4. (A) atoms of the GCH1 active sites subset (left) and corresponding partial charges (right). (B) atoms of the PTPS active site subset (left) and corresponding partial charges (right). (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020)). 69

Figure 3.5. Potential energy profiles of the GCH1 Zn^{2+} ion coordinating residues. The energy profiles are shown in black dots, and the theoretical MM fitting curve is shown as a red line. The QM PES scans were performed for 10 steps with an increment of 0.05 Å in the two directions forward and reversed (stretches and contracts) for three bonds A: Zn^{2+} -His-ND1(111), B: Zn^{2+} -Cys-SG(108), and C: Zn^{2+} Cys-SG(179). 10 steps with an increment of 1 degree for three angles, D: Cys-SG(108)- Zn^{2+} -Cys-SG(179), E: Cys-SG(108)- Zn^{2+} -His-ND1(111) and F: Cys-SG(179)- Zn^{2+} -His-ND1(111), G: Zn^{2+} -His-ND1(111)-His-CE1(111), H: Zn^{2+} -His-ND1(111)-His-CG(111), I: Zn^{2+} -Cys-SG(108)-Cys-CB(108), J: Zn^{2+} -Cys-SG(179)-Cys-CB(179)) and 90 steps with an increment of 1 degree for two dihedral angles (K: Cys-SG(108)- Zn^{2+} -His-ND1(111)-His-CE1(111) and L: Cys-SG(179)- Zn^{2+} -His-ND1(111)-His-CG(111)). (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020)). 70

Figure 3.6. Potential energy profiles of the PTPS Zn^{2+} ion coordinating residues. The energy profiles are shown in black dots, and the theoretical MM fitting curve is shown with a red line. The QM PES scans were performed for 10 steps with an increment of 0.05 Å in the two directions forward and reversed for three bonds (A: Zn^{2+} -His-NE2(43), B: Zn^{2+} -His-NE2(41), and C: Zn^{2+} -His-NE2(29)), 10 steps with an increment of 1 degree for three angles (D: His-NE2(29)- Zn^{2+} -His-NE2(43), E: His-NE2(29)- Zn^{2+} -His-NE2(41) and F: His-NE2(41)- Zn^{2+} -His-NE2(43)) and 90 steps with an increment of 1 degree for three dihedral angles (G: His-NE2(43)- Zn^{2+} -His-NE2(29)-His-CD2(29), H: His-NE2(43)- Zn^{2+} -His-NE2(41)-His-CD2(41) and I: His-NE2(29)- Zn^{2+} -His-NE2(41)-His-CD2(41)). (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020))..... 71

Figure 3.7. Assessment of the GCH1 stability by (A) RMSD, (B) RMSF, and (C) Rg. The black dashed lines separate each chain of the protein. The dashed red circles highlight the location of the active site residues responsible for the metal's coordination. (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020))..... 74

Figure 3.8. Assessment of the PTPS stability by (A) RMSD, (B) RMSF, and (C) Rg. The black dashed lines separate each chain of the protein. The dashed red circles highlight the location of the active site residues responsible for metal coordination. (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020))..... 75

Figure 3.9. RMSD values of the Zn^{2+} ions of (A) GCH1 (PDB ID: 1WUR) and (B) PTPS (PDB ID: 1Y13) over the 100 ns MD simulations. (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020)). 76

Figure 3.10. (A) GCH1 Zn^{2+} ion coordination distances during the 100 ns MD simulations. The black line represents the coordination distance of Zn^{2+} -His-ND1(111), the red line represents the coordination distance of Zn^{2+} -Cys-SG(108), and the green line represents the coordination distance of Zn^{2+} -Cys-SG(179). (B) PTPS Zn^{2+} ion coordination distances. The black line represents the coordination distance of Zn^{2+} -His-NE2(29), the red line represents the coordination distance of Zn^{2+} -His-NE2(41), and the green line represents the bond distance of Zn^{2+} -His-NE2(43). (C) GCH1 Zn^{2+} ion coordination angles during the 100 ns MD simulations. The black line represents the coordination angle of Cys-SG(108)- Zn^{2+} -His-ND1(111); the green line represents the coordination angle of Cys-SG(197)- Zn^{2+} -His-ND1(111), and the red line represents the coordination angle of Cys-SG(108)- Zn^{2+} -Cys-SG(179). (D) PTPS Zn^{2+} ion coordination angles. The red line represents the coordination angle of His-NE2(41)- Zn^{2+} -His-NE2(29), the green line represents the coordination angle of His-NE2(41)- Zn^{2+} -His-NE2(43), and the black line represents the coordination angle of His-NE2(29)- Zn^{2+} -His-NE2(43). (E) GCH1 Zn^{2+} ion coordination dihedral angles during the 100 ns MD simulations. The red line represents the dihedral angle of Cys-SG(108)- Zn^{2+} -His-ND1(111)-His-CE1(111); the black line represents the dihedral angle of

Cys-SG(108)-Zn²⁺-His-ND1(111)-His-CG(111). (F) PTPS Zn²⁺ ion coordination angles. The red line represents the dihedral angle of His-NE2(43)-Zn²⁺-His-NE2(41)-His-CD2(41), the green line represents the coordination angle of His-NE2(43)-Zn²⁺-His-NE2(29)-His-CD2(29), and the black line represents the dihedral angle of His-NE2(29)-Zn²⁺-His-NE2(41)-His-CD2(41). (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020))..... 77

Figure 3.11. Density histogram illustrating the distribution and occurrence of the Zn²⁺ ion coordination distances during the 100 ns MD simulations of (A) GCH1 and (B) PTPS. The red line is an overlaying density line drawn to show the kernel density of the histograms. (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020)). 79

Figure 3.12. MD snapshots of the last frame from both active sites. The black dashed circles show the water molecules arriving at the active sites and completing the tetrahedral geometry of the (A) GCH1 and (B) PTPS Zn²⁺ ions. (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020))..... 81

Figure 3.13. Dynamic properties of both the apo form (Black) and ligand-bound complexes over a simulation period of 100 ns. (A) The global stability of the different systems as determined by RMSD (B) The compactness of the apo system (Black) and ligand-bound complexes by Rg. (C) local fluctuation profile of the individual residues by RMSF. The Apo system is shown in black. (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020))..... 82

Figure 3.14. Snapshot of the last MD frame showing a docked ligand interaction with the Zn²⁺. The red circle highlights the ion-induced perturbation of the water hydrogen bonds resulting in a free hydroxyl that we believe is responsible for attacking the substrate C atoms. The Zn²⁺ interaction with the ligand hydroxyl group was the main reason for stabilizing the substrate throughout the MD simulations. (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020)). 83

Figure 4.1. (A) The RMSIP of the first 20 non-trivial modes subspaces (left) and the first 100 non-trivial modes subspaces from both *T. thermophilus* GCH1 crystal structure (PDB ID:1WUR) and *P. falciparum* GCH1 homology model structure (B) MSF profiles of the 20 lowest-frequency normal modes of *T. thermophilus* GCH1 crystal structure (PDB ID:1WUR) and *P. falciparum* GCH1 homology model. The grey blocks partition each chain of the GCH1 homo-decameric structure. 103

Figure 4.2. (A) The characterized global motions of the *P. falciparum* GCH1 homology model structure within the four non-degenerate low-frequency modes. Each arrow denotes the protein C_β atoms vector of movement. Each of the protein chains is coloured differently for illustration purposes. (B) Individual MSF profiles of the four non-degenerate low-frequency modes, from top to bottom

Mode 7, 12, 15, and 18. Residues of a notable fluctuation are labelled on the MSF plot of the one GCH1 chain. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2021)). 106

Figure 4.3. (A) Multiple sequence alignment of the *P. falciparum*, *D. melanogaster*, *H. sapiens*, *R. norvegicus*, *E. coli*, and *T. thermophilus* GCH1 sequences. The black stars show the catalytic and metal coordinating residues. The black block shows the loop region surrounding the active site pocket. The blue blocks show the BH4 binding site, and the green block shows the GFRP binding site. One chain representation of the *P. falciparum* GCH1 (B) 20 Lowest-frequency and (C) 20 highest-frequency normal modes MSF profiles. The solid red circle highlights the active site region. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2021)). 109

Figure 4.4. (A) MSF values over the 20 lowest-frequency modes mapped onto the *P. falciparum* GCH1 structure. The fluctuations colour scale is blue to red (low to high) atomic fluctuation (B) Atomic deformation energy values mapped onto the *P. falciparum* GCH1 structure. The deformation energy colour scale is blue to red (low to high) deformation energy..... 111

Figure 4.5. A heatmap showing GCH1 residue cross correlation. Correlated regions (red) show the residues moving in the same direction, e.g. upward and upward, and anti-correlated (blue) residues moving in directions that are linearly opposite to each other, e.g. upward and downward. The high diagonal values occur where $i = j$, in this case, the same atom..... 112

Figure 4.6. (A) The characterized global motions of the *R. norvegicus* stand-alone GCH1 structure within the four non-degenerate low-frequency modes. The obtained eigenvectors were projected as a set of arrows that denote the direction of motion of the protein C_{β} atoms. Each of the protein chains is coloured differently for illustration purposes. (B) Individual MSF profiles of the four non-degenerate low-frequency modes, from top to bottom Mode 7, 12, 15, and 18. Residues of a notable fluctuation are labelled on the MSF plot of the one GCH1 chain. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2021)). 117

Figure 4.7. Collective MSF profile of the (A) 20 lowest-frequency and (B) 20 highest-frequency normal modes. The grey blocks partition each chain of the GCH1 homo-decameric 3D structure. One chain representation of the *R. norvegicus* stand-alone GCH1 of the (C) 20 lowest-frequency and (D) 20 highest-frequency MSF profiles. The solid red circle highlights the active site catalytic residues, and the dashed red circle highlights neighbouring residues that form the active site pocket. The black arrow points towards the central five-helix bundle region. (E) Experimental *B*-factor of the *R. norvegicus* GCH1 crystal structure (PDB ID:1WPL). (F) MSF values over the 20 lowest-frequency modes mapped onto the *R. norvegicus* GCH1 structure. The fluctuations colour scale is from blue to red (low to high) atomic fluctuation. (G) Atomic deformation energies: the structure is coloured such

that atoms of the lowest deformation energy are in blue, and atoms of highest deformation energy are in red. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2021)). 119

Figure 4.8. (A) ProDy mechanical stiffness heatmap of the *R. norvegicus* stand-alone GCH1. The colour bar illustrates the effective force constant values (blue as a strong /stiff pair of interactions and red as a weak/flexible pair of interactions). (B) The mean value of the effective force constant of residue pairs, the GCH1 secondary structure, is shown along the upper abscissa. The mechanically stiff/rigid regions are shown in blue, representing the GCH1 20 stranded β -barrel. The black circle highlights the residues participating in forming the active site pocket, whereas the red circles indicate the residues of the BH4 binding site. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2021))..... 121

Figure 4.9. (A) The characterized global motions of the *R. norvegicus* GCH1-GFRP structure (PDB ID: 1WPL). Each arrow denotes the protein C_{β} atoms vector of motion. The obtained eigenvectors were projected as a set of arrows that indicate the direction of movement of the protein C_{β} atoms. Each of the protein chains is coloured differently for illustration purposes. (B) Individual MSF profiles of the six non-degenerate low-frequency modes, from top to bottom Mode 3, 4, 7, 12, 26, and 77. The grey blocks represent each chain of the GCH1 structure, and the yellow blocks represent each chain of the GFRP. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2021)). 124

Figure 4.10. Superimposition of the *R. norvegicus* GCH1-GFRP inhibitory complex (PDB ID:1WPL) and the *R. norvegicus* GCH1-GFRP stimulatory complex (PDB ID: 1IS7). The green arrows highlight structural differences, which are noted by an inward contraction of the protein complex as it transitions from stimulatory to inhibitory form. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2021)). 125

Figure 4.11. (A) From left to right: the experimental *B*-factors of the GCH1-GFRP crystal structure of the *R. norvegicus*, atomic fluctuation within the 1% lowest-frequency modes: regions of the lowest MSF are shown in blue. Higher MSF in green, yellow, and red and the atomic deformation energies mapped onto the structure and coloured such that the most deformed regions are in red, and the least deformed are in blue. The arrows show the wringing motion of the terminal helices. (B) MSF profiles over the 1% lowest-frequency modes of both complexes. The grey blocks partition each chain of the GCH1 homo-decameric 3D structure, and the orange blocks partition each chain of the GFRP protein. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2021)). 128

Figure 4.12. (A) The structure of the GCH1-GFRP stimulatory complex from *R. norvegicus* (PDB ID: 1IS7), the zoomed out blue box, shows the GFRP phenylalanine (Phe) binding site residues, and

the zoomed out black box shows the GCH1 BH4 binding site residues. (B) MSF profiles over the 1% lowest-frequency normal modes for both GCH1-GFRP inhibitory (blue) and stimulatory (orange) complexes. (C) MSF profiles over the 1% highest-frequency normal modes for both GCH1-GFRP inhibitory (blue) and stimulatory (orange) complexes. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2021)). 131

Figure 4.13. (A) A heatmap representation of the GCH1-GFRP complex residue pair-wise cross-correlation heatmap. The scale colour bar on the right indicates the extent of the correlation in which the red colour highlights correlated motions (residue pairs moving together in the same direction), while the blue colour highlights the anti-correlated movements (residue pairs moving in the opposite direction). The high diagonal values are occurring where $i = j$ shows the correlation within the same atom. The grey blocks partition each chain of the proteins. (B) The rotation of GCH1 in an opposing direction to the GFRP as identified in **Mode_{IC} 3**. (C) The inward contraction of GCH1 and the outward expansion/expelling of the GFRP as identified in **Mode_{IC} 4**. The black arrows show the GCH1 direction of motion, and the white arrows represent the GFRP direction of movement. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2021)). 133

Figure 5.1. The characterized global motions of the *P. falciparum* PTPS structure within the eight slowest non-degenerate modes. The obtained eigenvectors were projected as a set of arrows that denote the direction of motion of the protein C_{β} atoms. Each of the protein chains is coloured differently for illustration purposes. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2020)). 144

Figure 5.2. (A) The experimental *B*-factors mapped onto the *P. falciparum* PTPS structure (PDB ID: 1Y13) (B) Atomic mobility over the 20 lowest-frequency normal modes. Regions of the lowest fluctuation are shown in blue, with higher mobility in green, yellow, and red. The figure illustrates the high mobility of the protein terminal regions, central helices, and the rigidity of the protein core, indicating that the NMA has captured most of the experimentally determined atomic fluctuation. .. 147

Figure 5.3. Deformation energy values mapped onto the *P. falciparum* PTPS structure (A) top view (B) side view. The colours of the atoms indicate the amount of deformation. The dark blue sites are the least deformed/most mobile, whereas red areas are strongly deformed/ rigid. The highest deformation energy was observed within the protein tunnel region (red). 148

Figure 5.4. A heatmap representation of the *P. falciparum* PTPS residues pairwise cross-correlation. The scale colour bar on the right indicates the extent of the correlation in which the red colour highlights correlated motions (residue pairs moving together in the same direction), while the blue colour highlights the anti-correlated movements (residue pairs moving in the opposite direction). The

yellow blocks partition each chain of the protein. (Reproduced with permission form (Khairallah, Ross and Tastan Bishop, 2020)). 149

Figure 5.5. (A) The *P. falciparum* PTPS sequence motifs mapped onto its structure (right), the location of high-frequency vibrating residues within the structure is illustrated by the black arrows. The MEME sequence logo contains a stack of amino acids letters at every position in the motif (left). The black arrows indicate the location of the high-frequency vibrating residues within the motifs. MSF profiles of (B) 20 lowest-frequency and (C) 20 highest-frequency normal modes. Residues of a notable fluctuation within the slowest modes and the fastest modes are labelled on each protein chain. The dark green, purple, red, and lime green bars highlight the location of motifs 5, 6, 7, and 8, respectively. (Reproduced with permission form (Khairallah, Ross and Tastan Bishop, 2020)). 151

Figure 5.6. (A) The structure of *P. falciparum* PTPS and the metal centre (active site) in the circled zoom (B) boxed zoom showing the location of the identified high-frequency vibrating residues. (Reproduced with permission form (Khairallah, Ross and Tastan Bishop, 2020)). 153

Figure 5.7. GNM based identification of global hinge sites over the 20 lowest-frequency modes. Hinge residues are located at the crossover line where the eigenvectors' values are equal to zero. Residues surrounding the hinge residues fluctuated in the positive and negative direction of motion. The black arrows illustrate the identified hinge residues. K97 (indicated by a red arrow) is an example of a residue identified of notable fluctuation within the MSF profile of the 20-lowest-frequency normal modes. Thus, it contributes to driving the global protein motion and does not act as a hinge site. (Reproduced with permission form (Khairallah, Ross and Tastan Bishop, 2020)). 154

Figure 5.8. Colour-coded representation of the *P. falciparum* PTPS structure based on the mobility of the residues over the (A) 20 lowest-frequency and (B) 20 highest-frequency GNM modes. The colour scale varies from (blue) as the most rigid to (red) as most mobile sites. The boxed zoom shows the GNM kinetically hot residues/hotspots. The analyses and figures were obtained from the DynOmics portal (Li *et al.*, 2017). (Reproduced with permission form (Khairallah, Ross and Tastan Bishop, 2020)). 155

Figure 5.9. (A) ProDy mechanical stiffness heatmap of the *P. falciparum* PTPS. The colour bar illustrates the effective force constant values (blue as a strong /stiff pair of interactions and red as a weak/flexible pair of interactions). (B) The mean value of the effective spring constant over all pairs for each residue, the protein secondary structures, are shown along the upper abscissa. (D) A cartoon representation of a single chain of *P. falciparum* PTPS with the residues of the strongest interaction mapped onto the structure and shown as red spheres. (Reproduced with permission form (Khairallah, Ross and Tastan Bishop, 2020)). 156

List of Tables

Table 2.1. GCH1 sequences retrieval information	33
Table 2.2. PTPS sequences retrieval information	34
Table 3.1. GCH1 and PTPS Zn ²⁺ ion coordination bond length and angle values from the initial experimental crystal structures, QM optimized, and post-MD structures.....	67
Table 3.2. The derived force field parameters of both GCH1 and PTPS active site Zn ²⁺ ions.....	72
Table 4.1. Low-frequency normal modes of the <i>P. falciparum</i> GCH1 homology model, associated eigenvalues, and degeneracy level. The first six trivial modes were excluded, and therefore Mode 1 represents the computed Mode 7.	105
Table 4.2. Low-frequency normal modes of the stand-alone GCH1, associated eigenvalues, and degeneracy level. The first six trivial modes were excluded, and therefore Mode 1 signifies Mode 7.	113
Table 4.3. Low-frequency normal modes of the stand-alone GCH1, GCH1-GFRP inhibitory complex, 1WPL and GCH1-GFRP stimulatory, 1IS7 and their associated eigenvalues, and degeneracy level. The first six trivial modes were excluded, and therefore Mode 1 represents Mode 7. The subscripted IC and SC stand for GCH1-GFRP inhibitory complex and stimulatory complex, respectively.....	122
Table 4.4. The overlap between the predicted displacement of the inhibitory complex (PBD ID: 1WPL) in the normal modes and the experimental conformation of the stimulatory complex (PDB ID:1IS7) crystal structure. Notably, the overlap values within the identified four modes were rather small and can be reasoned to the large size of the complex and its extensive degeneracy.....	126
Table 5.1. PTPS first 20 non-trivial normal modes, associated eigenvalues, and level of degeneracy. The first six trivial modes were excluded; therefore, mode 1 represents mode 7.....	143

The 20 amino acids that makeup proteins and their respective abbreviations

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

Similarity amino acid grouping

***Aromatic: F Y W H**

***Aliphatic: V I L**

***Charged positive: R K H**

***Charged negative: D E**

***Small: S T A G C**

***Polar (Not charged): N Q T S Y C**

Research Outputs

Primary Publications

1. Afrah Khairallah, Caroline J. Ross, Özlem Tastan Bishop. 2021. “**GTP Cyclohydrolase I as a Potential Drug Target: New Insights into Its Allosteric Modulation via Normal Mode Analysis.**” *Journal of Chemical Information and Modeling* 61 (9), 4701-4719. <https://doi.org/10.1021/acs.jcim.1c00898>
2. Afrah Khairallah, Caroline J. Ross, Özlem Tastan Bishop. 2020. “**Probing the Structural Dynamics of the *Plasmodium falciparum* Tunneling-Fold Enzyme 6-Pyruvoyl Tetrahydropterin Synthase to Reveal Allosteric Drug Targeting Sites.**” *Frontiers in Molecular Biosciences* 7, 1-17. <https://doi.org/10.3389/fmolb.2020.575196>
3. Afrah Khairallah, Özlem Tastan Bishop, and Vuyani Moses. 2020. “**AMBER Force Field Parameters for the Zn (II) Ions of the Tunneling -Fold Enzymes GTP Cyclohydrolase I and 6-Pyruvoyl Tetrahydropterin Synthase.**” *Journal of Biomolecular Structure and Dynamics*. 39 (16), 5843-5860. <https://dx.doi.org/10.1080/07391102.2020.1796800>

Secondary Publications

4. Chebon-Bore, Lorna, Taremekedzwa Allan Sanyanga, Colleen Varaidzo Manyumwa, Afrah Khairallah, and Özlem Tastan Bishop. 2021. “**Decoding the Molecular Effects of Atovaquone Linked Resistant Mutations on *Plasmodium falciparum* Cytb-Isp Complex in the Phospholipid Bilayer Membrane.**” *International Journal of Molecular Sciences* 22 (4), 1–32. <https://doi.org/10.3390/ijms22042138>
5. Amamuddy, Olivier Sheik, Wayde Veldman, Colleen Manyumwa, Afrah Khairallah, Steve Agajanian, Odeyemi Oluyemi, Gennady M. Verkhivker, and Özlem Tastan Bishop. 2020. “**Integrated Computational Approaches and Tools for Allosteric Drug Discovery.**” *International Journal of Molecular Sciences* 21 (3), 847. <https://doi.org/10.3390/ijms21030847>

Conference Presentations

1. Afrah Khairallah, Vuyani Moses and Özlem Tastan Bishop. **The Characterization of the Plasmodial GTP Cyclohydrolase I Enzyme as a Potential Antimalarial Drug Target Using Computational Approaches.** The 27th Conference on Intelligent systems for Molecular Biology and the 18th European Conference on Computational Biology, (2019). Basel, Switzerland

2. Afrah Khairallah, Vuyani Moses and Özlem Tastan Bishop. **The Characterization of the Plasmodial GTP Cyclohydrolase I Enzyme as a Potential Antimalarial Drug Target Using Computational Approaches.** The International Society for Computational Biology, Student Council Symposium, (2019). Basel, Switzerland
3. Afrah Khairallah, Özlem Tastan Bishop, and Vuyani Moses. **CHARMM Force Field Parameters for the Zn²⁺ Centres of the 6-Pyruvoyl Tetrahydropterin Synthase Enzyme.** ISC2019 high-performance Conference, (2019). Frankfurt Main, Germany
4. Afrah Khairallah, Özlem Tastan Bishop, and Vuyani Moses. **CHARMM Force Field Parameters for the Zn²⁺ Centres of the Malarial Enzyme: 6-Pyruvoyl Tetrahydropterin Synthase.** The Centre for High-Performance Computing (CHPC) National Conference, (2019). Johannesburg, South Africa
5. Afrah Khairallah, Özlem Tastan Bishop, and Vuyani Moses. **Generation of CHARMM Force Field Parameters for the Zn²⁺ Centres of the Malarial Folate Pathway Enzyme: 6-Pyruvoyl Tetrahydropterin Synthase.** The South African Medical Research Council (SAMRC) Symposium, (2019). East London, South Africa
6. Afrah Khairallah, Caroline J. Ross, Özlem Tastan Bishop. **Normal Mode Analysis of the Malaria Parasite GTP Cyclohydrolase I Enzyme.** Rhodes University Interdisciplinary Postgraduate Conference, (2019). Grahamstown, South Africa
7. Afrah Khairallah, Özlem Tastan Bishop, and Vuyani Moses. **The Development and Validation of CHARMM Force Field Parameters for the Zn²⁺ Centres of GTP Cyclohydrolase Enzyme.** The Centre for High-Performance Computing National Conference (CHPC), (2018). Cape Town, South Africa
8. Afrah Khairallah, Özlem Tastan Bishop, and Vuyani Moses. **The characterization of the Plasmodial GTP Cyclohydrolase I (GCH1) Enzyme as a Potential Malarial Drug Target Using Computational Approaches.** Rhodes University Interdisciplinary Postgraduate Conference. (2018). Grahamstown, South Africa

Awards

- The Centre for High-Performance Computing (CHPC) National Conference PhD. Student Poster Presentation Award (2019). Johannesburg, South Africa

Thesis overview

The research presented in this thesis is organized as five separate research chapters, three of which have been published in peer-reviewed journals. **Chapter one** includes a comprehensive review of the malaria disease as well as the study background. The chapter is then closed with the study aim rationale and primary objectives. **Chapter two** describes the sequence analysis, phylogenetic, and motif discovery of guanosine-5'-triphosphate (GTP) cyclohydrolase I (GCH1) and 6-pyruvoyl tetrahydropterin synthase (PTPS) to characterize unique sequence and structure-based features of the malaria parasite enzymes. **Chapter three** involves the generation of force field parameters for the CCH1 and PTPS active site metal ions. This is followed by evaluating the newly generated parameters via all-atom Molecular Dynamics (MD) simulations and post-MD trajectory analysis to deduce the accuracy of the derived parameters. **Chapter four** describes the Normal Mode Analysis (NMA) calculated on the Anisotropic Network Model (ANM) of the *Plasmodium falciparum* GCH1 and further elucidates the structural dynamics that govern the enzyme large-scale collective motions as well as disclosing regions responsible for modulating conformational changes and hot spot residues. This chapter also invites attention to the structural dynamics of the mammalian GCH1 enzyme and its feedback regulatory protein (GFRP). **Chapter five** includes the NMA of the *Plasmodium falciparum* PTPS synthase enzyme to study the PTPS large-scale collective motions and conformational states to pinpoint potential allosteric drug targeting sites. Lastly, the thesis is closed by the concluding remarks and future work.

Chapter 1

Study Background

Chapter overview

This introductory chapter presents the malaria disease and reviews the literature concerning current malaria control and treatments, further detailing the Plasmodium parasite life cycle, the drug resistance problem, and the parasite *de novo* folate biosynthesis pathway.

1.1 Introduction

1.1.1 Malaria

Malaria is an infectious disease caused by a unicellular protozoan parasite of the genus *Plasmodium* and phylum Apicomplexa (Sato, 2021). Malaria is transmitted through an infected female Anopheles mosquito bite during its blood meal (Frischknecht and Matuschewski, 2017). Five species of *Plasmodium* are known to cause malaria in humans, namely: *Plasmodium falciparum* (*P. falciparum*), *Plasmodium vivax* (*P. vivax*), *Plasmodium ovale* (*P. ovale*), *Plasmodium malariae* (*P. malariae*), and *Plasmodium knowlesi* (*P. knowlesi*) (Müller-Sienerth *et al.*, 2020). Other species of *Plasmodium* are known to cause malaria in rodents, such as *Plasmodium berghei* (*P. berghei*), *Plasmodium yoelii* (*P. yoelii*), and *Plasmodium chabaudi* (*P. chabaudi*). *P. falciparum* is responsible for most malaria cases and deaths globally; therefore, it is the most pathogenic among the five human infective species (World Health Organization, 2020). As per the World Health Organization (WHO) 2020 malaria report, 229 million malaria cases were reported worldwide, and the estimated number of malaria deaths stood at 409 000. 215 million cases were reported in the African region. Thus, the region accounted for about 94% of global cases and regarded as malaria-endemic.

Despite the significant progress achieved to eradicate malaria, the mosquito-borne disease remains a public health problem (World Health Organization, 2020), and there remains an enormous burden of malaria infection that cannot go unnoticed. Furthermore, malaria morbidity and mortality rates are higher among pregnant women and children (World Health Organization, 2020). Countries with limited resources are more affected by the disease, mainly on its public health and socio-economic development, due to many factors besides malaria itself, such as the lack of access to adequate and affordable treatments, socio-economic instability, ecological conditions, and the absence of well-established health systems (Onwujekwe *et al.*, 2013; Galactionova *et al.*, 2015; Kibret *et al.*, 2019; Sarma *et al.*, 2019). Additionally, the emerging parasites' resistance against available antimalarial drugs hindered malaria control and elimination efforts.

1.1.2 The malaria parasite life cycle

Plasmodium is an obligate parasite that grows intracellularly (Horta *et al.*, 2020). The parasite's life cycle is divided into two main phases: an asexual phase in the human host and a sexual phase in the mosquito (Siciliano and Alano, 2015). The parasite is first transmitted to the host blood system through an infected female Anopheles mosquito bite. During its blood meal, the mosquito inoculates the parasite (sporozoites) into the host bloodstream (Siciliano and Alano, 2015). The sporozoites are nucleated highly motile cells with distinctive apical complex and microtubule-based structures (Tardieux and Baum, 2016). The sporozoites are then taken up by the liver cells in where they replicate asexually. This stage of infection is called the pre-erythrocytic stage or liver stage; at this particular stage, the infection does not present any clinical manifestations. As a result, the host immune defence mechanisms are not activated, allowing the merozoites to survive (Soulard *et al.*, 2015). In the liver cells, the sporozoites then develop to form a multinucleate stage known as schizonts. A schizont contains several cell nuclei; the daughter nuclei are surrounded by cytoplasm and organize themselves into single spindle-shaped individuals, known as merozoites. After schizogony is completed, the swollen liver cells rupture and releases thousands of mobile daughter merozoites into the host bloodstream. The released merozoites rapidly infect the erythrocytes, starting the erythrocytic stage. The erythrocytic stage is the main stage of the infection and is associated with the pathology (Vaughan, Aly and Kappe, 2008). After invading the erythrocyte, the merozoites undergo a trophic period followed by an asexual replication in which they enlarge and develop to form immature or ring stage trophozoites which then progress to mature trophozoites.

During the trophic period, the parasite ingests the host cell cytoplasm and breaks down the haemoglobin into amino acids. A by-product of haemoglobin digestion is the malaria pigment (hemozoin). The ring stage then develops into replicative schizont forms, releasing multiple invasive daughter merozoites from decayed erythrocyte cells. The newly released daughter merozoites then initiate another round of schizogony to infect further erythrocyte. Multiple rounds of the erythrocytic stage produce larger numbers of parasites that invade the blood, resulting in thousands of infected red blood cells in the host circulatory systems, consequently causing clinical illness (Soulard *et al.*, 2015).

Next begins the sexual phase when a portion of the merozoites reproduce sexually and differentiate into male and female gametocytes, then get taken up by the mosquito during the blood meal. The key events of the parasite life cycle are illustrated in Figure 1.1.

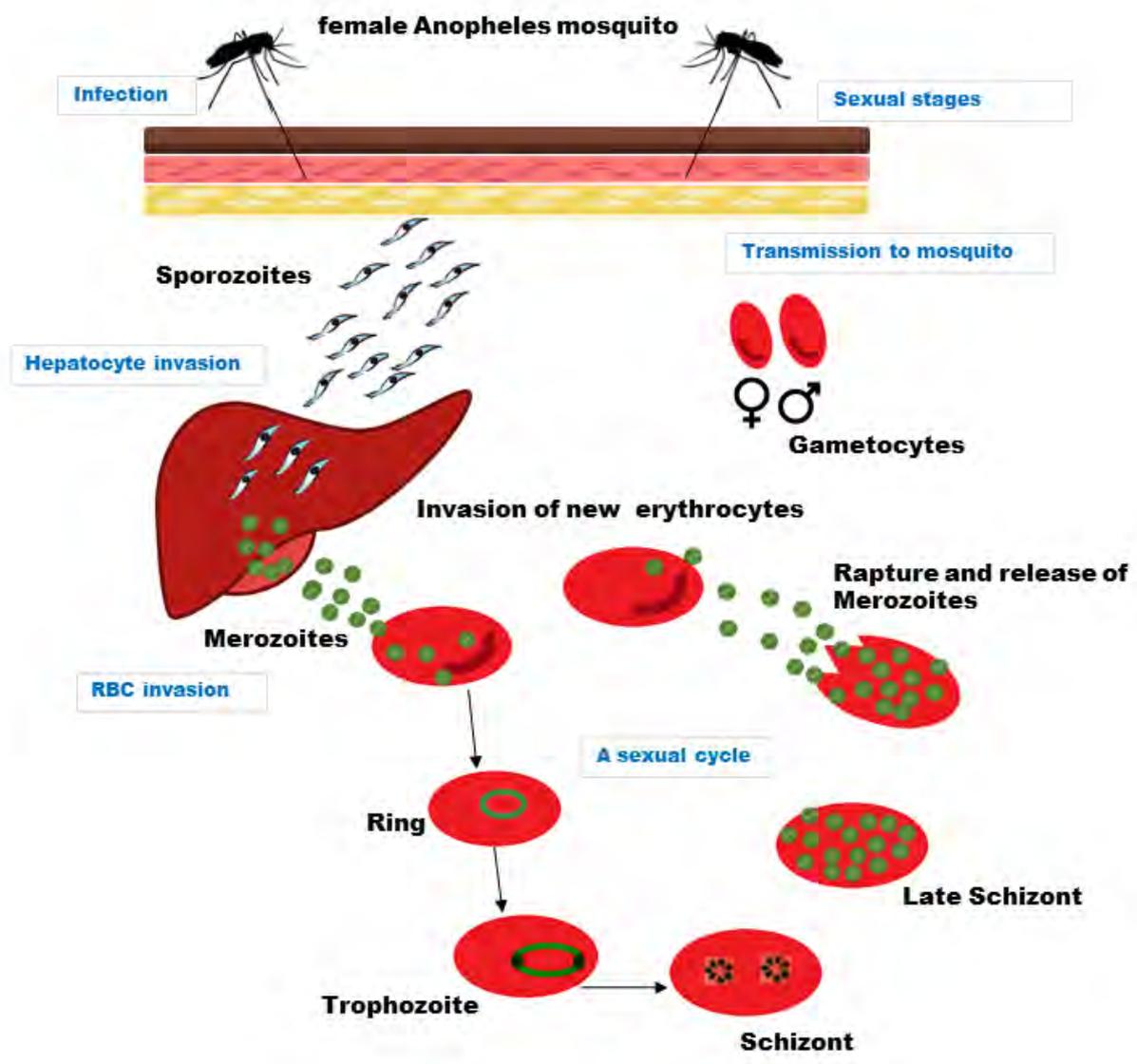


Figure 1.1. Key events of the malaria parasite life cycle in the human host. First, the sporozoites are injected into the host system through the mosquito bite during its blood meal. The sporozoites then travel to the liver and mature into schizonts. The schizonts rupture and release daughter merozoites into the host bloodstream. The merozoites infect the erythrocyte, then enlarge to ring stage trophozoites and progress to mature trophozoites. The mature trophozoites develop into replicative schizont forms and release multiple invasive daughter merozoites, starting a new infection cycle. A portion of the merozoites differentiates into sexual erythrocytic stages (male and female gametocytes) that can be transferred back to the female Anopheles mosquito during a blood meal to start the sexual

stage in the mosquito carrier. Inoculation of the sporozoites into a new human host repeats the malaria life cycle. (Adapted from (Molina-Franky *et al.*, 2020)).

1.1.3 The symptoms

The early signs and symptoms of malaria are flu-like symptoms of periodic alteration in body temperature, followed by shivers, fever, and sweating. If not treated on due time, the symptoms can progress and become more severe to include respiratory distress, severe anaemia, abstain from eating and drinking, repetitive vomiting, unconsciousness, failure of several body organs, coma, and death (Josling and Llinás, 2015; Basu and Sahi, 2017). Pregnant women infected with malaria are especially more vulnerable, in which they can exhibit more severe symptoms, including severe anaemia, higher rates of miscarriage, premature delivery, and maternal death (Bauserman *et al.*, 2019).

1.1.4 Classical antimalarial drugs and their mechanism of action

Antimalarial drugs can be divided into three main categories based on the stage of the parasite life cycle where they act, as tissue schizonticides that target the hypnozoites (the dormant stage of the parasite) in the liver, blood schizonticides that target the asexual erythrocytic stages of the parasite, and gametocides that target the sexual erythrocytic forms of the parasite in the host bloodstream to prevent their transmission back to mosquitoes (Sevene, González and Menéndez, 2010). Antimalarial drugs can also be classified based on the major metabolic pathways targeted by the antimalarial drugs as nucleic acid metabolism, heme toxification, oxidative stress, and fatty acid biosynthesis pathways as some of the novel targets (Shibeshi, Kifle and Atnafie, 2020). Also, antimalarials can be classified based on their chemical structures and mode of action (Kumar *et al.*, 2018). These include aryl aminoalcohol compounds, antifolate compounds, artemisinin compounds, and the antimalarial atovaquone as its own class with a unique mode of action. In the next section, the common antimalarial drugs are reviewed briefly. The structures of the below reviewed antimalarial drugs is well presented by Delves and colleagues (Delves *et al.*, 2018).

Aryl aminoalcohol compounds

Aryl aminoalcohol compounds are characterized by an aromatic and amino alcohol portion linked by a carbon chain of two or three atoms (Bhattacharjee and Karle, 1996). The aryl aminoalcohol compounds are fast-acting schizonticidal drugs, including quinine, quinidine, chloroquine, amodiaquine, mefloquine, halofantrine, lumefantrine, piperazine, and tafenoquine (Kumar *et al.*, 2018). Their mode of action is generally blocking the detoxification of the toxic byproducts of heme degradation (Müller and Hyde, 2010). The malaria parasite resistance against aryl aminoalcohol compounds first appeared in Southeast Asia in the late 1950s, then South America, and finally Africa (Cowell and Winzeler, 2019). The aryl aminoalcohol compounds (e.g., chloroquine) are still used as the first-line drug of choice in most sub-Saharan African countries, mainly in combination with other antimalarial drugs (Achan *et al.*, 2018).

Artemisinin compounds

Artemisinin, also called qinghaosu, is sesquiterpene lactone derived from the Chinese herb Qinghao (*Artemisia annua*) (Krungkrai and Krungkrai, 2016). The activity of artemisinin is linked to its unusual endoperoxide trioxane moiety (Aderibigbe, 2017). The drug is active against the parasite early ring, later ring forms, blood schizonts, trophozoites, and early gametocyte stages (Bridgford *et al.*, 2018). Artemisinin compounds are first activated by the intra-parasitic heme-iron, which catalyses the cleavage of its endoperoxide, resulting in the generation of free radical intermediates that can kill the parasite by causing protein damage to one or more of its essential proteins as well as the inhibition of newly synthesized proteins folding (Meshnick, 2002). The main therapeutic limitation of artemisinin is its low solubility in both oil and water. In order to address this problem, semi-synthetic drugs and synthetic artemisinin derivatives with higher solubility were derived; some of the common artemisinin derivatives include dihydroartemisinin, artemether, arteether, and artesunate (Aderibigbe, 2017).

Artemisinin and its derivatives are often used in combinations with other conventional antimalarial drugs such as lumefantrine, piperazine, amodiaquine, and sulfadoxine-pyrimethamine. The use of artemisinin combination therapy (ACT) was adopted to mitigate the multidrug resistance to

artemisinin monotherapy. The ACT was shown to be safe, well-tolerated, highly effective, and can delay the onset of drug resistance (Meshnick, 2002), whereas artemisinin monotherapy was considered an inappropriate treatment (Menard *et al.*, 2005). To date, the efficacy of Artemisinin combination therapy has remained high outside Southeast Asia (Woodrow and White, 2017). However, a recent study reported some evidence that resistance to artemisinin combination therapy may already be developing in parts of East Africa (Dokomajilar *et al.*, 2006; Uwimana *et al.*, 2020).

Antifolate compounds

Antifolate compounds, also known as folate antagonists, disrupt the balance of metabolites in the parasite folate synthesis pathway, thus the depletion of folate cofactors, which affects the parasite nucleic acid biosynthesis and cellular metabolism. The antifolates act against different stages of the parasite's life cycle due to their ability to inhibit nucleic acid synthesis. Antifolate compounds include pyrimethamine, trimethoprim, proguanil, and chlorproguanil (Nzila, 2006), to mention a few. The antifolate drug therapy was shown to be effective, but the parasite still developed resistance against available antifolates (Heinberg and Kirkman, 2015). To further stall the development of resistance, combination therapy was introduced for malaria treatment, in which pyrimethamine and proguanil were first used in synergy (Plowe *et al.*, 1995), followed by sulfonamides drugs, such as sulfadoxine and sulfalene and the sulfone, dapson (collectively known as the sulfa drugs) in combination with pyrimethamine (White, 1999). Resistance to combination therapy such as sulfadoxine and pyrimethamine was soon encountered in Southeast Asia and later spread worldwide. The combination of sulfadoxine and pyrimethamine is still widely used in Africa for therapy and malaria prevention, especially in children and pregnant women (Mlugu *et al.*, 2020). Another common type of combination therapy is proguanil-atovaquone (Malarone), major chemoprophylaxis for nonimmune travellers in endemic areas (Schlagenhauf *et al.*, 2018).

Atovaquone

Atovaquone is classified in its own class due to its unique mode of action (Baggish and Hill, 2002). The drug targets the parasite mitochondrial cytochrome bc1 complex, a component of the electron transport chain, by acting as an analogue of ubiquinone, resulting in a collapse to the mitochondrial

membrane potential, thus arresting the parasite respiration and essential pyrimidine biosynthesis (Hyde, 2007). The atovaquone was first used in monotherapy, and parasite resistance rapidly developed, reducing the drug's efficacy (Berry *et al.*, 2006). To overcome the resistance problem, Atovaquone was used in combination with other antimalarials, such as proguanil marketed with the trademark Malarone[®]. The Atovaquone-proguanil combination therapy was effective until cases of resistance had been reported (Kuhn, Gill and Kain, 2005; Berry *et al.*, 2006; Hyde, 2007; Staines *et al.*, 2018).

1.1.5 The current state of the fight against malaria

Most of the classical antimalarial drugs have been used for decades, but in many cases, their use was restricted by the emergence of parasite resistance (Hyde, 2007; Blasco, Leroy and Fidock, 2017). Furthermore, the mechanisms of resistance are poorly understood for most of these drugs. It is also important to mention the collateral impact of the new coronavirus disease (COVID-19) on malaria, especially in Africa, in which the fight against malaria is likely to be hampered (Nghochuzie *et al.*, 2020).

An essential part of fighting malaria is developing vaccines and antimalarials of a novel mechanism of action to target the parasite at multiple points in its life cycle and further include innovative gametocyte-targeted transmission-blocking drugs (Delves *et al.*, 2019; Reader *et al.*, 2021). In recent years, the pace of antimalarial research progress was relatively high, in which several promising antimalarial candidates have made it either to pre-clinical or clinical trials (Tse, Korsik and Todd, 2019). Most of these compounds are of a novel mechanism of action, including compounds of the Medicines for Malaria Venture (MMV), University of Cape Town, Anacor Pharmaceuticals, BIOTEC Thailand P218, St Jude Children's Research Hospital, and Rutgers University ATP4 inhibitor SJ733 (Delves *et al.*, 2019; Chughlay *et al.*, 2020; Reader *et al.*, 2021), to mention a few popular ones.

1.1.6 Malaria vaccines

Besides the ongoing antimalarial drug discovery efforts, effective vaccine development is essential to control the disease. Several vaccines have been developed and are divided into three broad categories

based on the targeted stages of the parasite life cycle a pre-erythrocytic vaccine (PEV), blood-stage vaccines (BSV), and transmission-blocking vaccines (TBV) (Duffy and Patrick Gorres, 2020). PEV vaccines remain the most promising among all, in which the PEV vaccine RTS, S (Mosquirix™) has advanced to implementation in humans. The ongoing implementation of the programs will be completed in the year 2022, and the outcomes will be reviewed for broader use in Africa (Duffy and Patrick Gorres, 2020). The TBV has progressed to phase 2 clinical trials over the past decade, whereas the BSV vaccines targeting the merozoite invasion proteins was not successful (Duffy and Patrick Gorres, 2020); this has shifted the attention to find alternative novel BSV antigens and other BSV targets such as infected red cell surface proteins, schizont egress antigens, or intact infected erythrocytes that have been attenuated (Draper *et al.*, 2018; Duffy and Patrick Gorres, 2020; Molina-Franky *et al.*, 2020).

1.1.7 The *de novo* folate synthesis pathway

Folate is the natural form of vitamin B9; it is water-soluble and naturally found in many foods such as vegetables of green leaves (Rosenberg, 2012). The name originated from the Latin word folium, meaning leaf. Folate derivatives are enzyme cofactors responsible for transferring one-carbon units in various cellular metabolism to synthesise several amino acids and nucleic acid metabolism (Nzila, Steve A. Ward, *et al.*, 2005; Ducker and Rabinowitz, 2017; Tjong and Mohiuddin, 2019). The folate molecule is found naturally as conjugated pterins, composed of a pteridine ring attached to a *p*-aminobenzoate (pABA) moiety with one or several glutamates linked to it (Figure 1.2). Depending on the oxidation level, the transferable one-carbon unit can be attached to either the N5 of the pteridine ring or the N10 of the pABA structure (de Crécy-Lagard *et al.*, 2007).

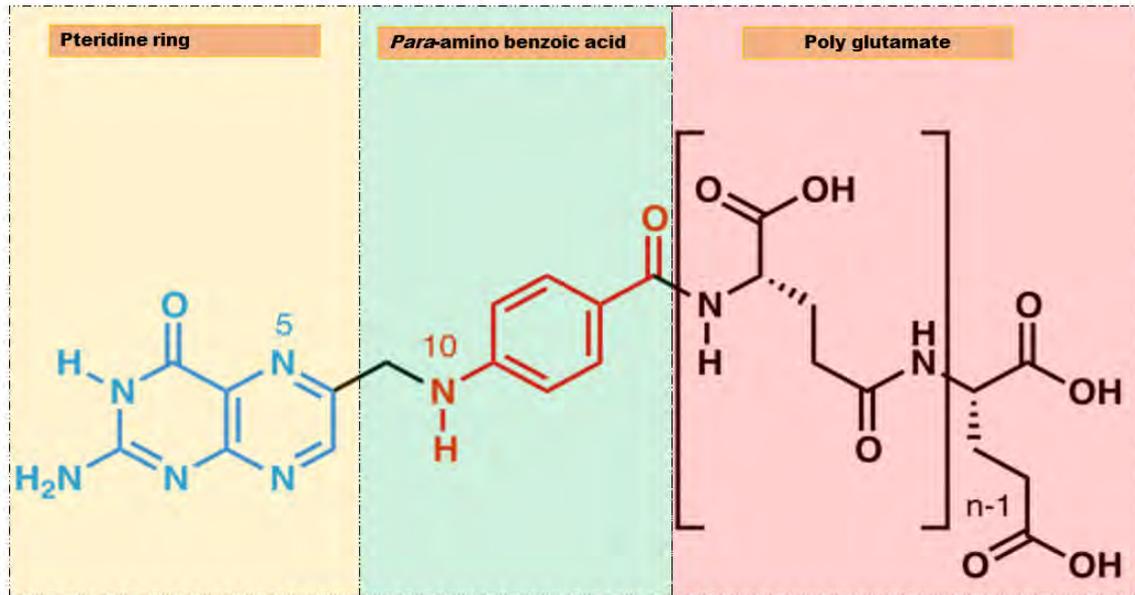


Figure 1.2. Structure of the folate molecule. (Adapted from (Rosenberg, 2012)).

In addition to their ability to salvage folate externally, most prokaryotes and unicellular eukaryotes can make their own folate from the simple precursors guanosine-5'-triphosphate (GTP), pABA, and glutamate, whereas higher eukaryotes, except for plants, do not have the biosynthetic capability of folate synthesis and depend primarily on external dietary uptake of folate (Hyde, 2005; de Crécy-Lagard *et al.*, 2007).

The malaria parasite *de novo* folate biosynthesis pathway consists of six enzymes that convert GTP to the final product tetrahydrofolate. The first reaction step of the *de novo* folate biosynthesis pathway is catalysed by GTP cyclohydrolase (GCH1), which converts the GTP moiety to the 7,8-dihydroneopterin (DHNP). The second step of the pathway is catalysed by 6-pyruvyl tetrahydropterin synthase (PTPS), which converts the DHNP product to 6-pyruvyl tetrahydropterin (PTP) via an internal redox transfer and final elimination of the DHNP phosphate tail (Bürgisser *et al.*, 1995), in addition to producing PTP, the enzyme also produces 6-hydroxymethyl dihydropterin (HDMP) as a by-product (Dittrich *et al.*, 2008). HDPM is then processed by the enzyme 6-hydroxymethyl dihydropterin pyrophosphokinase (HPPK), affording 6-hydroxymethyl dihydropterin pyrophosphate (HDMP-P2), an active pyrophosphorylated intermediate. The resulting product HDMP-P2 is then combined with pABA to generate dihydropteroate (DHP), mediated by dihydropteroate synthase (DHPS). The pABA is principally obtained either by *de novo* synthesis via the shikimate pathway or

salvage from the host plasma (Mather and Ke, 2019). Dihydrofolate synthase (DHFS) then adds one glutamate moiety to the DHP product forming dihydrofolate. The last reaction step is catalysed by dihydrofolate reductase (DHFR), which converts dihydrofolate to the reduced form tetrahydrofolate (Colloc'h, Poupon and Mornon, 2000; de Crécy-Lagard *et al.*, 2007; Kümpornsin, Kotanan, *et al.*, 2014). The overall reactions of the *de novo* folate biosynthesis pathway are shown in Figure 1.3.

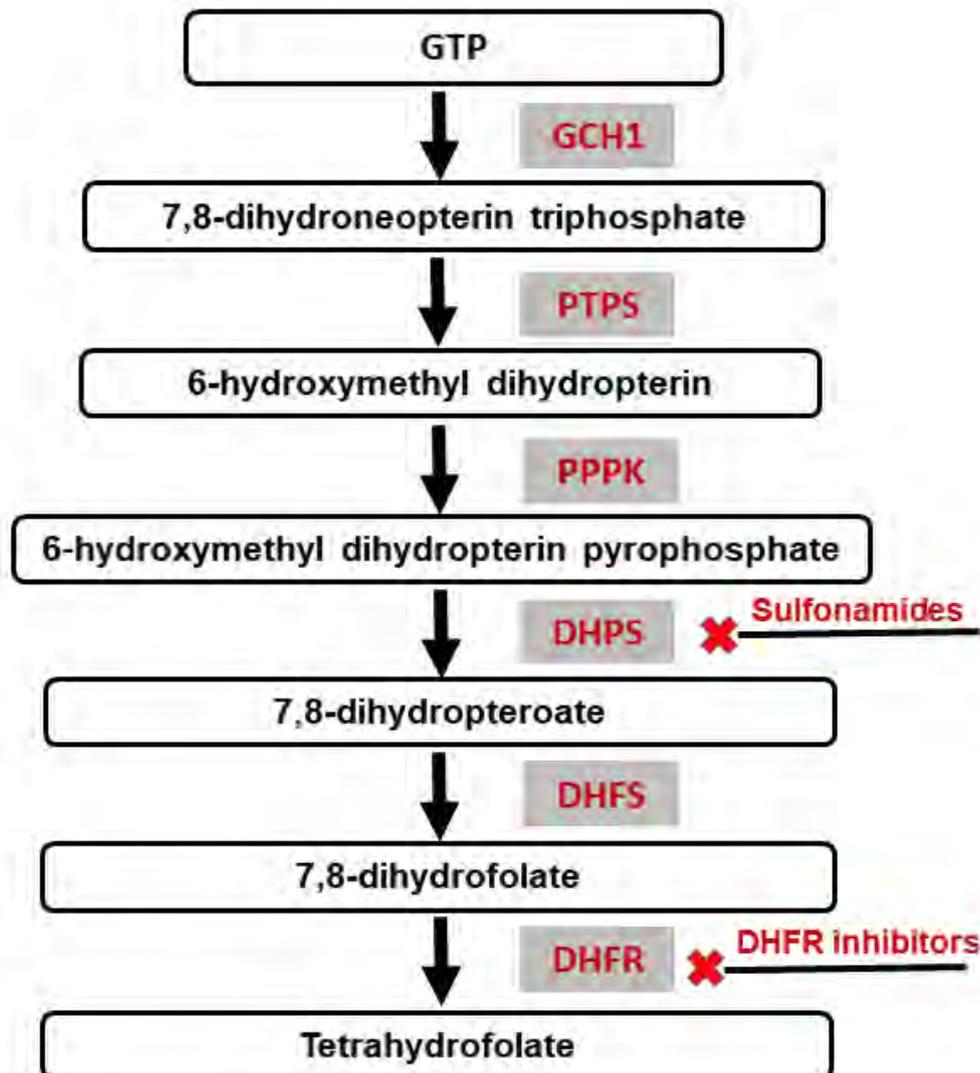


Figure 1.3. Illustration of the malaria parasite *de novo* folate biosynthesis pathway. The pathway is constituted by six enzymes: GTP cyclohydrolase (GCH1); 6-pyruvoly tetrahydropterin synthase (PTPS); 6-hydroxymethyl dihydropterin pyrophosphokinase (PPPK); dihydropteroate synthase (DHPS); dihydrofolate synthase (DHFS); dihydrofolate reductase (DHFR). Para-aminobenzoic acid (pABA) enters as a substrate for DHPS. Enzymes inhibited by the antifolate drugs are indicated by a red cross, whereas the Inhibitors of folate synthesis are shown in red at the right of the pathway. The

salvaged folate enters the pathway upstream of DHFR. (Adapted from (Heinberg and Kirkman, 2015)).

1.1.8 Antifolate drug resistance

When the resistance against the classical antimalarial drugs such as chloroquine, primaquine, quinine, mefloquine, and artemisinin derivatives emerged, the antifolate class became the most common alternative. Antifolates were affordable, safe, well-tolerated, and continued to be useful and effective for preventing and treating malaria (Müller and Hyde, 2010). The antifolates reduce the availability of folate derivatives by targeting two critical enzymes of the malaria parasite *de novo* folate biosynthesis pathway, which are the DHFR and DHPS (Nduati *et al.*, 2008). The DHFR is inhibited by methotrexate, trimethoprim, cycloguanil, and pyrimethamine (Schweitzer, Dicker and Bertino, 1990; Nzila, Steve A. Ward, *et al.*, 2005), and the DHPS enzyme, which is exclusive to prokaryotes, is inhibited by sulfonamide drugs (Sköld, 2000).

Despite the success of antifolates therapy in malaria treatment, the parasite still developed resistance against the available antifolates resulting in the reduction of their efficacy, hence causing a delay in clearing the parasite from the infected host blood system (Heinberg and Kirkman, 2015). The antifolates resistance first appeared in South America and the Southeast in the 1970s and was later noted in Africa in the 1980s (Heinberg and Kirkman, 2015). Some of the factors that contributed to the emergence and progress of the resistance are the intensive use of antifolates as a preventative therapy, antifolates monotherapy, and inadequate or incomplete treatments (White, 2004; Hyde, 2007). The drug resistance persisted even with increased drug dosage, and the parasite was also able to escape the effect of combination therapy in which a combination of antifolate drugs was used to collectively inhibit both DHFR and DHPS (Nzila, 2006; Nyunt and Plowe, 2007; Mohapatra *et al.*, 2014).

The reported antifolates resistance occurred due to the acquisition of point mutations in genes coding the parasite's DHFR and DHPS enzymes (Müller and Hyde, 2013; Heinberg and Kirkman, 2015). The point mutations that confer resistance arise in a stepwise manner, and the accumulation of point mutations leads to greater degrees of parasite resistance and higher tolerance against the antifolates

(Heinberg and Kirkman, 2015). There are five well-characterized point mutations, three in the DHFR enzyme (N51I, C59R, S108N) and two in the DHPS enzyme (A437G, K540E) (Heinberg and Kirkman, 2015). Additional mutations such as I164L and A581 of the DHFR and DHPS, respectively, are becoming more common than before and could evolve to comprise further the antifolates activity (Conrad and Rosenthal, 2019).

Structural studies involving the DHPS showed that the discrimination between the binding of sulfa drugs compared to the natural substrate pABA is a result of a steric hindrance causing the displacement of the drug in the active site and prevents the tight coupling (Griffith *et al.*, 2018). In contrast, the presence of mutations in the DHFR enhances its structural flexibility and weakens the binding affinity of pyrimethamine to the DHFR (Amusengeri, Tata and Bishop, 2020). In addition to the reported mutations, the efficacy of antifolates is further compromised by the ability of the parasite to salvage folate and increase its flux through other enzymes of the *de novo* folate biosynthesis pathways (Nair *et al.*, 2008; Heinberg and Kirkman, 2015).

1.1.9 Alternative metabolic targets: the search for new antifolates

Antifolate drugs have been long known to be effective in malaria treatment, but the emerged parasite resistance limited their efficacy and, consequently, their use. Given the previous success in targeting the parasite *de novo* folate biosynthesis pathway and the vast store of knowledge about this pathway, other enzymes within the same pathway stand as good alternatives and can be exploited as potential drug targets for the treatment of malaria. The malaria parasite DHFR structure is different from its human counterpart in which the human DHFR is inhibited much less by the antimalarial DHFR inhibitors (Yuthavong *et al.*, 2005). The DHPS, on the other hand, is absent in humans, which forms the basis for the DHPS inhibitors selectivity. As a result of such differences, the current antifolates have been effectively used in the treatments of malaria and other infectious diseases. However, the parasite antifolate emerged resistance reduced the current antifolates efficacy and hindered their use.

1.1.10 Enzymes of the tunnelling fold superfamily as alternative antifolate drug targets

The Tunnelling fold (T-fold) protein superfamily includes a group of homo-multimeric proteins crossed by a distinct central tunnel. Protein members of the T-fold superfamily similarly bind to flat substrates of purine or pterin families at their active sites. The active sites of the T-fold enzymes are formed at the interface of every three adjacent monomeric units. A striking feature characterizing the known T-fold proteins is that they share a high level of structural conservation and a low level of sequence similarity (often below 10% sequence identity). Furthermore, their functions and catalytic mechanisms are unrelated (Colloch, Poupon and Mornon, 2000). The monomeric topology of the T-fold superfamily is characterized by antiparallel β -sheets of four sequential strands flanked by antiparallel helices on their concave side. The monomeric units then assemble through tight hydrogen bonds of their β -sheets, resulting in a single compact domain of α and β barrel-like structure. The complete barrel-like functional assembly is formed via two barrels face to face association, leaving the tunnel running through the multimeric units (Colloch, Poupon, and Mornon 2000; Thöny, Auerbach, and Blau 2000; Tanaka *et al.*, 2005). Both malaria parasite *de novo* folate pathways enzymes GCH1, and PTPS enzymes belong to this exclusive superfamily and will be discussed in detail in the next section.

Guanosine-5'-triphosphate cyclohydrolase

As discussed earlier, the GCH1 is the first and rate-limiting enzyme of the malaria parasite *de novo* folate biosynthesis pathway (Kompis, Islam and Then, 2005). Products of the *de novo* biosynthesis pathway are essential cofactors that act as a source of the one-carbon unit for the biosynthesis of purine, pyrimidine, and methionine, all of which are important for DNA replication of the parasite cells (Nzila, Steve A Ward, *et al.*, 2005). However, GCH1 is the initiating and rate-limiting enzyme of the tetrahydrobiopterin (BH4) biosynthesis pathway in mammals. BH4 is a key cofactor for nitric oxide-producing enzymes, melanin, and neurotransmitters (Paranagama *et al.*, 2017). In that way, BH4 is responsible for maintaining cardiovascular, inflammatory, and neurovascular homeostasis (McNeill and Channon, 2012; Ichinose *et al.*, 2013; Bendall *et al.*, 2014; Rivera *et al.*, 2017).

Guanosine-5'-triphosphate cyclohydrolase 3D structure

GCH1 is a homo-decameric enzyme with 5-fold symmetry (Nar, Huber, Meining, *et al.*, 1995). The GCH1 structure is therefore composed of 10 identical monomeric/chain units. Each monomer folds into an α and β structure with predominant anti-parallel N-terminal helix pairs. Every five-equivalent monomeric/chain unit then associate together symmetrically along their β -sheets, forming one pentameric unit. Nar and colleagues described the GCH1 pentamer structure as a crab shape with five legs, in which the legs are the description of the protein N-terminal helices. The GCH1 functional assembly unit is formed by a face to face association of two pentameric units, leaving the tunnel running through the homo-decameric structure (Nar, Huber, Meining, *et al.*, 1995). The C-terminal helices of each monomer make the central cavity; they are also called the central five-helix bundle. The overall structure of the GCH1 is presented in Figure 1.4-A. The *E. coli* GCH1 enzyme was reported to have a toroidal shape with an approximate height of 65 Å and a diameter of 100 Å (Nar, Huber, Meining, *et al.*, 1995). The GCH1 tunnel dimensions are 30 Å x 30 Å x 15 Å, and it is only accessible through the centre opening and not the protein equator (Nar, Huber, Meining, *et al.*, 1995).

GCH1 has 10 identical zinc-containing active sites, each buried in a deep pocket of 10 Å (1 nm). The active site is formed at the interface of every three adjacent monomers (Nar, Huber, Meining, *et al.*, 1995; Colloc'h, Poupon and Mornon, 2000; Khairallah, Tastan Bishop and Moses, 2021). The GCH1 Zn²⁺ ion in the active site is coordinated to a histidine residue through its ND1 atom, two cysteine residues, and a water molecule that completes the tetrahedral coordination environment of the ion (G Auerbach *et al.*, 2000; Tanaka *et al.*, 2005; Khairallah, Tastan Bishop and Moses, 2021).

The GCH1 core is highly conserved among species at both the sequence and structure levels (Nar, Huber, Meining, *et al.*, 1995; Colloc'h, Poupon and Mornon, 2000; Tazawa *et al.*, 2000; Gräwert, Fischer and Bacher, 2013; Kämpornsin, Kotanan, *et al.*, 2014). In contrast, the GCH1 N-terminal domain is highly diverse and of variable lengths (Funderburk *et al.*, 2006; Swick and Kapatos, 2006; Higgins and Gross, 2011).

In a study by Funderburk and colleagues, the N-terminal regions were anticipated to exert a strong regulation effect on the protein catalytic domain via an unknown regulatory interaction/coupling between the N-terminal helices and the active sites (Funderburk *et al.*, 2006). Furthermore, Kämpornsin and colleagues performed a bacterial complementation assay to identify the components of GCH1 necessary for its activity (Kämpornsin, Kotanan, *et al.*, 2014). The study showed that deletion of the N-terminal regions did not affect the protein genetic complementation/activity. Therefore, it was suggested that these N-terminal helices are not directly required for the catalytic activity of GCH1. The GCH1 N-terminal helices were also described as docking sites for protein-protein interaction (Krishnakumar *et al.*, 2000; Bowling *et al.*, 2008). Lastly, Swick and Kapatoss reported that deletion of the human GCH1 N-terminal residues 1 to 42 decreased the strength of GCH1 interaction with its feedback regulatory protein (GTP cyclohydrolase I feedback regulatory protein; GFRP) and the truncation of C-terminal (central five-helix bundle) amino acids 237 to 250 eliminated the GFRP binding to GCH1 (Swick and Kapatoss, 2006).

GCH1 feedback regulatory protein

Altered levels of BH₄ due to abnormalities in the regulation mechanisms of the mammalian GCH1 activity results in several clinical syndromes and neuronal disorders (Latremoliere and Costigan, 2011; H *et al.*, 2020). To ensure that this balance is maintained, the activity of the mammalian GCH1 is controlled by the regulatory partner GFRP. The GFRP binds to each side of the GCH1 structure to mediate allosteric feedback inhibition by BH₄ and feed-forward activation by phenylalanine (Harada, Kagamiyama and Hatakeyama, 1993; Maita *et al.*, 2004; Hussein *et al.*, 2015). In prokaryotic and lower eukaryotic organisms, the activity of GCH1 is not regulated by GFRP, in which the latter has not been found and reasoned to the GCH1 distinct functional role in the *de novo* folate biosynthesis (Maita *et al.*, 2004; Kämpornsin, Kotanan, *et al.*, 2014). Additionally, the N-terminal regions of prokaryote and lower eukaryote organisms were reported as heterogeneous and more extended relative to the mammalian GCH1; thus, the N-terminal regions were proposed to play a role in regulating the GCH1 within these organisms (Kämpornsin, Kotanan, *et al.*, 2014).

The GFRP structure is arranged in two pentameric units with 5-fold symmetry (Figure 1.4-A). The pentameric unit consists of a six-stranded anti-parallel β -sheet and two α -helices (Maita *et al.*, 2004). The overall structure of the GCH1-GFRP complex is illustrated in Figure 4-B and Figure 4-C. The GCH1-GFRP complex exists in either a stimulatory or inhibitory form. The inhibitory complex is identified to contain 10 allosteric effector molecules of BH4 located at the interface of every two neighbouring GCH1 chains; the BH4 binding site consists of the residues D118-M120, L148 -N150, and the C-terminal residues T220 to E234 (Maita *et al.*, 2004). The stimulatory complex contains 10 allosteric effector molecules of phenylalanine (Phe) located at the interfaces formed by two neighbouring GFRP subunits; the Phe binding sites are made of the following residues Q9-E13 and V73-V78 (Harada, Kagamiyama and Hatakeyama, 1993; Maita *et al.*, 2002).

GCH1 as a potential drug target

GCH1 is crucial for BH4 biosynthesis; mutations in the gene encoding the GCH1 enzyme were reported to cause rare humans genetic disorders such as malignant hyperphenylalaninemia (Dayasiri *et al.*, 2019), dopamine responsive dystonia, and Parkinson's disease (Mencacci *et al.*, 2014), to mention a few. As a result, the GCH1 stands as an attractive drug target for the development of novel drugs directed against humans rare hereditary disorders, pulmonary hypertension, atherosclerosis, diabetes mellitus, Alzheimer's disease, depression, and neuropathic pain (Alp *et al.*, 2003; Lötsch *et al.*, 2010; Latremoliere and Costigan, 2011; Pan *et al.*, 2011; Bendall *et al.*, 2014; Chuaiphichai *et al.*, 2014; Simonet *et al.*, 2021).

In prokaryotes and lower eukaryotes (e.g., parasitic protozoa), GCH1 plays a distinct role as the initiating and rate-limiting enzyme of the *de novo* folate biosynthesis pathway (Kompis, Islam and Then, 2005). Given that rapidly dividing cells rely heavily on the availability of folate derivatives, the *de novo* folate biosynthesis pathway is primarily pursued as a drug target in the treatment of several infectious diseases, including malaria (Nzila, 2006; Anderson and Wright, 2014; Bourne, 2014). Besides being the rate-limiting enzyme of this pathway, GCH1 was reported to influence the course of drug resistance evolution and was found to be overexpressed in resistant parasites to compensate for the loss of folate (Heinberg *et al.*, 2013; Kümpornsin, Modchang, *et al.*, 2014). Furthermore, the

transcription of GCH1 was reported to peak during the early trophozoite stage of the malaria parasite, indicating the crucial role of the enzyme in folate synthesis and consequently cell proliferation (Krungkrai, Yuthavong and Webster, 1985; Nzila, Steve A Ward, *et al.*, 2005). Hence, GCH1 can be pursued as a potential drug target for developing novel antimalarial and antibacterial drugs.

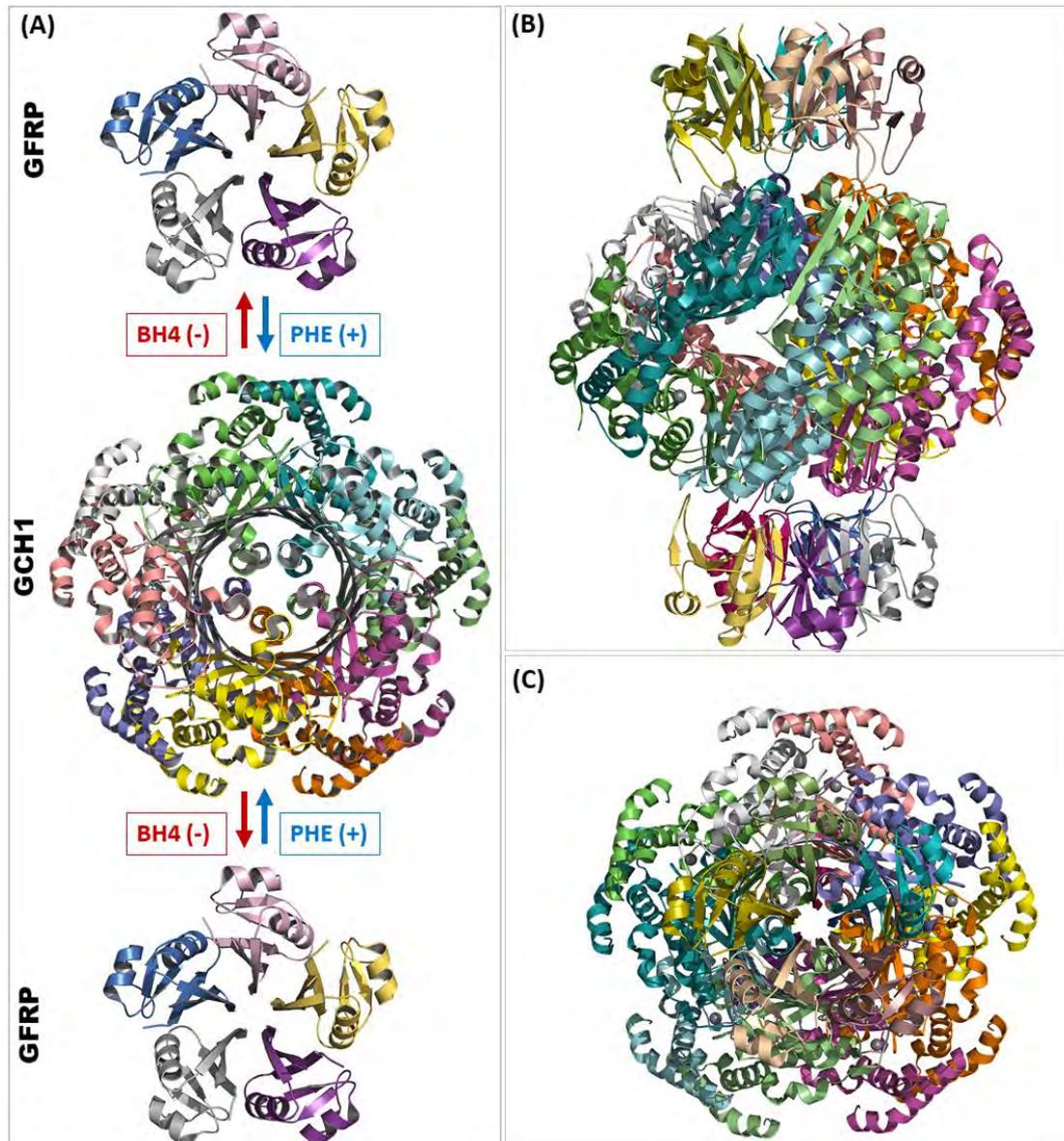


Figure 1.4. Coloured by chain the 3D structure of the GFRP two pentameric units that bind to both sides of the GCH1 homo decameric structure feedback inhibition (-) of GCH1 activity by BH4 (indicated by the red arrow) and feed-forward stimulation (+) by phenylalanine (indicated by the blue arrow). GCH1-GFRP complex of the *R. norvegicus* (PDB ID: 1WPL) (B) Side view of the GCH1-GFRP complex structure and (C) Top view of the GCH1-GFRP complex structure. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2021)).

6-pyruvoyl tetrahydropterin synthase as a potential drug target

The subsequent enzyme of the malaria parasite *de novo* folate biosynthesis pathway PTPS is a homohexameric lyase that has 3-fold symmetry with an overall dimension of $60 \times 60 \times 60 \text{ \AA}$ (Thöny, Auerbach, and Blau 2000; Burgisser *et al.*, 1994). The six monomers assemble via tight hydrogen bonds to form two trimer structures (Nar *et al.*, 1994). The enzyme functional assembly unit is formed by a face-to-face association of the two trimeric units, leaving the tunnel running through the homohexameric structure (Figure 1.5). The PTPS hexameric structure encloses a central large solvent-filled cavity of the dimensions $20 \times 20 \times 15 \text{ \AA}$. The distinct central cavity accumulates a cluster of basic and aromatic residues (Nar, 2011). PTPS contains six identical zinc-containing active sites, and each site is buried in a deep cavity of 12 \AA formed between every three adjacent monomers. The active site Zn^{2+} ion is coordinated to three histidine residues through their NE2 atoms and one water molecule (Khairallah, Tastan Bishop and Moses, 2020). The active sites of the PTPS are accessible to the substrate through the central opening along the trimer axis (Bürgisser *et al.*, 1995).

The malaria parasite PTPS enzyme has never been targeted before and presents as an attractive alternative drug target due to its distinct role in the malaria parasite *de novo* folate biosynthesis pathway. Previous studies also reported a distinct difference of the *P. falciparum* active site relative to the humans, in which a conserved active site residue (cysteine) was substituted by glutamate in the *P. falciparum* PTPS (Dittrich *et al.*, 2008). Furthermore, the unique structural features of the malaria parasite PTPS can provide an opportunity to discover new allosteric sites.

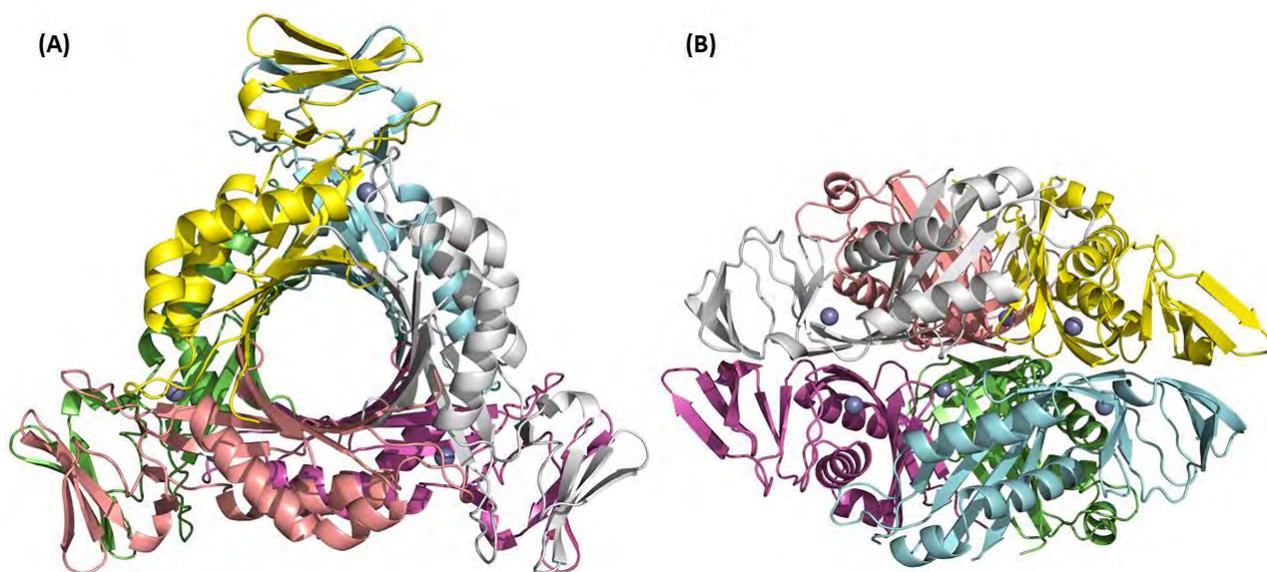


Figure 1.5. Coloured by chain (A) top and (B) side views of *P. falciparum* PTPS 3D structure (PDB: 1Y13). The homo-hexameric functional assembly unit is formed by the face to face association of two trimeric units. Each trimer comprises three monomeric units that assemble via tight hydrogen bonds of the protein N- and C-terminal β -strands.

1.2 Knowledge gap

Despite the significant progress achieved to eradicate malaria, the mosquito-borne disease remains a major public health problem (World Health Organization, 2020). The developed and on-gonging drug resistance against the available antimalarial antifolates drives an urgent need for continued research to discover new metabolic drug targets and drugs of a novel mechanism of action to stay one step ahead. The class of antifolates was effective in treating and preventing malaria for over decades. However, the parasite resistance against the available antifolate drugs emerged rapidly and persisted even with increased drug dosage or in combination therapy, rendering these drugs ineffective in many cases (Nzila, 2006; Grimberg and Mehlotra, 2011; Hanboonkunupakarn and White, 2015; Antony and Parija, 2016; Kumar, 2017). The emerged resistance allowed the parasite to escape the effects of antifolates and reducing their efficacy in clearing the parasite, thus resulting in prolonged illness and a higher risk of death. Looking ahead to the anticipated constraints on antifolates resistance, new antifolate drug targets and drugs are needed.

1.3 Study rationale

Folate derivatives are essential for the malaria parasite cell division and growth, and this has led to the pathway being recognized as an attractive metabolic target for treating several infectious diseases, including malaria. However, due to the loss of effectiveness of current antifolates to the parasite resistance, other enzymes within the pathway should be explored as alternative drug targets to overcome the resistance problem. There is a good rationale for developing a new antifolate drug, given the previous success in targeting this pathway, it being well characterized, and the mechanism of drug resistance in the previously targeted enzymes is fairly well understood. Additionally, the pathway initiating enzyme GCH1 and the subsequent enzyme PTPS belong to the T-fold superfamily, a superfamily characterized by a unique structural topology that can be exploited to reveal potential allosteric drug targeting sites. The two enzymes have never been targeted before and stand as potential drug targets. Furthermore, the *P. falciparum* GCH1 has been reported to influence the course of drug resistance evolution in the parasite (Kümpornsin, Modchang, *et al.*, 2014), and it appears to be vital for erythrocytic-stage parasites as suggested by the failure to make a GCH1 knockout line in *P. falciparum* (Müller and Hyde, 2013).

1.4 Research aim and objectives

Aim

The antimalarial antifolates drug resistance motivates our efforts to search for new drug targets and drugs of a novel mechanism of action. The primary aim of this study is to contribute to the existing and growing knowledge in antimalarial drug discovery by characterizing the malaria parasite *de novo* folate biosynthesis pathway enzymes GCH1 and PTPS as alternative metabolic drug targets for malaria treatment. More specifically, to illustrate their sequences and structural dynamics to disclose new allosteric drug targeting sites within their structures.

Main objectives

- (1) To perform sequence, phylogenetic, and motif analyses to identify the parasite enzymes' unique sequences as well as structure-based features that ensure future selective inhibition.

- (2) To derive force field parameters that adequately describe the two enzymes' active site metal centres during all-atom molecular dynamic simulations.
- (3) To study the structural dynamics of the two enzymes using normal mode analysis of elastic network model:
- To characterize large-scale motions of functional relevance.
 - To identify key regions that govern the large-scale motion and disclose potential allosteric drug targeting sites.

The objectives mentioned above can help decipher the two enzymes' sequence, structure, global dynamics, and hence function. It can further highlight important and unique parasite regions that regulate/drive critical global motions, this in turn can aid in the rational design of allosteric drugs to escape drug resistance.

Chapter 2

Sequence and Structural Characterization of the Tunnelling-Fold Enzymes GTP Cyclohydrolase I and 6-Pyruvoyl Tetrahydropterin Synthase

Chapter overview

The work presented in this chapter describes the sequence, phylogenetic, and motif analyses of the malaria parasite *de novo* folate biosynthesis pathway enzymes, guanosine-5'-triphosphate cyclohydrolase I and 6-pyruvoyl tetrahydropterin synthase, to characterize and pinpoint unique sequence and structure-based features that can ensure a selective inhibition of the two enzymes.

2.1 Introduction

Enzymes carry out various functional roles in the living organism. They control and implement various biological processes such as catalysis, cell signalling, immunological responses, structural support and transportation, endocrine function, and storage (Buxbaum, 2007; Mäntsälä and Niemi, 2009). Studying the underlying amino acid sequences of enzymes is essential to understand their function. The first step of characterizing an enzyme's structure and functional domains involves comparing its sequence against sequence databases to find similar sequences (Hatherley *et al.*, 2016). High sequence similarity indicates sequence homology, meaning that sequences are evolutionarily related and share a common ancestral sequence, consequently similar functional domains (Sander and Schneider, 1991; Özlem Tastan Bishop, de Beer and Joubert, 2008). There are three types of homologs, namely: orthologs, paralogs, and xenologs. Orthologs result from evolutionary speciation events, paralogs result from a duplication event, and xenologs result from horizontal gene transfer between different species (Koonin, 2005). Among the three, orthologs retain similar functional characteristics as they evolve, thus being central in predicting protein structure and function (Koonin, 2005; Xiong, 2006).

2.1.1 Database similarity search tool

Several public biological databases are made available and are divided into three broad categories: primary of raw sequence data, secondary containing curated and annotated data derived from the primary databases, and specialized focusing on data of specific organisms of particular research interest (Xiong, 2006). The National Centre for Biotechnology Information (NCBI) database (<https://www.ncbi.nlm.nih.gov/>), the Universal Protein Resource (UniProt) (<http://www.uniprot.org>) and the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) (<https://www.rcsb.org/>) are popular main primary biological databases. PROSITE (<https://prosite.expasy.org/>) and PlasmoDB (<http://plasmodb.org/plasmo/>) are examples of common secondary and specialized databases, respectively.

In order to search the available biological databases, different search tools based on similarity searching are made available (Pearson and Lipman, 1988; Altschul *et al.*, 1990; Brandt and Heringa,

2009; Finn, Clements and Eddy, 2011; Jaroszewski *et al.*, 2011; Remmert *et al.*, 2012). The Basic Local Alignment Search Tool (BLAST) (Altschul *et al.*, 1990) is one of the most common and fundamental sequence similarity search tools. The BLAST can be used as a stand-alone program or through a variety of web servers, such as the NCBI (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>), UniProt (<http://www.uniprot.org/blast>) and the PlasmoDB (<http://plasmodb.org/plasmo/app/search/transcript/UnifiedBlast>, to mention a few. The output of BLAST is a list of sequence matches ordered according to a statistical significance assessed by the alignment score and the expectation value (E-value). The E-value defines the likelihood of a match occurring by chance from the database. Thus, the lower the E-value is the higher the biological significance of the alignment. The BLAST tool adopts heuristic algorithms for rapid and optimal local alignments within biological databases (Boratyn *et al.*, 2013). For remote sequence homology detection, one can use the Position-Specific Iterative Basic Local Alignment Search Tool (PSI-BLAST), which constructs and carries out a BLAST search using a custom position-specific scoring matrix from the multiple sequence alignments (Bhagwat and Aravind, 2007).

Another common sequence similarity search tool for identifying protein homology is the HHpred (<http://toolkit.tuebingen.mpg.de/hhpred>). The tool uses Hidden Markov models (HMM-HMM comparison) (Söding *et al.*, 2005) for more sensitive and speedy searches/detection of remote homologs and predicting protein secondary structures. The HMM profiles are more accurate than the traditional position-specific scoring matrices and achieve better alignments; due to their position-specific gap penalties (So ding 2005). HHpred results display the alignment and alignment statistics, including the probability of homology to the query, percentage identity, similarity, and the E-value.

2.1.2 Sequence alignment strategies and algorithms

Sequence comparison is achieved through alignment, in which regions of similarity and residue to residue correspondence is searched (Xiong, 2006). The sequence alignment analysis is based on sequence similarity, which designates the alignment of residues of similar physicochemical properties and the sequence identity indicating the alignment of matching residues.

The sequence comparison can either be carried out globally or locally across the sequences. The global alignment involves aligning the sequences across their entire length; in that way, every sequence is aligned (Needleman and Wunsch, 1970). Whereas in local alignment, only local regions of similarity are aligned (Smith and Waterman, 1981). The global alignment is most useful when query sequences set are similar and of roughly equal size whereas, the local alignment is ideal for sequences that are often long and widely divergent but only contain similar domains. Various sequence alignment algorithms have been developed, including slow yet accurate algorithms like dynamic programming and fast but less accurate heuristic or probabilistic algorithms (Xiong, 2006; Finn, Clements and Eddy, 2011). As the number of sequences increases, dynamic programming quickly becomes impractical, hence considered computationally expensive and time-consuming for aligning many sequences.

2.1.3 Alignment scoring schemes

During sequence alignment, the sequences are arranged so that a maximum number of residues from each sequence are matched up according to a scoring function. The sequence alignments are usually scored using either a protein scoring matrix or a consistency-based algorithm (Chatzou *et al.*, 2016). The protein scoring matrix contains predefined empirical values on substitution rates of amino acids over time. It can determine how favourable a substitution is based on similarity in physicochemical property deduced from observed alignments. The most commonly used protein scoring matrices are the Point Accepted Mutations (PAM) and Blocks of Amino Acid Substitution Matrix (BLOSUM) (Xiong, 2006). PAM matrices are based on scoring all amino acid positions; thus, it represents the likelihood of an amino acid substitution to another, whereas the BLOSUM matrices are based on substitutions and conserved positions in blocks; thus, it represents the most conserved common regions in related sequences (Mount, 2008).

On the other hand, the consistency-based methods harness information on the substitution rates from regions that are consistently aligned among a set of pairwise superpositions to create substitution matrices (Ebert and Brutlag, 2006; Chatzou *et al.*, 2016). The score of the alignment is then calculated as the sum of substitution and the and gap scores. The gap scores are calculated as the sum of gap

opening and extension penalties. Thus, for a gap of length n , the gap cost would be $G+L_n$. Where G is the gap opening cost and L is the gap extension cost of n length; in this manner, the total gap penalty is a linear function of gap length (Xiong, 2006).

2.1.4 Multiple Sequence Alignment

Multiple sequence alignment (MSA) is a natural extension of the pairwise alignment; it involves the alignment of multiple sequences often identified through the database similarity searching tool to determine the degree of similarity across multiple sequences (Chowdhury and Garai, 2017). The information obtained from the MSA can be essential for the prediction of unknown protein structures and function, identification of functionally conserved domains, unique sequence patterns, and sequence variants.

Different approaches and scoring functions are developed to attain a fast and more accurate MSA and can be broadly divided into exhaustive and heuristic approaches. The exhaustive alignment approach involves examining all possible aligned positions simultaneously, similar to dynamic programming in pairwise alignment. However, in this approach, a multidimensional search matrix is created for aligning N sequences; consequently, an N -dimensional matrix is required to be filled with alignment scores. Back-tracking is then applied through the N -dimensional matrix to find the highest scored path representing the optimal alignment. As the number of sequences increases, computational and memory usage increases exponentially (exponential running time). Hence, the exhaustive approach is computationally demanding and limited to a small number of relatively short sequences. In order to overcome the problem and further reduce the computational cost, heuristic approaches are used as alternatives. The heuristic approaches can be divided into three main categories: progressive, iterative, and block-based approaches (Chowdhury and Garai, 2017).

Progressive approach

In the progressive approach, the most similar sequences are first aligned based on a heuristic pairwise alignment method of the Needleman–Wunsch global alignment (Needleman and Wunsch, 1970). This is followed by recording the similarity scores from the pairwise comparisons to build a guide tree

describing the sequence relatedness, then successively adding less related sequences or groups to the alignment until the entire query set has been incorporated into a solution. The drawback of this approach is its “greedy” nature, in which the alignment results are reliant on the initial pairwise alignments. Meaning if any errors were introduced in the initial steps of the alignment, they remain fixed and can not be corrected. Nevertheless, current progressive methods can reduce the likelihood of making a poor choice of initial sequences by weighting sequences in the query set according to their relatedness, position-specific gap penalties, and weight matrix choice (Thompson, Higgins and Gibson, 1994; Rédei, 2008). Popular alignment programs that employ this algorithm include KAlign (Lassmann and Sonnhammer, 2005), Clustal Omega (Sievers *et al.*, 2011), ClustalW (Thompson, Higgins and Gibson, 1994), PROMALS (Pei and Grishin, 2007), and T-COFFEE (Notredame, Higgins and Heringa, 2000).

Iterative approach

An initial global alignment is first performed in the iterative approach, followed by realigning sequence subsets iteratively to an ideal optimal solution. This is based on the idea that an optimal solution can be found by iteratively modifying existing suboptimal solutions. In that way, iterative realignment can gradually improve an initial low-quality alignment until no further improvements in the alignment scores. MAFFT (Kato, 2002) and MUSCLE (Edgar, 2004) are some of the most popular programs that employ this approach. The iterative approach can provide more accurate alignment results, but they can only align a few hundred sequences (Daugelaite, O’ Driscoll and Sleator, 2013). Since both progressive and iterative approaches are global alignment-based, they can fail to identify local regions of high similarity, hence may not be suitable for comparing highly divergent sequences of varying lengths.

Block-based approaches

Instead of starting from low order pairwise alignments, the block-based approach aligns conserved regions into blocks rather than continuous sequences, then aligning the regions between successive blocks to form a final alignment. DIALIGN2 (Morgenstern, 1999) and Match-Box (Depiereux and Feytmans, 1992) are examples of block-based alignment algorithms. Lastly, the MSA can also be

constructed by using existing protein structural information. The structural information can increase the accuracy of the final MSA and provide better alignment since structures evolve more slowly (far more conserved than sequences) (Illergård, Ardell and Elofsson, 2009). T-COFFEE (O'Sullivan *et al.*, 2004), Espresso (Armougom *et al.*, 2006), and MICAAlign (Xia *et al.*, 2009) are examples of popular structure-based MSA programs.

2.1.5 Phylogenetic analysis

The phylogenetic analysis allows studying the ancestral origin and evolutionary divergence/relationship among various species/sequences over time. Phylogenetic trees are often used to graphically represent the evolutionary relationships by showing the evolutionary divergence or similarity of the species/sequences involved (Sarma *et al.*, 2018). The phylogenetic analysis requires an accurate MSA and suitable evolutionary model to approximate the evolutionary distance between lineages (Xiong, 2006). A suitable evolutionary model is selected by scoring a set of sequence evolution models to the data with some optimality criteria (Tamura *et al.*, 2011). After selecting a suitable evolutionary that best describes the observed pattern of sequence variation, tree construction follows. In order to construct a phylogenetic tree, two main methods can be employed: distance-based or character-based. The distance-based method is further subdivided into clustering-based or optimality-based. The maximum parsimony and maximum likelihood methods are the most popular character-based approaches. An explanation of the different tree-building approaches is well detailed by Xiong (Xiong, 2006).

Lastly, a phylogenetic tree can be either rooted or unrooted, but it is most common to provide unrooted trees. To root a tree, one should carefully choose an outgroup that is not very distant from the sequences because it can cause topological errors/distortion to the data, nor a closely related one to the sequences, which in this case it may not be a true outgroup (Opperdoes and Lemey, 2012). The Molecular Evolutionary Genetics Analysis (MEGA) is one of the most widely used software packages for phylogenetic analysis (Tamura *et al.*, 2011). The program includes tools to align sequences, estimate the evolutionary distance between the sequences and infer phylogenetic trees. MEGA further allows for visualization of the constructed tree in a tree explorer. It employs various empirical models

of protein evolution (amino acids substitution) for phylogenetic inference from the amino acid sequence data, such as the Dayhoff model (Dayhoff, Eck and Park, 1972; Dayhoff, Schwartz and Orcutt, 1978), the Bishop–Friday model, Jones, Taylor, and Thornton amino acid substitution models (Jones, Taylor and Thornton, 1992) and the Whelan and Goldman (WAG) model (Whelan and Goldman, 2001), to mention a few. All models differ at the amino acid substitution rates, which are experimentally determined. MEGA also employs different tree-building algorithms of character and distance-based.

2.1.6 Motif discovery

Motifs are evolutionarily conserved and unique sequence patterns that underlie biological functions (Xiong, 2006). It can further highlight key residues that are unique and typically required to retain protein function and structural stability (Mackenzie and Grigoryan, 2017; Ross *et al.*, 2017; Zheng and Grigoryan, 2017). Locating sequence motifs within protein structures can shed light on sites that modulate the function of proteins. The identification of sequence motifs location in a global MSA is known as motif finding or profile analysis. Motif finding tools identify short and highly conserved sequence patterns within the global MSA by constructing a matrix similar to a substitution matrix that reflects the amino acid or nucleotide composition of each position in the putative motif. Several algorithms are developed for motif findings and classified into four main classes: enumerative, probability, nature-inspired, and combinatorial, with each algorithm containing several subclasses (Cieplak *et al.*, 1995). The Multiple Expectation Maximisation for Motif Elicitation (MEME) is one of the most popular methods based on the probabilistic approach. It utilizes statistical techniques to describe each ungapped motif identified, length, occurrence frequency, and position(s) within the queried sequences (Bailey *et al.*, 2015). Another common method is the GenomeNet web server motif finding, which accompanies the motif finding process with functional annotation by further querying the discovered motifs through a sequence annotation database such as the protein family's (Pfam) database (Kanehisa *et al.*, 2002).

2.2 Methods

2.2.1 Sequence retrieval

The amino acid sequences of the *Plasmodium falciparum* (*P. falciparum*) guanosine-5'-triphosphate (GTP) cyclohydrolase I (GCH1) and 6-pyruvoyl tetrahydropterin synthase (PTPS) enzymes were first retrieved from PlasmoDB database (<http://plasmodb.org/plasmo/>); with the accession numbers PF3D7_1224000 and PF3D7_0628000, respectively. The retrieved sequences were then used as a query to search for other Plasmodium homologs sequences within the PlasmoDB database and homologs sequences from mammals, fungi, and bacteria within the NCBI (<https://www.ncbi.nlm.nih.gov/>) and UniProt (<http://www.uniprot.org>) databases. The most probable homolog sequences exhibiting low E-values and high sequence identity to the query sequences were selected for the subsequent analysis.

2.2.2 Multiple sequence alignment

The MSA was performed using three alignment tools: The Profile Multiple Alignment with Local Structures and 3D constraints (PROMALS3D), Tree-based Consistency Objective Function Evaluation (T-COFFEE), and the Multiple Alignment using Fast Fourier Transform (MAFFT) (Kato, 2002). The ESPript webserver (Robert and Gouet, 2014) was used to render sequence similarities and visualise the sequence alignments. The resultant MSAs from the different tools were compared to evaluate the alignment accuracy and find consensus. The MAFFT sequence alignments were selected for the phylogenetic analysis and the calculation of the pairwise sequence identity. Lastly, all versus all sequence identity of the aligned sequences was calculated using an in-house Python script (Hatherley *et al.*, 2015) and MATLAB software.

2.2.3 Motif discovery

Motif discovery was performed on both GCH1 and PTPS retrieved sequences to identify unique sequence patterns belonging to these enzymes then identify their location on the 3D structures. The *de novo* Motif discovery was performed using the MEME suite version 4.11.2 (Bailey *et al.*, 2015). The search size of unique motifs was set to a range between 6 and 20 residues. The Motif Alignment

Search Tool (MAST) was then used to detect overlapping motifs (Bailey and Gribskov, 1998). A heat map representing the occurrence and length of motifs at each position among the different sequences was generated using an in-house Python script (Hatherley *et al.*, 2015) and the MAST and MEME log data files as inputs. The identified unique motifs of the *P. falciparum* GCH1 and PTPS were then mapped onto their respective structures and visualized using PyMOL Molecular Graphics System (DeLano, 2014).

2.2.4 Phylogenetic analysis

Phylogenetic analysis was carried out on both *P. falciparum* GCH1 and PTPS MSA datasets using MEGA version 7.0 (Kumar, Stecher and Tamura, 2016). The best evolutionary/substitution models for phylogenetic tree construction were selected according to the lowest three Bayesian Information Criterion (BIC) scores. The top three evolutionary models with the lowest BIC scores for each of the top three evolutionary models with the lowest BIC were then evaluated using a 100%, 95%, and 90% sequence gaps deletion. Phylogenetic trees were then constructed of the best three models at each gap deletion using the character-based Maximum likelihood method. A thousand bootstrap replicates and a strong branch swap filter was used in each tree construction. The bootstrap test provides a measure to evaluate the confidence levels of the tree topology. Lastly, a comparison between each generated tree and the corresponding bootstrap tree was performed to observe the overall branching pattern and determine the most reliable phylogenetic tree.

2.3 Results and Discussion

2.3.1 Sequence retrieval

Protein sequences of GCH1 from 22 different species, including 10 *Plasmodium* species, four mammalian species, four bacterial species, and four fungal species (Table 2.1), were retrieved using the *P. falciparum* GCH1 sequence (Accession ID: PF3D7_1224000) via BLAST search. Protein sequences of PTPS from 20 different species, including nine *Plasmodium* species, four mammalian species, four bacteria species, and three fungal species (Table 2.2), were retrieved using the *P. falciparum* PTPS sequence (Accession ID: PF3D7_0628000). Both tables below show information of the retrieved sequences, E-value, and corresponding sequence identities.

Table 2.1. GCH1 sequences retrieval information

Species	Accession Number	E-value	Identity
<i>Plasmodium falciparum</i>	PF3D7_1224000	0.00	100.00%
<i>Plasmodium vivax</i>	PVX_123830	2.00×10^{-82}	45.00%
<i>Plasmodium malariae</i>	PmUG01_14058300	2.00×10^{-85}	68.00%
<i>Plasmodium ovale</i>	PocGH01_14049700	6.00×10^{-95}	70.00%
<i>Plasmodium knowlesi</i>	PKNH_1443200	7.00×10^{-87}	69.00%
<i>Plasmodium chabaudi</i>	PCHAS_1440900	2.00×10^{-85}	70.00%
<i>Plasmodium yoelii</i>	PY17X_1441400	3.00×10^{-87}	70.00%
<i>Plasmodium gaboni</i>	PGSY75_1224000	0.00	86.00%
<i>Plasmodium berghei</i>	PBANKA_1438900	4.00×10^{-88}	68.00%
<i>Plasmodium reichenowi</i>	PRCDC_1223300	0.00	98.00%
<i>Colletotrichum orchidophilum</i>	A0A1G4BKI0	6.80×10^{-30}	44.10%
<i>Blumeria graminis</i>	A0A061HDS3	1.60×10^{-31}	43.20%
<i>Pseudogymnoascus</i> sp. VKM F-4517 (FW-2822)	A0A094GQ82	4.30×10^{-30}	43.20%
<i>Fusarium oxysporum</i>	X0M7M4	1.90×10^{-30}	44.40%
<i>Rhodothermus profundus</i>	A0A1M6TV79	5.40×10^{-35}	47.80%
<i>Gramella</i> sp. LPB0144	A0A1L3J3H7	3.10×10^{-35}	45.00%
<i>Thermus thermophilus</i>	Q5SH52	2.90×10^{-32}	39.2%
<i>Escherichia coli</i> (strain K12)	P0A6T5	5.60×10^{-19}	33.8%
<i>Pan troglodytes</i> (Chimpanzee)	H2RBI2	3.40×10^{-28}	35.40%

<i>Mus musculus</i> (Mouse)	Q05915	1.70x10 ⁻²⁹	35.30%
<i>Rattus norvegicus</i> (Rat)	P22288	5.30x10 ⁻²⁸	36.00%
<i>Homo sapiens</i> (Human)	P30793	3.40x10 ⁻²⁸	35.40%

Table 2.2. PTPS sequences retrieval information

Species	Accession Number	E-value	Identity
<i>Plasmodium falciparum</i>	PF3D7_0628000	1.00x10 ⁻¹²³	100.00%
<i>Plasmodium reichenowi</i>	PRCDC_0626400	4.00x10 ⁻¹²¹	98.00%
<i>Plasmodium gaboni strain</i>	PGABG01_0626800	1.00x10 ⁻¹¹⁶	96.00%
<i>Plasmodium malariae</i>	PmUG01_11035600	3.00x10 ⁻⁹²	73.00%
<i>Plasmodium ovale curtisi</i>	PocGH01_11029300	1.00x10 ⁻⁸⁹	73.00%
<i>Plasmodium knowlesi</i>	PKNH_1121800	2.00x10 ⁻⁸⁶	68.00%
<i>Plasmodium vivax</i>	PVP01_1121500	7.00x10 ⁻⁸⁶	69.00%
<i>Plasmodium chabaudi</i>	PCHAS_1126200	5.00x10 ⁻⁸⁰	65.00%
<i>Plasmodium berghei</i>	PBANKA_1126700	1.00x10 ⁻⁷⁹	65.00%
<i>Plasmodium yoelii</i>	PY17X_1128200	2.00x10 ⁻⁷⁸	64.00%
<i>Gemmata sp. SH-PL17</i>	A0A142XE24	2.60x10 ⁻²⁵	34.60%
<i>Gemmatimonadetes bacterium</i>	A0A2V7MAP3	7.40x10 ⁻²²	33.50%
<i>Acidobacteria bacterium</i>	A0A2V8BFN7	2.90x10 ⁻¹⁹	30.50%
<i>Actinobacteria bacterium</i>	A0A2M7TB79	1.80x10 ⁻²²	34.20%
<i>Rhizophagus irregularis</i>	A0A2I1GC34	1.90x10 ⁻¹⁰	25.10%
<i>Rhizopus microsporus</i>	A0A0A1ND78	0.00	24.81%
<i>Spizellomyces punctatus</i>	A0A0L0HI22	3.10x10 ⁻²⁵	23.35%
<i>Homo sapiens</i> (Human)	Q03393	3.80x10 ⁻¹	26.20%
<i>Rattus norvegicus</i> (Rat)	P27213	1.70x10 ⁻¹	28.20%
<i>Pan troglodytes</i> (Chimpanzee)	H2R4Y8	0.00	26.10%
<i>Ailuropoda melanoleuca</i> (Giant panda)	G1LBC1	0.00	27.50%

2.3.2 Multiple sequence alignment

The MSA was performed to identify highly conserved sequence regions, often assumed to be structurally and functionally essential and Plasmodial unique sequence regions that can be exploited to design and develop drugs of high selectivity against the malaria parasite enzymes. Both GCH1 and PTPS MSAs are presented and discussed below.

GCH1 MSA

The Plasmodium GCH1 sequences were compared to other homologs, including the human GCH1. The GCH1 MSA result is shown in Figure 2.1. The MSA of the *P. falciparum* GCH1 showed high conservation of the protein active site and surrounding residues, including the key catalytic active site residues H280, C277, and C348. Notably, these residues are also the coordinating residues of the GCH1 active site metal ion. Previous studies reported that mutations involving any of these residues result in the loss of protein activity (Nar, Huber, Auerbach, *et al.*, 1995).

Other highly conserved residues include the positively charged active site residues that interact with the substrate phosphate group, such as R231, R306, R352, and K303. H346 was also conserved in all the studied sequences; H346 was previously reported as a catalytically important residue that causes the protonation of the substrate ring N-7, resulting in the cleavage of N-7/C-8 (ring-opening) (Kümpornsin, Kotanan, *et al.*, 2014). The active site residue K278 was also conserved among the Plasmodium species and substituted by a Glu residue in all other species. Other residues near the active site region R313 to E319 were also conserved across all species except for the *E. coli* GCH1 sequence in which the R313 was substituted with a Gln residue. The conserved residue E319 is known to form two hydrogen bonds with the GTP guanine ring N1 and N2 atoms. Another highly conserved catalytic residue is S302 which interacts with the 2' and 3' hydroxy groups of the GTP ribose ring (Rebelo *et al.*, 2003). S302 was conserved in all sequences. Overall, the GCH1 active site region was highly conserved; the residues responsible for key interactions with the GTP substrate were also well conserved, which indicates their crucial role/importance in maintaining the protein function (Figure 2.1).

From the MSA, sequence regions that are unique to the Plasmodium were also characterized. This includes K355, E356, and H357 located in the loop region near the active site. K330, K331, and Y332 were conserved in the Plasmodium species and substituted in the mammalian sequences by a Thr, Glu, and Ala, respectively (Figure 2.1). The human GCH1, D127 to M129, L157 to N159, and the C-terminal residues T229 to E243 constitute the effector molecule tetrahydropterin (BH4) binding site

(Maita *et al.*, 2004), in the Plasmodium sequence, this site showed high sequence variation and was not conserved. The Human GCH1 BH4 binding site residues correspond to the *P. falciparum* N263 to I265, I293 to N295, and A365 to H377. The sequence variation within the BH4 binding site residues can be reasoned to the distinct role of GCH1 in higher eukaryotes as a part of the BH4 biosynthesis pathway (details about the role of GCH1 in higher eukaryotes are presented in section 1.1.10 of [Chapter 1](#)). Additionally, two key residues of the human GCH1, R235 and R241, were reported to interact with the BH4 and further induce notable conformational changes (Ebenhoch *et al.*, 2020); the two residues were substituted in the *P. falciparum* GCH1 by K371 and H377. Lastly, high sequence variation was identified at the GCH1 N-terminal and C-terminal regions, thus demonstrating the protein terminus as the least conserved sites (Figure 2.1). Notably, the Plasmodium species have an extended N-terminal domain of about 135 residues absent in the human GCH1 (Kümpornsin, Kotanan, *et al.*, 2014). The unique and extended N-terminal region of the Plasmodium GCH1 was proposed to exert a strong regulation effect on the catalytic domain (Funderburk *et al.*, 2006).

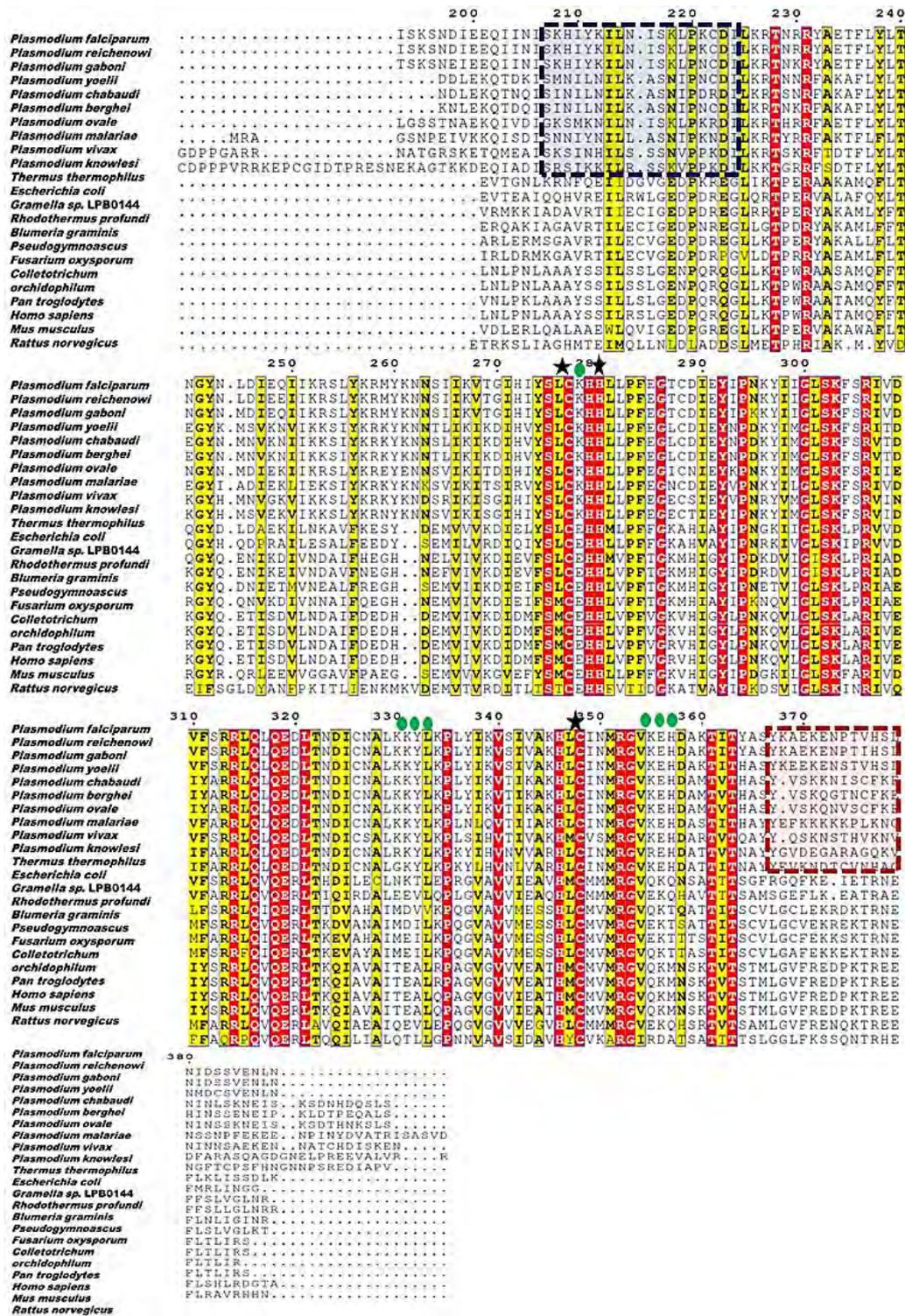


Figure 2.1. MSA of the GCH1 retrieved sequences. The black stars illustrate the catalytic and metal-coordinating residues. The green circles show the functionally important active site neighbouring residues. Identical/highly conserved residues are printed in white text and highlighted in red and similar residues are printed in black text and highlighted in yellow. The blue and red dashed boxes show Motif 5 and Motif 8 of the Plasmodium sequences, respectively.

PTPS MSA

The MSA of the *P. falciparum* PTPS shown in Figure 2.2 also displayed the high level of conservation of the active site residues in all the studied PTPS sequences. This includes the catalytically essential residues H29, H41, and H43 which are also responsible for the metal ion coordination. Other catalytically key residues were also found among the highly conserved residues, such as E131, H80, E107, G42, N44, Y45, D89, and E161. T127, located at the bottom of the active site pocket, was only conserved in the Plasmodium and mammalian sequences and substituted by Val or Ala in the bacterial and fungal PTPS, respectively. S126, T127, E128, located near the active site pocket, constitute an acceptor site for the substrate ring during catalysis were conserved in all sequences, except S126 substituted by Thr in the mammalian PTPS sequences and T127 replaced by Val/Ala in the bacterial sequences.

The plasmodial PTPS sequences were notably characterized by the presence and absence of inserts at specific regions in the MSA profile. The mammalian sequences had an insertion lacking in the Plasmodium PTPS sequences between residue E38 and T39. At the same time, the Plasmodium PTPS sequences had a sequence insertion between S88 to P113, which was absent in the mammalian PTPS sequences. Overall, the PTPS MSA profile illustrates key differences between the Plasmodium PTPS and the mammalian PTPS, such as insertion and absence of residues stretches and functional key residues substitutions. To sum up, the characterized sequence conservation and differences of both Plasmodium GCH1 and PTPS can aid in developing parasite-specific inhibitors.

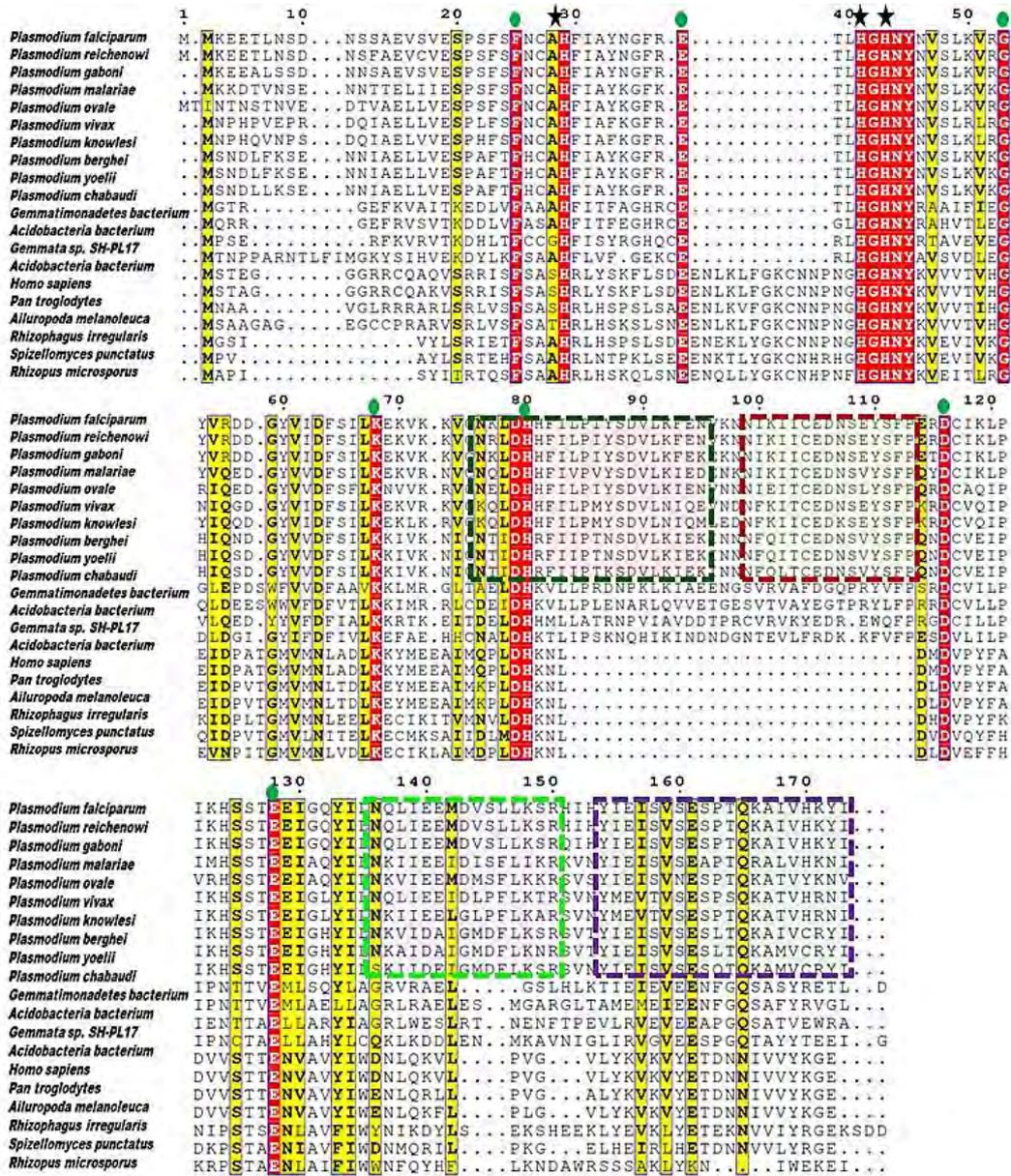


Figure 2.2. MSA of the PTPS retrieved sequences. The black stars illustrate the catalytic and metal-coordinating residues. The green circles illustrate the functionally important active site neighbouring residues. Identical/highly conserved residues are printed in white text and highlighted in red and similar residues are printed in black text and highlighted in yellow. The forest green, red, lime green and purple dashed boxes show Motif 5,6,7 and 8 of the Plasmodium sequences, respectively.

2.3.3 Pair-wise sequence identity

From the GCH1 pairwise sequence identity values presented in a heatmap (Figure 2.3-A), all Plasmodium GCH1 sequences showed sequence identity above 34%. The *P. vivax* showed the least sequence identity relative to the *P. falciparum*, whereas *P. reichenowi* and *P. gaboni* shared the

highest sequence identity with the *P. falciparum* of 98% and 88%, respectively. The varied sequence identity values between the Plasmodium GCH1 sequences illustrate a high level of variation/diversity within the Plasmodium GCH1 sequences that can be attributed to the extended heterogeneous nature of the GCH1 N-terminal regions. On the other hand, the mammalian GCH1 showed sequence identities below 28% relative to the Plasmodium sequences.

The Plasmodium PTPS sequences shared a sequence identity above 73%, indicating more similarity within the Plasmodium PTPS species. Notably, the *P. falciparum*, *P. reichenowi*, and *P. gaboni* PTPS shared the highest sequence identity above 95%. The mammalian PTPS sequences shared the lowest sequence identity of 30% relative to the Plasmodium sequence; the sequence features contributing to the low similarity can be further explored to obtain selective targeting of the Plasmodium enzymes. The PTPS pairwise sequence identity heatmap is shown in (Figure 2.3-B).

2.3.4 Phylogenetic analysis

The authenticity of the generated phylogenetic tree to represent the phylogeny of the data was assessed statistically via a bootstrap test. Bootstrapping involves the iterative alteration of the data-set and creating trees that are then compared against the original tree to test for robustness. In other words, the bootstrap tree values are indicative of the confidence levels of the topology. The Le and Gascuel model with gamma distribution and invariant sites (LG + G + I) model at a 100% gap deletion calculated the best phylogenetic tree. The resultant GCH1 phylogenetic tree shown in Figure 2.1-C displayed four distinct clusters of the plasmodial, bacterial, fungal, and mammalian GCH1. The rodent infective Plasmodium species *P. berghei*, *P. yoelii* and *P. chabaudi* clustered together within the Plasmodium species, demonstrating their close relatedness. The *P. vivax* and *P. knowlesi*, on the other hand, were seemingly divergent from other Plasmodium species as they formed a separate cluster together. *P. falciparum* shared its closest ancestor with the chimpanzee parasites *P. reichenowi* and *P. gaboni* and clustered together. The mammalian GCH1 sequences were clustered together and showed the least similarity to the plasmodial sequences (Figure 2.3-A and Figure 2.3-C).

The PTPS phylogenetic tree in Figure 2.3-D also displayed the distinct clustering of the Plasmodium PTPS sequences. Within the Plasmodium PTPS sequences, the rodent infective Plasmodium clustered together, the *P. ovale* known for causing a relatively mild form of malaria clustered alone and the *P. knowlesi*, and *P. vivax* previously reported to share a close phylogenetic relationship (Verzier *et al.*, 2019) clustered together. Lastly, the *P. falciparum*, *P. reichenowi*, and *P. gaboni* all clustered together. The mammalian PTPS sequences were also clustered together and showed the least similarity to the plasmodial sequences (Figure 2.3-B and Figure 2.3-D).

In summary, the phylogenetic analysis showed the distinct clustering of both plasmodial GCH1 and PTPS sequences, suggesting that the Plasmodial sequences have evolved at a different rate relative to the mammals and prokaryotes orthologs. The clustering patterns of both GCH1 and PTPS sequences were also in agreement with the MSAs and pairwise sequence identity heatmaps. Hence, we can conclude that the Plasmodial sequences are related by structural and evolutionary attributes and considerably different from their mammalian orthologs. These findings can be valuable for designing inhibitors to target the plasmodial GCH1 and PTPS proteins more selectively.

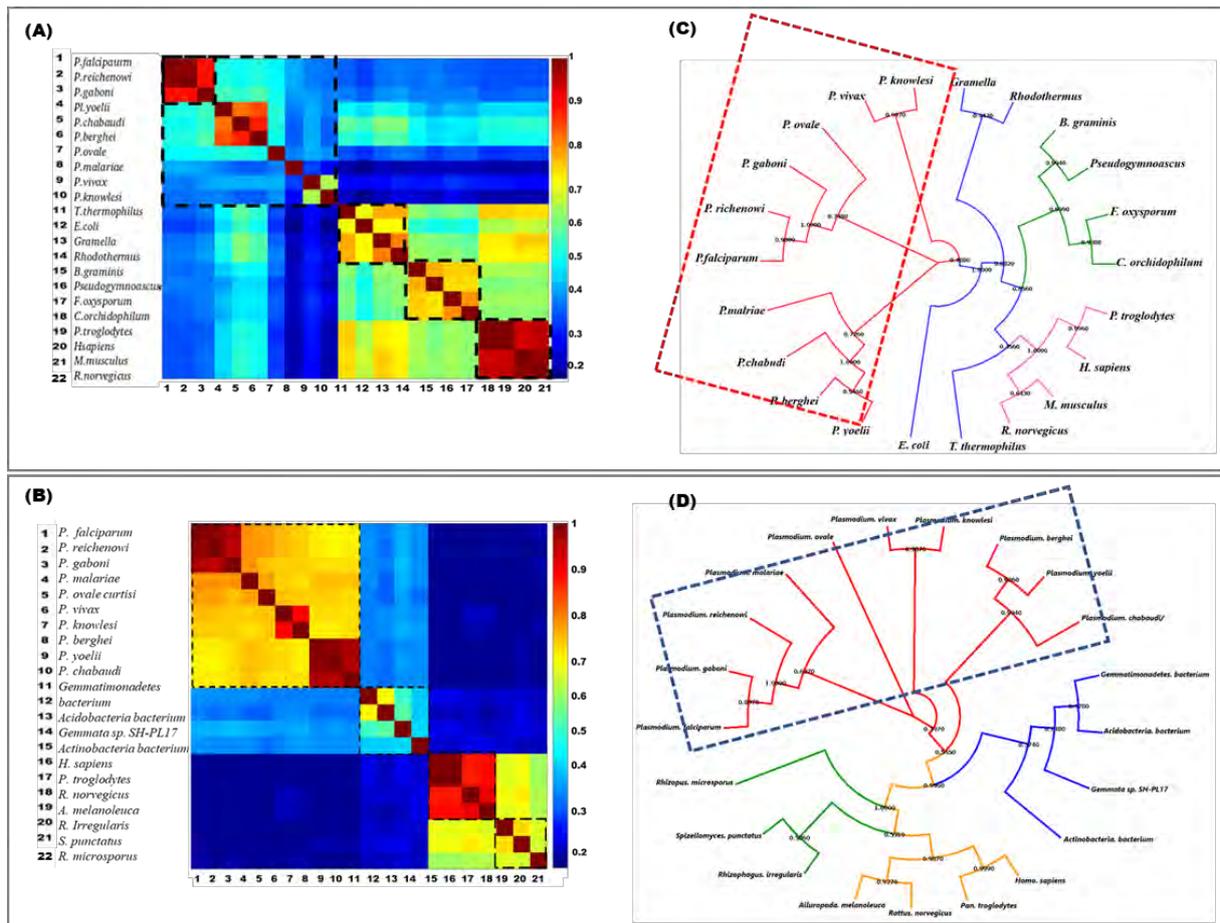


Figure 2.3. (A) GCH1 pair-wise sequence identity heat map. (B) PTPS pair-wise sequence identity heat map. The heat map shows the pairwise sequence identity scores of the MSA as a colour-coded matrix; the identity score increases from blue to red /least to most conserved sequences. (C) Phylogenetic tree of the retrieved GCH1 sequences. The red dashed square highlights the Plasmodium GCH1 sequences. (D) Phylogenetic tree of the retrieved PTPS sequences. The red dashed square highlights the Plasmodium PTPS sequences.

2.3.5 Motif analysis

The MEME suite returned sequence motifs as sequence logos with their corresponding scores and positions and the E-value of each motif. The provided E-value indicates the statistical significance of the discovered motifs based on the log-likelihood ratio of the returned motif with the same width and site count (Bailey *et al.*, 2015). The sequence logo shows a stack of the one-letter amino acids code, the height of the letters indicates the probability (in bits) of the amino acid/residue occurring at that position multiplied by the number of times that residue occurs within that site in each motif site in the total dataset (Bailey and Elkan, 1994; Bailey *et al.*, 2015). Thus, the height of the stack is reduced for

residues that are not well conserved at a particular position. The colour of the individual letters is based on the hydrophobicity properties of the amino acids (Kyte and Doolittle, 1982).

MAST calculation is done parallel to the MEME jobs within the MEME suite in which the pairwise correlation between each pair of motifs is calculated to determine the probability that the two motifs are significantly different. Consequently, when the correlation value of the motif pairs is high, they are considered similar and will not be treated as separate motifs (Bailey and Elkan, 1994; Bailey *et al.*, 2015).

GCH1 motif analysis

The motif discovery was carried out on the retrieved GCH1 sequences from the four mammalian species, including human, four bacterial species, four fungal species, and 10 other Plasmodium species (the same dataset of sequences used in MSA). A total of 21 motifs were identified (Figure 2.4-A). Except for the *E. coli* GCH1 sequence, seven common motifs were identified across all sequences (Motif 1 to 6 and Motif 9), the *E. coli* GCH1 lacked Motif 4. Motif 11 was unique to the rodent infective Plasmodium species *P. berghei*, *P. yoelii*, and *P. chabaudi*, and Motif 12 was only conserved among the *P. falciparum*, *P. reichenowi*, and *P. gaboni*. As we focus our motif study on the most virulent malaria parasite, the *P. falciparum* GCH1, two unique motifs belonging to the *P. falciparum* GCH1 were identified. The two motifs were well conserved in the *P. falciparum* GCH1 enzyme sequence and were not detected in any mammalian species. A heat map showing the occurrence of the GCH1 motifs is presented in Figure 2.4-B.

The identified unique motifs of the *P. falciparum* GCH1 were mapped onto its structure to determine their location (Figure 2.4-C). Motif 8 was located in the N-terminal helices region, whereas Motif 12 was situated in the GCH1 tunnel lining/central helices. Both motifs were unique to the Plasmodium sequences and exhibited notable sequence variability compared to the mammalian sequences (Figure 2.1). Notably, Motif 8 is located in the N-terminus, a previously reported region for sequence variability and further proposed as a regulatory site (Nar, Huber, Meining, *et al.*, 1995). Motif 12, however, located at the central five-helix bundle/central tunnel cavity, was not previously reported for

sequence variability. Here we show, the sequence variability relative to the mammalian sequences (Figure 2.1). Such unique sites of key functional residues can thus be considered in antimalarial drug design to ensure a more selective inhibition of the parasites GCH1 enzyme.

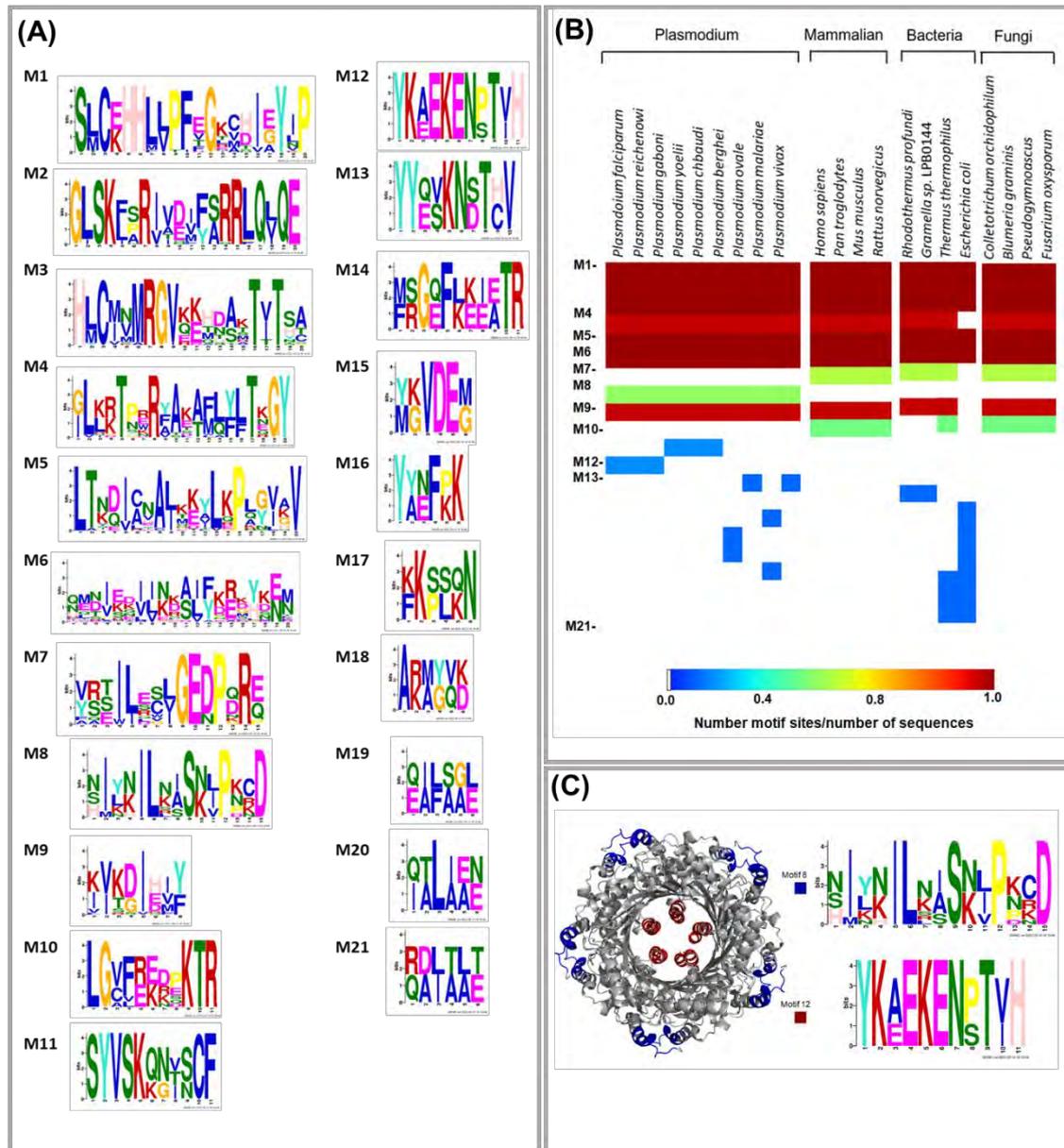


Figure 2.4. Motif analysis of the retrieved GCH1 sequences. (A) Motif numbering and web logos. The sequence logo contains a stack of letters at every position in the motif. The height of the letters demonstrates the probability (in bits) of the letter occurring at that position multiplied by the number of times that residue occurs within that site in each motif site in the total dataset (Bailey *et al.*, 2015). Residues are coloured based on their chemical properties: blue–hydrophobic residues (A, C, F, I, L, M, V and W); red–positively charged residues (K and R); green–polar, non-charged and non-aliphatic residues (N, Q, S and T); magenta–most acidic residues (D and E); light pink, orange, turquoise and

yellow are for H, G, Y and P respectively. (B) MEME heat map summarizing motif information for GCH1 homologue sequences. The white regions show sequences lacking a motif, and the level of conservation increases from blue to red. (C) The identified unique motifs of the *P. falciparum* GCH1 mapped onto its 3D structures. The sequence logos of the two unique motifs are shown next to the 3D structure.

PTPS motif analysis

The PTPS motif discovery was carried out on the retrieved sequences from four mammalian species, including humans, four bacterial species, three fungal species, and ten other Plasmodium species (the same dataset of sequences used in MSA). The PTPS motif analysis resulted in the identification of 27 motifs (Figure 2.5-A). Some of the motifs were common across all species, while others were conserved in certain species. Notably, motifs 1-4 were conserved among all the studied PTPS sequences. The analysis revealed four motifs (Motifs 5-8) conserved in the Plasmodium PTPS sequences and not detected in any mammalian species. Motifs 6-8 were uniquely conserved in the Plasmodium species, while Motif 5 was also found among the bacterial and fungal PTPS enzymes. A heat map illustrating the occurrence of the PTPS motifs is shown in (Figure 2.5-B). The conserved *P. falciparum* motifs were mapped onto its crystal structure to identify their location (Figure 2.5-C). Motif 5 was located in the central β -sheet strands forming the tunnel cavity, motif 6 and 7 were located in the N-terminal antiparallel β -strands, as well as a loop linking the β -sheet strands, and Motif 8 was located in the central α -helices region. Overall, the motif results illustrated notable sequence variation between the Plasmodium GCH1 and PTPS enzymes relative to their human homologs (Figure 2.2). As a result, the characterized key differences can be exploited in antimalarial drug design and development studies to ensure more drug selectivity.

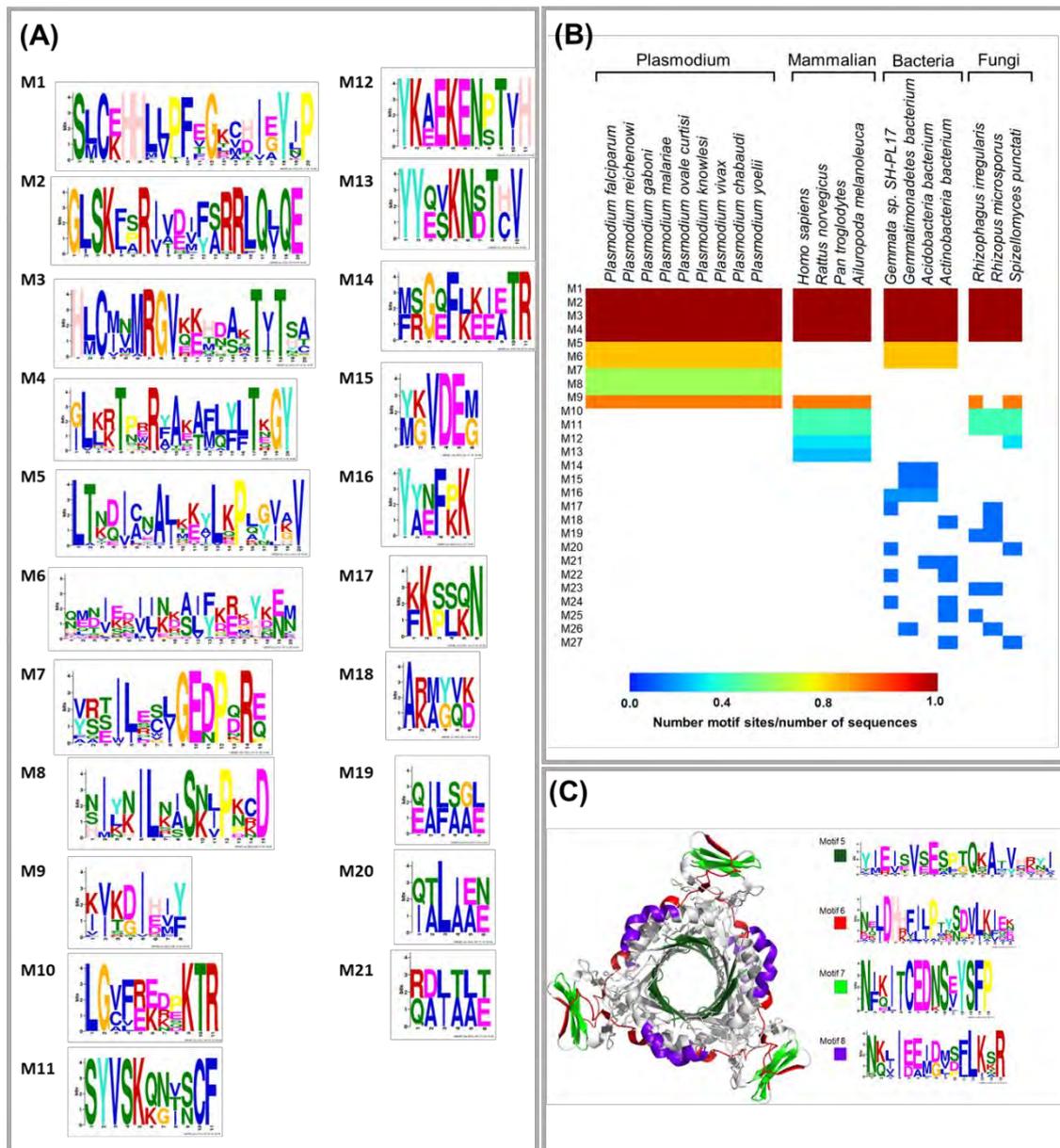


Figure 2.5. Motif analysis of the retrieved PTPS sequences. (A) Motif numbering and web logos. The sequence logo contains a stack of letters at every position in the motif. The height of the letters demonstrates the probability (in bits) of the letter occurring at that position multiplied by the number of times that residue occurs within that site in each motif site in the total dataset (Bailey *et al.*, 2015). Residues are coloured based on their chemical properties: blue–hydrophobic residues (A, C, F, I, L, M, V and W); red–positively charged residues (K and R); green–polar, non-charged and non-aliphatic residues (N, Q, S and T); magenta–most acidic residues (D and E); light pink, orange, turquoise and yellow are for H, G, Y and P respectively. (B) MEME heat map summarizing motif information for GH1 homologue sequences. The white regions show sequences lacking a motif, and the level of conservation increases from blue to red. (C) The identified unique motifs of the *P. falciparum* PTPS mapped onto its 3D structures. The sequence logos of the two unique motifs are shown next to the 3D structure.

2.4 Conclusions

In this chapter, sequence analysis was carried out involving the malaria parasite *de novo* folate biosynthesis pathway GCH1 and PTPS. The database search identified Plasmodial homolog sequences from different species and other orthologs from mammalian, bacterial, and fungal GCH1 sequences. The MSA was performed on the retrieved sequences and has disclosed key features that distinguish plasmodial homologs from humans. In both GCH1 and PTPS, the active site residues remained highly conserved across all species, demonstrating a well-conserved catalytic mechanism. The phylogenetic analysis showed the unique clustering of the Plasmodium sequences, illustrating the distinct evolutionary distance between the Plasmodium enzymes and their human homologs. The analysis further overlapped with predicted pairwise sequence identities. Lastly, the motif discovery disclosed motif 8 and 12 of the GCH1 and Motif 5, 6, 7, and 8 of the PTPS to be uniquely conserved in Plasmodium sequences but not in humans. The characterized motifs were mapped onto the structures and located at sites of unique structural features such as the proteins' central tunnels and terminus regions. Overall, the analysis highlighted key sequence and structural differences that can be further explored to design and develop antifolate antimalarial drugs with good selectivity toward the parasite enzymes.

Chapter 3

Force Field Parameters for the Zn^{2+} Ions of the Tunnelling-fold Enzymes GTP Cyclohydrolase I and 6-Pyruvoyl Tetrahydropterin Synthase

This chapter describes the development of force field parameters for the guanosine-5'-triphosphate cyclohydrolase I and 6-pyruvoyl tetrahydropterin synthase active site Zn^{2+} ions. The force field parameters were derived from the quantum mechanics scans of potential energy surfaces. The newly generated Zn^{2+} force field parameters of both enzymes were then evaluated via all-atom molecular dynamic simulations.

Chapter 3 is reproduced in part with permission from the following publication:

AMBER Force Field Parameters for the Zn^{+2} Ions of the Tunnelling-Fold Enzymes GTP Clohydrolase I and 6-Pyruvoyl Tetrahydropterin synthase. Afrak Khairallah, Özlem Tastan Bishop, and Vuyani Moses. *Journal of Biomolecular Structure and Dynamic*. 2020. 39 (16), 5843-5860. DOI: 10.1080/07391102.2020.1796800.

Authors Contribution: Ö.T.B. and V.M. conceived the project. A.K. performed the calculations and analysis under the supervision of V.M. and Ö.T.B.

Chapter overview

Both guanosine-5'-triphosphate cyclohydrolase I and 6-pyruvoyl tetrahydropterin synthase are metal-dependent enzymes that share the common fold topology of the tunnelling fold superfamily. The two enzymes have never been targeted before and stand as attractive drug targets. To accurately study and reproduce the catalytically important metals through all-atom molecular dynamics (MD) simulations, appropriate force field parameters describing the active site metal ions need to be developed and validated for use. Here, the bonded approach was used to derive the force field parameters, in which the parameters used to describe the metals bonded terms were obtained from a high level of theory quantum mechanics scans of the potential energy surface of the bonded terms (bonds, angles, and dihedrals). The resultant energy profiles were then fitted via the least square fitting approach to terms in AMBER potential, followed by MD simulations in an explicit water environment for a total of 100 ns in triplicates to ensure the localization of the metal ions in the active site and the steadiness of their coordination environment.

3.1 Introduction

3.1.1 Metalloproteins

Metal-containing proteins, also known as metalloproteins, make up around one-third of all known proteins (Lu *et al.*, 2012; Zhang and Zheng, 2020). Metal ions play key roles during the catalysis of many reactions as electron donors or acceptors and act as structural regulators (Riordan, 1977; Prejanò *et al.*, 2020). They also appear to play an essential role in determining the rate of enzyme activity and specificity (Okrasa and Kazlauskas, 2006; Sparta and Alexandrova, 2012; Kim *et al.*, 2020). Both guanosine-5'-triphosphate (GTP) cyclohydrolase I (GCH1) and 6-pyruvoyl tetrahydropterin synthase (PTPS) are metalloproteins. GCH1 has 10 zinc-containing active sites buried in a deep pocket at the interface of every three adjacent monomeric units of its homodecameric structure (Nar, Huber, Auerbach, *et al.*, 1995). The GCH1 enzymatic reaction is initiated with the attack of the GTP ring C-8 by a zinc-activated water molecule. The Zn^{+2} ion enhances the water molecule acidity to generate a hydroxyl nucleophile close to the imidazole ring of GTP, resulting in the attack of the GTP C-8 and ring-opening/hydrolysis, followed by an Amadori rearrangement of the sugar moiety and a final ring closure via a Schiff base reaction (Nar, Huber, Auerbach, *et al.*, 1995; Nar, Huber, Meining, *et al.*, 1995; Günter Auerbach *et al.*, 2000; Schramek *et al.*, 2002; Tanaka *et al.*, 2005). The GCH1 Zn^{+2} ion adopts a tetrahedral coordination geometry; it is coordinated to one histidine (His) residue through its ND1 atom and two cysteine (Cys) residues; the fourth coordination is occupied by a water molecule (Tanaka *et al.*, 2005).

The PTPS, on the other hand, have six zinc-containing active sites also buried in a deep pocket at the interface of every three adjacent monomeric units of its homo-hexameric structure. The PTPS Zn^{+2} ion is also important for the enzymatic reaction. It is responsible for activating two key residues in the active site: Cys42/Glu38 by lowering their pKa via electrostatic effects. The activated residues then abstract a proton from the substrate carbon atom, resulting in its ring-opening (Nar, 2011). The Zn^{+2} ion is also responsible for bringing the side-chain protons of the substrate closer to the active site residues by binding to the substrate two hydroxyl groups (Bürgisser *et al.*, 1995; Nar, 2011). The

PTPS Zn²⁺ ion adopts a tetrahedral coordination environment; it is coordinated to three His residues through their NE2 atoms and a water molecule that completes its tetrahedral coordination geometry.

3.1.2 Methods of molecular energy calculation

The energy of a molecule can be calculated by using either high-level quantum mechanical (QM) or low-level molecular mechanical (MM) methods. In the next section, we discuss both methods in detail.

Quantum Mechanics

The QM method is mainly based on the electronic structure theory (Stewart, 1970). In this method, molecules are treated as a collection of nuclei and electrons without reference to their chemical bonds. Then laws of QM are then applied to approximate the wave function of electrons by solving the time-independent Schrödinger equation (Equation 3.1). The Schrödinger equation is a linear partial differential equation that describes the wave function of a quantum mechanical system (Schrödinger, 1926). The calculations directly involve electronic interactions to describe molecules at the level of atoms and sub-atomic particles. Thermodynamic properties can then be deduced from the wavefunctions using the harmonic approximation. Based on the principle of QM, the physical properties of a given molecule are calculated directly from its fundamental physical quantities; thus, QM is regarded as first-principles calculations that do not require any empirical parameters.

$$H\psi = E\psi \quad \text{Equation 3.1}$$

H is the Hamiltonian operator of the system (corresponds to the system's total energy), **ψ** is the many-electron wave function, and **E** is the system's total energy.

The Hamiltonian operator can be generalized as:

$$E\psi = \psi \left[\frac{h^2}{2m} \sum_{i=1}^N \nabla_i^2 + \sum_{i=1}^N V(r_i) + \sum_{i=1}^N \sum_{j<i}^N U(r_i, r_j) \right] \quad \text{Equation 3.2}$$

Kinetic energy

Interaction between each electron and the collection of atomic nuclei

Interaction between different electrons

*The equation is given by taking a collection of atoms where \mathbf{M} is the nucleus and \mathbf{N} is the electrons.

The Schrödinger equation is further simplified by the Born–Oppenheimer approximation (Born and Oppenheimer, 1927; Wartak *et al.*, 2019). The approximation proposes that electronic and nuclear motions can be uncoupled because the mass of an atomic nucleus is much larger than the mass of an electron (Wartak *et al.*, 2019). As a result, one can neglect the motion of the atomic nuclei when describing the electrons, and the electronic wave function can be found as a solution to the electronic Schrödinger equation. Even though the nuclear kinetic energy terms are neglected, the resulting electronic wavefunction still depends upon the nuclear positions because the Born-Oppenheimer approximation still considers the variation in the initial positions of the nuclei (Scherrer *et al.*, 2017). The Schrödinger equation cannot be solved (not feasible) for any but a one-electron system that is the hydrogen atom and some trivial simple case. This is because complications arise in the many-electron systems (electrons spatial configuration) due to the inability to separate the electronic degrees of freedom $3 \times N$ or even $4 \times N$ (if we consider electronic spins) into N single-body problems. Therefore, an approximation is needed to provide a numerical solution to the Schrödinger equation while retaining as much of the key physics as possible.

In 1964, Hohenberg and Kohn introduced the density functional theory (DFT) as a numerical solution to the Schrödinger equation (Hohenberg and Kohn, 1964). The method proposed using the electron density as a fundamental parameter instead of the many-electron wave function. Thus, the total energy of a system is regarded as a unique function of its electron density. It was also proposed that the energy content reaches its absolute minimum when the charge density is that of the ground state; hence all properties of a certain system can be determined by the ground state charge density. In that way, the DFT takes the many-body problem and reorganizes it as a single-body problem. The DFT was not regarded as accurate enough for calculations in QM until it was able to better model the electrons exchange and correlation interactions in the 1990s (Gross and Kohn, 1990). Ever since, the DFT method has gained popularity and became the most widely used method for electronic structure calculations to recognize characteristic properties of molecules such as vibrational frequencies, dipole moments, and free energy of reactions (Arodola and Soliman, 2017; Chan, 2017; Mardirossian and

Head-Gordon, 2017; Lu and Gao, 2018; Postils *et al.*, 2018; Zhang, Shen and Yang, 2019; Hossain *et al.*, 2020). The DFT methods are included in various software packages such as Gaussian (Frisch *et al.*, 2009), GAMESS (Schmidt *et al.*, 2016), HyperChem (Froimowitz, 1993), Jaguar (Bochevarov *et al.*, 2013), VASP (Hafner, 2008), and Q-Chem (Krylov and Gill, 2013), to mention few.

From the 1960s to the 1980s, the use of DFT calculations was mostly limited by the inaccessibility to powerful computers. Nowadays, with the current advancements in computing power, one can perform such calculations in a single computer for relatively small and medium systems. Nevertheless, QM calculations of more than 100 atoms exceed the present capacity. Attempts to increase the capacity of these calculations are made, including the implementation of the graphics processing units (GPU) and parallelization in DFT calculation (Walker and Götz, 2016; Seritan, Bannwarth, B. Scott Fales, *et al.*, 2020). The attempts were further extended to an open-source, user-friendly, GPU-enabled quantum chemistry software by the open-source Quantum Interaction Computational Kernel (QUICK) program (Manathunga *et al.*, 2020). Lastly, QM is not yet able to capture the dynamics of molecules. As a result, newer methods such as *ab initio* molecular dynamics (AIMD) are employed, but they can only be applied to tens of atoms over periods of under 1 nanosecond (Bylaska *et al.*, 2009, 2017; Seritan, Bannwarth, Bryan S. Fales, *et al.*, 2020). One can also reduce the computational cost by linking the QM calculations with faster approximations via a QM/MM hybrid method to enable more extended simulations while retaining the high-level QM calculations (Senn and Thiel, 2009; Groenhof, 2013).

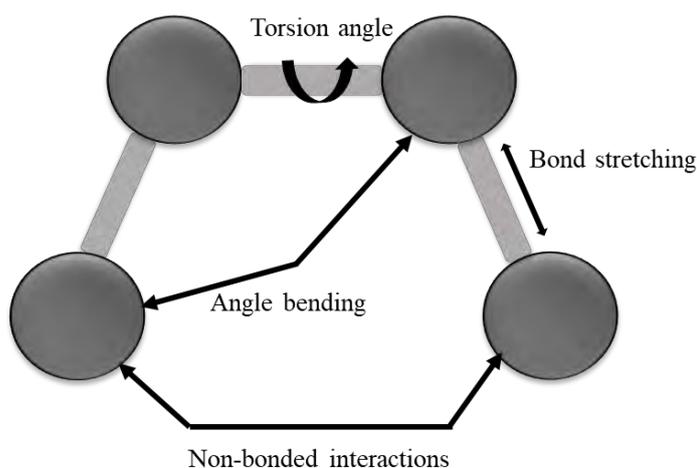
Molecular Mechanics

Molecular mechanics (MM) uses classical mechanics to model molecular systems. In MM, the electrons of a studied system are not explicitly considered; instead, the atomic nucleus and associated electrons are treated as single-particle /perfect spheres (Poltev, 2017). The exclusion of electrons in MM is also based and justified by the Born–Oppenheimer approximation. However, MM does not involve solving the Schrödinger equation; instead, it consists of solving Newton’s second law of motion equation (Equation 3.3), resulting in a mathematical prediction about the position and momentum of the physical system as a function of an external force. It starts by assigning an initial and random velocity to calculate forces acting on the atom in its new position, then calculates the

acceleration the atom will experience due to the force acting on it. MM then let the atoms fly like a Newtonian particle. MM is an alternative to QM when the chemical reactions are not considered; thus, the method may not capture chemical reactions or explain any bond breaking/formation.

$$F_i = m_i a_i \quad \text{Equation 3.3}$$

In MM, atoms are treated as masses/spheres and bonds between them as springs with appropriate force constants. The forces are accounted for by potential energy functions that are equipped with parameters to govern these interactions. The parameters include the bonded terms: bond stretching, angle bending, torsions, and the non-bonded terms: Van der Waals and electrostatic interactions (Figure 3.1). The bond and angle terms follow a harmonic potential and are very close to their equilibrium. In contrast, the dihedral potential can adopt a wide range of values that gives proteins the flexibility to undergo large-scale conformational changes.



$$E_{\text{tot}} = E_{\text{str}} + E_{\text{bend}} + E_{\text{tor}} + E_{\text{vdw}} + E_{\text{elec}}$$

Figure 3.1. The potential energy of a molecule is described as the sum of its bonded and non-bonded terms. The total potential energy (E_{tot}) is decomposed as follows: Energy required to stretch/compress a covalent bond (E_{str}), the energy required to bend a bond from its equilibrium angle (E_{bend}), the energy required to rotate a planar group of atoms from its equilibrium dihedral angle (E_{tor}). Lastly, the energy contribution of interactions between unbound atoms, the latter constitutes Van der Waals interactions (E_{vdw}): the steric exclusion, this potential is attractive at long distance, but becomes repulsive to prevents atoms from overlapping with each other, lastly the electrostatic interactions of charges (E_{elec}).

MM is less computationally expensive than QM and can be used to calculate the energy of large molecular systems like proteins. Unlike QM, MM is not a first principle calculation as it relies on a set of initial empirical parameters. Moreover, MM is not as accurate as QM. Ideally, one would want to move to full QM simulations, but the associated computational cost makes QM simulations untenable with increasing system size. Nevertheless, MD still provides a convenient and best alternative that is cost and time effective and has been used widely in various studies to better understand biomolecules' structural dynamics and conformation changes, especially for the *in-silico* drug discovery.

3.1.3 Force fields

Force fields are a collection of equations and associated parameter sets used to describe the potential energy function (V) that governs interactions between atoms in MM. The force field parameters for a given energy function may be derived from experiments of physics and chemistry such as vibrational bond-spectra or melting points of solvents or quantum mechanical calculations, often in the gas phase. AMBER, GROMOS, OPLS, and CHARMM, are among the widely used force fields biomolecules and differ in the functional and derivation of the parameters (Lopes, Guvench and Mackerell, 2015; Fröhlking *et al.*, 2020). Within function V, both bonded and non-bonded potential terms are defined. A typical MD potential describing bonded and non-bonded interactions in the GROMACS software package is determined by the equation shown below:

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

Equation 3.4

The energy function presented in equation (3.4) describes both bonded and non-bonded potential terms. The bonded terms include bonds, valence angles, dihedral angles where l , l_0 , and k_b , describe the bond distance, equilibrium bond length, and bond stretch force constant. θ , θ_0 , k_a describes the equilibrium bond angle and the angle bend force constant. V_n , n , and γ represent the dihedral angles force constant, periodicity, and phase angle. The last term defines the non-bonded Van der Waals interactions, where r_{ij} is the distance between two non-bonded atoms, ϵ is the energy well, r_{0ij} is the

radius in the Lennard-Jones term that is used to treat the Van der Waals interactions, and last is the non-bonded coulomb interactions where q_i and q_j represent point charges (Singh *et al.*, 2018).

3.1.4 Force field parametrization of metal ions

Most standard force field packages lack parameters of metal ions. Metal sites present challenges for the force field for many reasons, such as the ability of metal ions to accommodate different coordination environments; they thus could have other types of ligands/coordinating residues. Furthermore, the strength of the metal ions' bonds by ligands or coordinating residues can be intermediate between covalent bonds and non-bonded interactions. Nevertheless, several parametrization approaches have been developed and made available to derive standard harmonic force fields that can incorporate metal ions into classical MM (Hu and Ryde, 2011). While providing valuable information, each metal ion parametrization approach presents some shortcomings, and the choice of a parametrization approach can compromise speed and accuracy. The metal ions parametrization approaches can be divided into three broad categories: the non-bonded model approach, the semi-bonded/dummy model approach, and the bonded/covalent model approach (Li and Merz, 2017).

The non-bonded model approach is the simplest among others. It describes the metal ions interactions simply and exclusively by non-bonded potential (electrostatic and Van der Waals potential terms) by using either the formal charge of the metal or charges derived from QM calculations (Stote and Karplus, 1995; Li, Song and Merz, 2015). The non-bonded approach has been used successfully to describe alkali and alkaline-earth ions (Duarte *et al.*, 2014); it is, however, limited by the long-range nature of the non-bonded potentials, in which these potentials are often insufficient to maintain transition metals coordination environment, resulting in the escaping of the metal from its coordination centre or neglecting its correct geometry (Donini and Kollman, 2000; Hu and Ryde, 2011).

The semi-bonded approach describes the metal by a set of cationic dummy atoms placed around it. The dummy atoms are used to mimic covalent interactions and are connected to the metal depending on its predefined coordination geometry (Pang *et al.*, 2000). In the cationic dummy model, the central

atom (the metal) is modelled as a Van der Waals mass with little or no charge. On the other hand, the metal's dummy atoms are modelled as particles carrying partial charges relative to the central metal and no mass, neither Van der Waals potential terms (Li and Merz, 2017). In that way, the total charge of the metal ion is distributed over the cationic dummy atoms and retained over the entire complex. The semi-bonded approach neglects ion-induced dipole interactions, which should not be ignored, especially for highly charged metal ions; it thus fails to reproduce key experimental properties of a series of divalent metal ions. However, Li and Merz developed a new type of Lennard-Jones (LJ) potential by adding an additional r^{-4} term, also called the 12-6-4 LJ-type non-bonded model, to describe the ion-induced dipole interactions (Li and Merz, 2014). The 12-6-4-type LJ potential was successfully used to develop non-bonded models for a series of highly charged metal ions and further extended by Qinghua Liao and colleagues to other biologically relevant ions such as Al^{3+} , Fe^{3+} , and Cr^{3+} (Liao *et al.*, 2017).

Lastly, the bonded approaches describe the metal interactions as covalent bonds and part of a bonded system (Hoops, Anderson and Merz, 1991). In that way, it preserves the metal in its correct coordination geometry. The bonded approach describes the bond, angle, and dihedral terms (although the latter are often ignored for metal sites). The corresponding parameters of the bonded terms can be obtained either experimentally or calculated quantum mechanically (Li and Merz, 2017). The parameters can be derived from QM scans of the potential energy surface of the bonded terms; this method is regarded as accurate but somewhat tedious, especially for more than a few parameters. Another way of deriving the parameters is from a projection of the Hessian matrix (obtained from a QM frequency calculation) into internal coordinates, whereas connectivity is specified (Gresh, Piquemal and Krauss, 2005). This method is more common and more straightforward; however, the internal coordinates are not independent in which different choices of the internal coordinates can produce different parameters. In order to overcome the interdependency problem (the dependence of the force field on the choice of internal coordinates), this method was modified by Seminario to obtain force constants directly from the quantum mechanical Hessian matrix, thereby avoiding any use of internal coordinates (Seminario, 1996). The Seminario method has been fully automated and made

available through the AMBER suite of programs, the MCPB.py program (Li and Merz, 2016), the VFFDT (Zheng *et al.*, 2016), and Hess2FF software (Nilsson *et al.*, 2003). Lastly, the bonded approach can be limited because it describes a predefined bond between the metal and coordinating residues/ligands. Thus, in some cases, it may not allow interconversion between different coordination geometries.

3.2 Methods

3.2.1 Structures retrieval

The GCH1 crystal structure of the *Thermus thermophilus* (*T. thermophilus*) (strain HB8 / ATCC 27634 / DSM 579) (PDB ID: 1WUR) was retrieved from the PDB online database (<https://www.rcsb.org/>) as well as the PTPS crystal structure of the *Plasmodium falciparum* (*P. falciparum*) (PDB ID: 1Y13). The resolution of the retrieved structures is 1.82 Å and 2.2 Å for the GCH1 and PTPS enzymes, respectively. In this study, all residue numbers refer to the selected crystal structures.

3.2.2 Homology modelling of the *P. falciparum* GCH1

Due to the absence of crystal structure for the *P. falciparum* GCH1, a homology model was constructed via homology modelling (HM). The process of HM involves four main steps, template structure identification, target template sequence alignment, model building, and validation. The *P. falciparum* GCH1 amino acids sequence was first retrieved from the PlasmoDB database (<http://plasmodb.org/plasmo/>), with the accession number PF3D7_1224000. The obtained sequence was then used to find an optimal template structure for the homology modelling. The selection of the template structure was based on the sequence identity, coverage and resolution. The search for a suitable template structure was performed using the HHpred web server (<http://toolkit.tuebingen.mpg.de/hhpred>) (Hildebrand *et al.*, 2009). HHpred makes use of Hidden Markov models (HMM-HMM comparison) and incorporates both sequence and structure information into the search. The GCH1 crystal structure of the *T. thermophilus* HB8 (PDB ID: 1WUR) (Tanaka *et al.*, 2005) was then selected as a template structure for the HM process. The *T. thermophilus* GCH1 structure consists of 10 identical chains, a total of 1848 residues.

The *P. falciparum* GCH1 homology model was constructed using MODELLER version 9.16 (Webb and Sali, 2017). The model building was carried out using automodel script with the slow refinement option of MODELLER, and 100 models were generated. The quality of the generated models was then assessed based on their normalised discrete optimized protein energy (z-DOPE) scores (Shen and

Sali, 2006). This score is a statistical measure of a score's relationship to the mean in a group of scores of the DOPE. The DOPE energy profile is normalized over the number of DOPE restraints acting on each residue.

The top three models with the lowest z-DOPE scores were further evaluated using different quality assessment programs, including PROCHECK (Laskowski *et al.*, 1993), ProSA (Wiederstein and Sippl, 2007), and QMEAN (Benkert, Biasini and Schwede, 2011), from which the overall model of the best quality was selected.

3.2.3 Subsets selection

Due to the large nature of the studied proteins, representative subset structures were selected for the QM calculations to substantially reduce the associated computational cost and time with such large protein structures. Subset structures of the active sites containing the Zn^{2+} ion and residues coordinating to it were extracted from each enzyme's X-ray crystal structures. The subsets were selected such that all the metal coordinating residues are present, and the geometry of the active site Zn^{2+} ion was maintained. All free valences in the truncated subsets were capped with hydrogen atoms. Water molecules were included in the subset structures to ensure the completion of the Zn^{2+} ions tetrahedral geometry.

3.2.4 Subsets geometry optimization

The selected subset structures were then subjected to geometry optimization to obtain an ideal low energy starting geometry for the QM scans of the PES. The subsets geometry optimization was carried out using Gaussian 09 (Frisch *et al.*, 2010) with the Becke three-parameter and Lee Yang Parr (B3LYP) hybrid exchange functional (Vosko, Wilk and Nusair, 1980; Becke, 1988; Lee, Yang and Parr, 1988). The B3LYP hybrid exchange functional approximates the electrons exchange-correlation energy by incorporating features from ab initio methods, more specifically the Hartree-Fock methods (experimental) with the DFT approximations. The Los Alamos national laboratory 2 double zeta (LanL2DZ) pseudopotential basis set was used to describe the Zn^{2+} ion, and the 6-31G (d) basis set was for the organic atoms (Ditchfield, Hehre and Pople, 1971; Hehre, Ditchfield and Pople, 1972).

The 6-31G* is a split-valence double-zeta plus polarization basis set, where the diffuse function is described by the (+) and the polarization function on non-hydrogen atoms is described by the (*). The (**) is used if the polarization function is further extended to the hydrogen atoms, such as in the 6-31G** basis set. During the gas phase geometry optimization, no constraints were applied. However, the final optimized subset structures were superimposed onto the initial crystal structure subsets to check if any geometric distortions have occurred during the process. The metal bonded terms from the crystal structure and QM optimized subsets were measured and tabulated.

3.2.5 The restrained electrostatic potential charge evaluation

The partial atomic charges of the active site Zn^{2+} ions were calculated using the restrained electrostatic potential (RESP) charge-fitting approach (Cieplak *et al.*, 1995). The RESP approach involves fitting the molecular electrostatic potential obtained from the QM calculations using the atom-centered point charge model (Bayly *et al.*, 1993; Cornell *et al.*, 1993; Cieplak *et al.*, 1995). The lowest energy/optimized subset structures were submitted for the RESP charges calculation at the DFT/B3LYP level of theory, using the 6-31G* basis set for all non-metal atoms and the LANL2DZ basis with pseudopotential for the Zn^{2+} ions. The resultant partial atomic charges and their corresponding atom were then mapped onto the subset structures for illustration.

3.2.6 Force field parameter determination

The optimized subset structures were then used in the QM PES scans via Gaussian09 (Frisch *et al.*, 2010) at the DFT/B3LYP level of theory on the coordinates defining the bonds stretch, angles bend, and the torsional twist of atoms coordinating the Zn^{2+} ions. In the GCH1 subset, a total of three bonds, seven angles, and three dihedrals were considered for the PES scan. In the PTPS, three bonds, nine angles, and three dihedrals were considered. The features that were scanned are summarized in Table 2. The PES scans were carried out for a total of 10 steps for the bond and angle parameters and 90 steps for torsion parameters, in two directions as forward and reverse at 0.05 Å increment to stretch and contract bonds and 1-degree increment to bend the angles and twist the dihedral. Once the energy

profiles were produced from the PES scan, the least-squares fitting method was used for fitting to terms in the AMBER energy profile (Equation 3.4), and the force field parameters were derived.

3.2.7 Force field parameters evaluation

The derived force field parameters were validated through all-atom MD simulations of the GCH1 and PTPS crystal structures (PDB ID:1WUR) and PTPS (PDB ID: 1Y13), respectively, in addition to the *P. falciparum* GCH1 homology model. The derived parameters were incorporated into the GRONingen MACHine for Chemical Simulations (GROMACS) simulation and analysis computer software package by adding the newly generated parameters to the AMBER03 ffbonded.itp file and manual modification of the topology files to specify the bonded terms related to Zn²⁺ ions. The non-bonded potential such as the electrostatics and Lennard-Jones terms were accounted for using the charges derived from the QM RESP calculations. It is important to note that the fourth coordination position occupied by a water molecule in both systems was not modelled/scanned to allow the water molecule to move freely and permit exchange during the MD simulations. The MD simulations were performed on the GCH1 entire homo-decameric structure, including the 10-equivalent metal-containing active sites and the PTPS entire homo-hexameric structure, including the six-equivalent metal-containing active sites. For each protein, the simulations were carried out for 100 ns of 0.002 ps timestep and periodic boundary conditions in all directions using the GROMACS (Abraham *et al.*, 2015), with AMBER03 force field (Duan *et al.*, 2003). The simulations were carried using a Linux-based cluster and the computational resources available at the Centre for High-Performance Computing (CHPC) in Cape Town, South Africa.

3.2.8 MD simulation protocol

Topology and structure files of the GCH1, PTPS crystal structures, and *P. falciparum* GCH1 homology models were first created. The enzymes' structures were then solvated using the simple point charge (SPC126) water model in a rhombic dodecahedron (RHDO) of the following dimensions 126.33 Å × 126.33 Å × 126.33 Å for GCH1 and 117.61 Å × 117.61 Å × 117.61 Å for PTPS. At least a 1.5 nm distance between any protein atom and the wall was allowed. Each system was naturalized

by using 0.15 M Na^+ and Cl^- counter ions. Then energy minimization was carried out for 5000 of the steepest descent minimizers with a 100 kJ/mol/nm tolerance value, followed by equilibration of temperature and pressure using the modified Berendsen thermostat. The systems were equilibrated for a total of 100,000 steps under the NVT ensemble (canonical/ constant number of particles) at 300 K using the Berendsen temperature coupling algorithm and the NPT ensemble (isothermal-isobaric/ constant number of particles, pressure, and temperature) at 1 atm in all directions and a constant temperature of 300 K using the Parrinello–Rahman barostat algorithm. Bonds on all atoms were constrained using the LINCS algorithm and the particle-mesh Ewald (PME) (a variant of the Ewald sum method) to treat the long-range interactions (Darden, York and Pedersen, 1993). The short-range cut-off was set for 0.8 nm for both coulombic and Van der Waals interactions and a cut-off distance of 14 Å for long-range electrostatic interactions. A final MD production for a total of 100 ns on each system was performed with trajectories generated at a time step of 0.002 ps and timestep, and periodic boundary conditions in all direction frames saved every 10 picoseconds.

3.2.9 Post-MD trajectory analysis

The resultant MD trajectories were analysed concerning the root mean square deviation (RMSD), root mean square fluctuation (RMSF), and radius of gyration (R_g). The distance between the Zn^{2+} ions and coordinating atoms were measured throughout the MD simulations and plotted to check if the coordination environment of the Zn^{2+} ions was maintained and that it did not escape the active site pocket. Furthermore, the metals coordination bond, angle, and dihedral values of the QM optimized and post-MD subset structures were captured and compared to the initial X-ray structure subset structures to evaluate the derived force field parameters. The graphical representations, diagrammatic visualizations, and analyses were generated using the PyMOL Molecular Graphics System (DeLano L, 2002), R (R Core Team, 2019), and GRACE software (Turner, 2005).

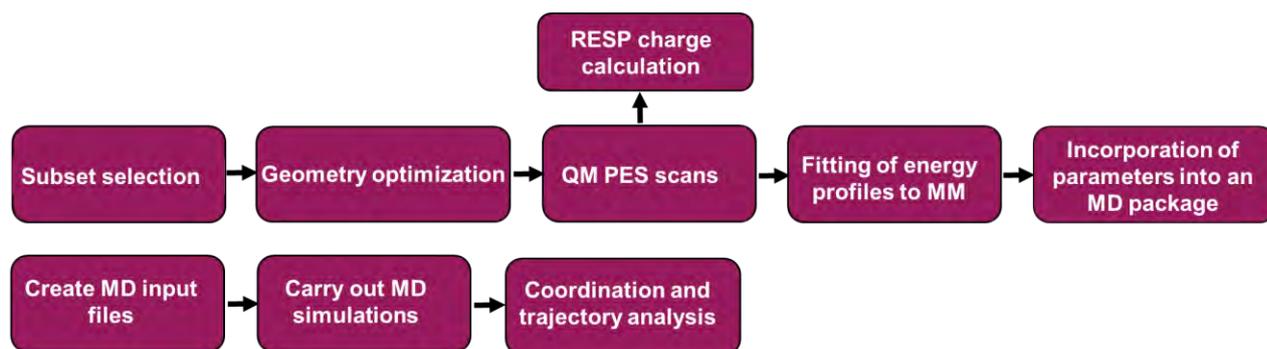


Figure 3.2. A flowchart illustrating the steps followed to derive and evaluate the force field parameters of both GCH1 and PTPS Zn^{+2} ions.

3.3 Results and Discussion

3.3.1 Subsets selection

The QM method is often limited to the study of small molecular systems due to the high computational cost and time. As a result, subset structures of both GCH1 and PTPS active sites were selected. The subsets were obtained from the available experimentally determined structures of the *T. thermophilus* GCH1 (PDB ID: 1WUR) and *P. falciparum* PTPS (PDB ID:1Y13). The QM calculations were carried out on the subset from the crystal structures to ensure the reliability and accuracy of the subsequent QM calculations. The calculated parameters were then applied to the *P. falciparum* GCH1 homology structure to confirm their validity for future use. The selected active site subset structures contained the Zn^{+2} ion and residues accountable for its coordination in each protein. In the GCH1, this included the two Cys and one His residues Cys-SG(108), Cys-SG(179), His-ND1(111), as well as a water molecule. The PTPS active site subset structure consisted of the three coordinating His residues His-NE2(29), His-NE2(41), His-NE2(43), and one water molecule occupying its fourth coordination. The GCH1 and PTPS selected active site subset structures for the QM calculations are shown in Figure 3.3.

The *T. thermophilus* crystal structure (PDB ID: 1WUR) was selected as a template structure to model the *P. falciparum* GCH1. The selected template structure exhibited the highest resolution of 1.82 Å, best quality and sequence identity of 34% to the *P. falciparum* GCH1 (according to the wwPDB validation panels). The template structure information, HHpred search query-template alignment, and the model quality assessment results are provided in Appendix 1.1-Appendix 1.5. The generated homology model was validated and shown to be within the range of reliable experimental native crystal structures and suitable for subsequent simulations.

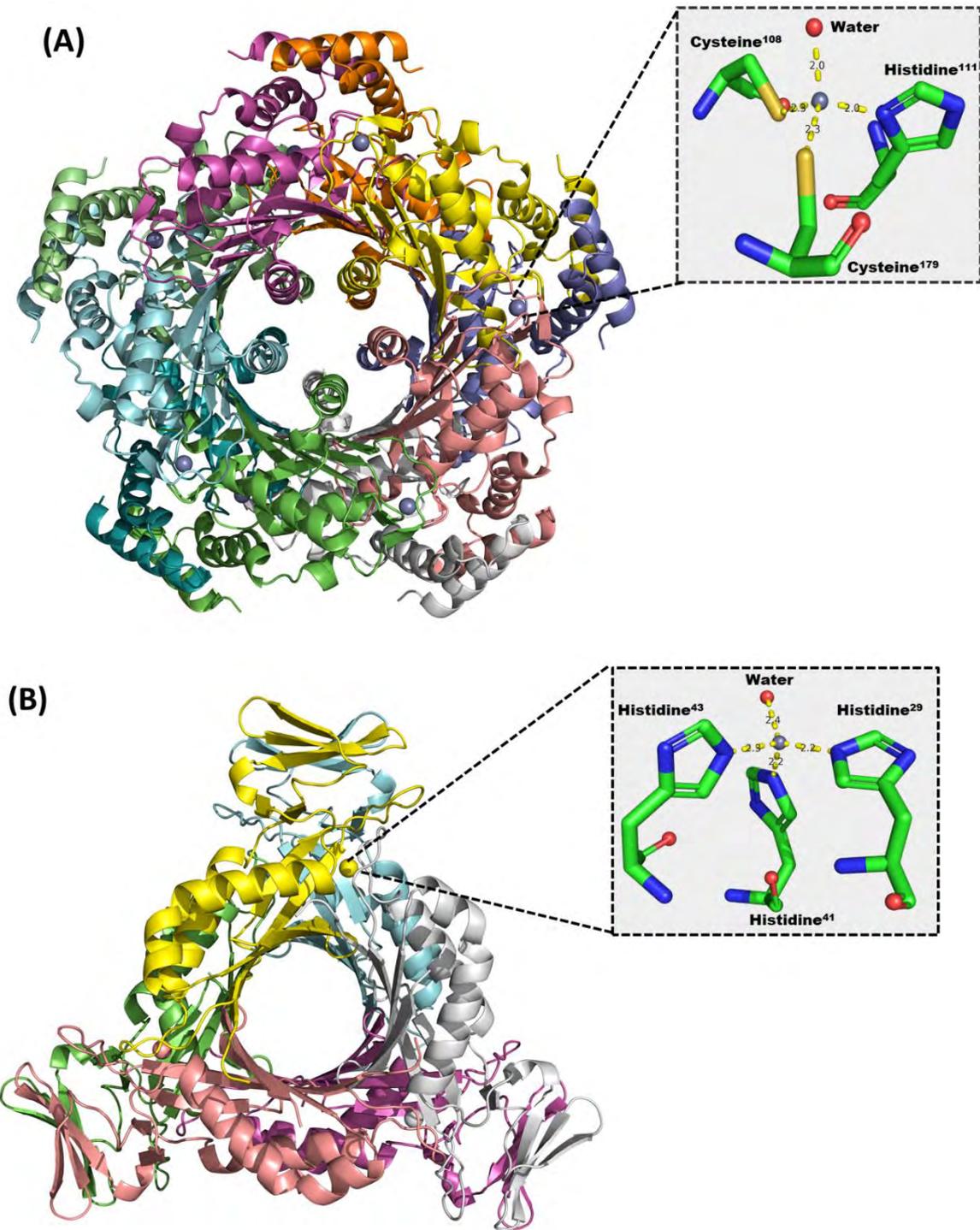


Figure 3.3. A diagram illustrating the overall X-ray crystal structures and the active sites. (A) GCH1 of the *T. thermophilus* (strain HB8 / ATCC 27634 / DSM 579) (PDB ID: 1WUR) (B) PTPS of the *P. falciparum* (PDB ID: 1Y13). The boxed zoom shows the active site metal ions as grey spheres, coordinating residues in the sticks, coordinating water molecules in red spheres, and the bonds in yellow dashes. (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020)).

3.3.2 Subsets geometry optimization

The selected active site subset structures were subjected to geometry optimization at the DFT/B3LYP level of theory to obtain an ideal starting geometry. The hybrid DFT method B3LYP (Vosko, Wilk and Nusair, 1980; Becke, 1988; Lee, Yang and Parr, 1988) was selected for the QM calculations; this method is generally regarded as the most common choice for either optimization and treatment of transition metal-containing systems (Elstner *et al.*, 2003; Peters *et al.*, 2010). The GCH1 and PTPS subset structures were successfully optimized to an ideal geometry, and the initial Zn²⁺ ion tetrahedral coordination environment was shown to be maintained in both metal centres (Table 3.1). The angle values of the QM optimized subsets exhibited some deviations from those of the crystal structures. This observation is attributed to the fact that crystal structures can only provide a static snapshot in addition to the absence of solvent effects during the QM gas-phase optimization. Nevertheless, the attained values indicate that the optimized subset structures did not undergo any notable geometry distortions or alteration during the optimization process. We further expect the values to be more in agreement, at least with the experimental crystal structure data, once the parameters are introduced into a complete system, and all-atom MD simulations in an explicit solvent are carried out.

Table 3.1. GCH1 and PTPS Zn²⁺ ion coordination bond length and angle values from the initial experimental crystal structures, QM optimized, and post-MD structures.

	Bond length (Å)			Angle (°)		
	Zn ²⁺ -His-ND1(111)	Zn ²⁺ -Cys-SG(108)	Zn ²⁺ -Cys-SG(179)	His-ND1(111)-Zn ²⁺ -Cys-SG(108)	His-ND1(111)-Zn ²⁺ -Cys-SG(179)	Cys-SG(108)-Zn ²⁺ -Cys-SG(179)
GCH1						
Crystal structure 1WUR	2.21	2.32	2.30	113.16	107.34	119.29
QM (B3LYP/6-31G*)	2.01	2.24	2.22	129.43	123.81	86.73
MD	1.99±0.0041	2.31±0.0040	2.31±0.0040	91.69±3.98	92.91±4.09	85.87±3.90
PTPS						
Crystal structure 1WUR	Zn ²⁺ -His-NE2(29)	Zn ²⁺ -His-NE2(41)	Zn ²⁺ -His-NE2(43)	His-NE2(41)-Zn ²⁺ -His-NE2(29)	His-NE2(41)-Zn ²⁺ -His-NE2(43)	His-NE2(29)-Zn ²⁺ -His-NE2(43)
Crystal structure 1WUR	2.29	2.13	2.25	91.22	94.43	97.55
QM (B3LYP/6-	2.03	2.03	2.04	115.26	114.14	117.96

31G*)

MD	1.99± 0.0042	2.00± 0.0041	2.00±0.0041	81.04±3.08	86.17±2.93	90.74±3.37
----	-----------------	--------------	-------------	------------	------------	------------

* MD values are reported as average \pm 1 standard deviation.

3.3.3 The restrained electrostatic potential charge evaluation

Metal ions' formal charges can be manipulated based on their coordination environments; hence, metal ions can be cationic, anionic, or neutral (Haas and Franz, 2009). Metal ions' charge is important for MD simulations to obtain an accurate potential electrostatic interaction with surrounding residues or substrates. In this study, the partial atomic charges of the Zn⁺² ions were calculated using the restrained electrostatic potential (RESP) approach at the B3LYP level of theory using the 6-31* basis set for all non-metal atoms and the LANL2DZ basis (with pseudopotential) for the Zn⁺² ions. The RESP approach involves fitting the molecular electrostatic potential obtained from the QM calculations using the atom-centred point charge model (Bayly *et al.*, 1993; Cornell *et al.*, 1993; Cieplak *et al.*, 1995). The Zn²⁺ ions' partial atomic charges were found to be +0.87 and +0.80 for GCH1 and PTPS, respectively. The calculated partial atomic charges and corresponding atom types are shown in Figure 3.4.

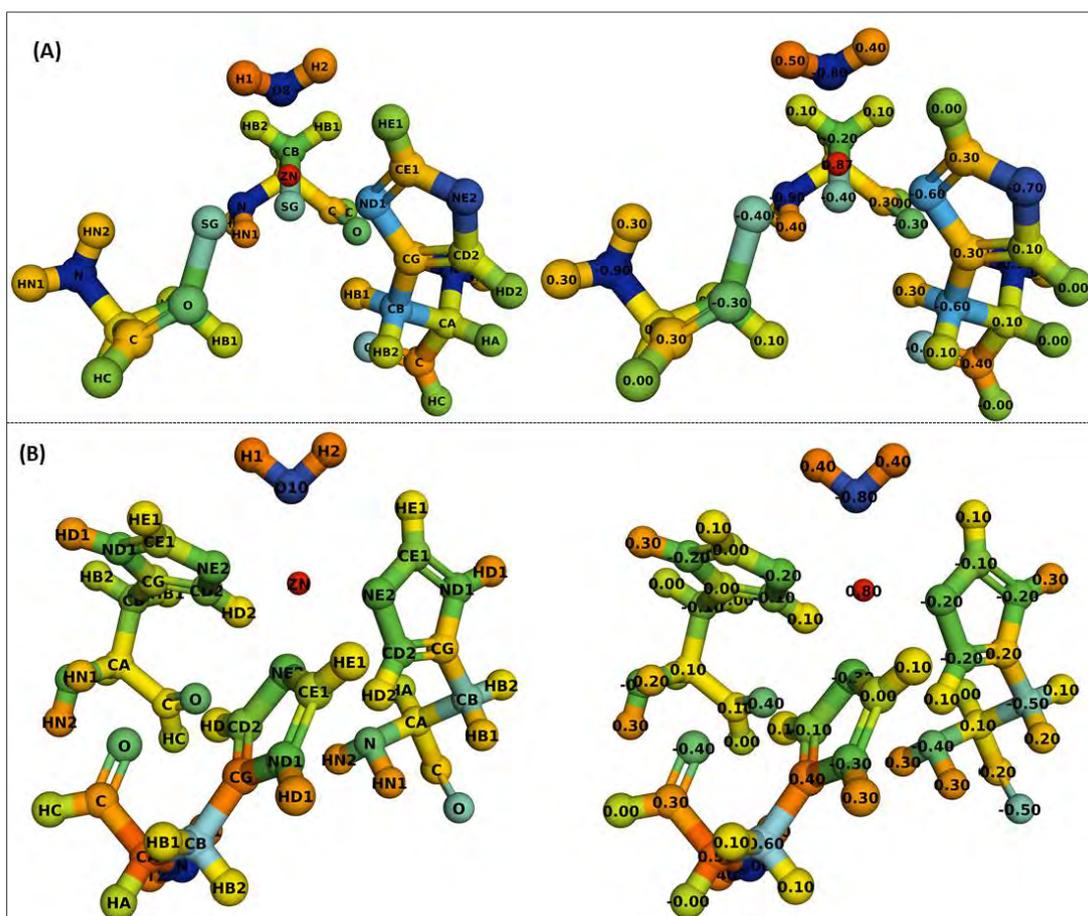


Figure 3.4. (A) atoms of the GCH1 active sites subset (left) and corresponding partial charges (right). (B) atoms of the PTPS active site subset (left) and corresponding partial charges (right). (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020)).

3.3.4 Force field parameter determination

Metal ions can adopt different coordination environments; hence, they can be coordinated to various ligands or residues. It is imperative to use appropriate force field parameters to cater to the metal's coordination environments and coordinating residues. In our systems, the bonded approach was used based on QM PES scans to derive force field parameters that account for the coordination environment of the GCH1 Zn^{+2} ion by two Cys residues and one His through its ND1, and the coordination environment of the PTPS Zn^{+2} ion by three His residues through their NE2 atoms. Force field parameters describing the latter are not available in the Zinc AMBER force field (ZAFF) ZAFF neither the extended ZAFF (EZAFF) (Peters *et al.*, 2010; Burger *et al.*, 2012; Yu, Li and Merz, 2018). A series of PES scans were carried out on coordinates that define the bond stretch, angle bend, and torsional twist of atoms coordinating the Zn^{+2} ions in each protein (Table 3.2). The resultant energy

profiles followed a harmonic profile model (Figure 3.5 and Figure 3.6). The resultant energy profiles were fitted to terms in AMBER potential energy function using the least square method. The fitting procedure is shown in Figure 3.5 and Figure 3.6. From the fitting procedure, the equilibrium bonded terms were derived and presented in Table 3.2; the table summarizes the equilibrium bond length, bond angle, dihedral angle, and associated force constants of the coordination atoms.

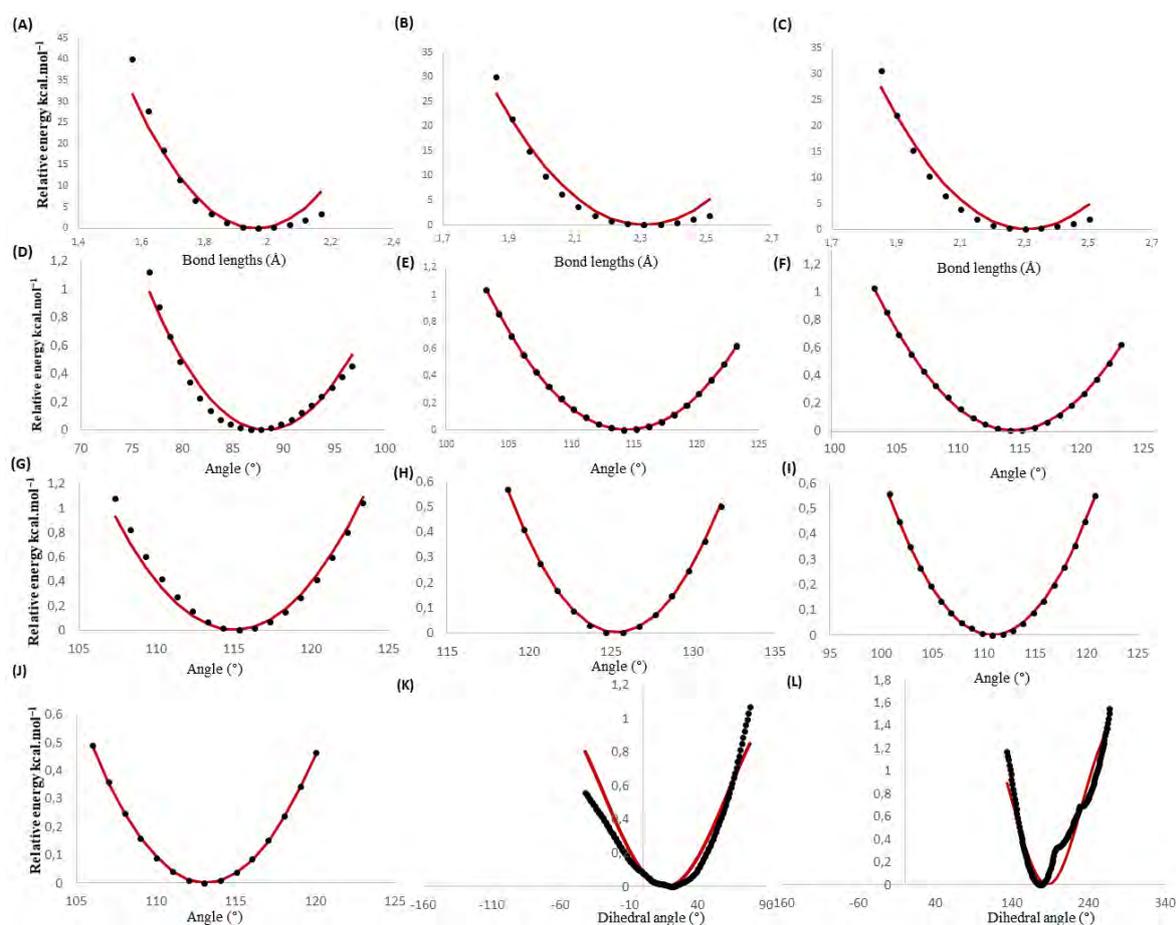


Figure 3.5. Potential energy profiles of the GCH1 Zn^{2+} ion coordinating residues. The energy profiles are shown in black dots, and the theoretical MM fitting curve is shown as a red line. The QM PES scans were performed for 10 steps with an increment of 0.05 Å in the two directions forward and reversed (stretches and contracts) for three bonds A: Zn^{2+} -His-ND1(111), B: Zn^{2+} -Cys-SG(108), and C: Zn^{2+} Cys-SG(179). 10 steps with an increment of 1 degree for three angles, D: Cys-SG(108)- Zn^{2+} -Cys-SG(179), E: Cys-SG(108)- Zn^{2+} -His-ND1(111) and F: Cys-SG(179)- Zn^{2+} -His-ND1(111), G: Zn^{2+} -His-ND1(111)-His-CE1(111), H: Zn^{2+} -His-ND1(111)-His-CG(111), I: Zn^{2+} -Cys-SG(108)-Cys-CB(108), J: Zn^{2+} -Cys-SG(179)-Cys-CB(179)) and 90 steps with an increment of 1 degree for two dihedral angles (K: Cys-SG(108)- Zn^{2+} -His-ND1(111)-His-CE1(111) and L: Cys-SG(179)- Zn^{2+} -His-ND1(111)-His-CG(111)). (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020)).

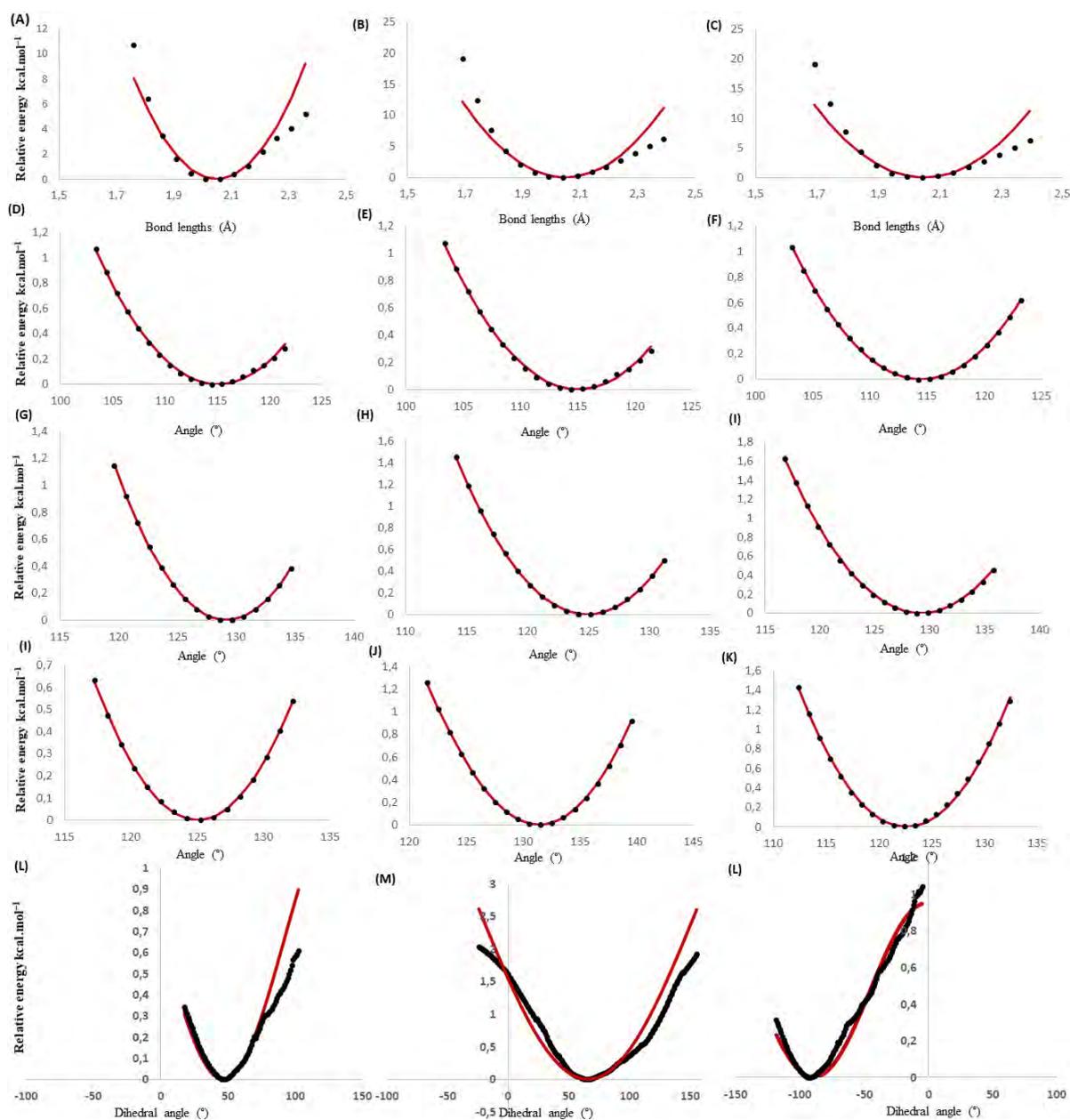


Figure 3.6. Potential energy profiles of the PTPS Zn²⁺ ion coordinating residues. The energy profiles are shown in black dots, and the theoretical MM fitting curve is shown with a red line. The QM PES scans were performed for 10 steps with an increment of 0.05 Å in the two directions forward and reversed for three bonds (A: Zn²⁺ -His-NE₂(43), B: Zn²⁺ -His-NE₂(41), and C: Zn²⁺ -His-NE₂(29)), 10 steps with an increment of 1 degree for three angles (D: His-NE₂(29)- Zn²⁺-His-NE₂(43), E: His-NE₂(29)-Zn²⁺-His-NE₂(41) and F: His-NE₂(41)-Zn²⁺-His-NE₂(43)) and 90 steps with an increment of 1 degree for three dihedral angles (G: His-NE₂(43)- Zn²⁺-His-NE₂(29)-His-CD₂(29), H: His-NE₂(43)-Zn²⁺-His-NE₂(41)-His-CD₂(41) and I: His-NE₂(29)- Zn²⁺-His-NE₂(41)-His-CD₂(41)). (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020)).

From the obtained results, His-ND1(111) showed an equilibrium bond length of 1.99 Å to the Zn²⁺ ion and a force constant value of 203 kcal mol⁻¹ Å⁻², whereas GCH1 Cys-SG(108) and Cys-SG(179) exhibited similar equilibrium bond length values of 2.31 Å to the Zn²⁺ ion and constant force values of 128.74 and 131.73 kcal mol⁻¹ Å⁻² respectively. As expected, this finding indicates that the two Cys residues share similar binding properties to the Zn²⁺ ion (Figure 3.5). The energy profiles of the PTPS Zn²⁺ ion coordinating His residues exhibited similar energy profiles consequently same force constants (Table 3.2), suggesting that the three residues have similar binding properties to the Zn²⁺ ion (Figure 3.6).

Table 3.2. The derived force field parameters of both GCH1 and PTPS active site Zn²⁺ ions.

GCH1			
Bonds	<i>K_r</i> (kcal mol ⁻¹ Å ⁻²)	<i>r_{eq}</i> (Å)	
Zn ²⁺ -Cys-SG(108)	128.74	2.31	
Zn ²⁺ -His-ND1(111)	203.25	1.99	
Zn ²⁺ -Cys-SG(179)	131.73	2.31	
Angles	<i>K_θ</i> (kcal mol ⁻¹ rad ⁻²)	<i>θ_{eq}</i> (degrees)	
Cys-SG(108)-Zn ²⁺ -His-ND1(111)	26.70	114.46	
Cys-SG(108)-Zn ²⁺ -Cys-SG(179)	24.33	88.23	
Cys-SG(179)-Zn ²⁺ -His-ND1(111)	26.70	114.46	
Zn ²⁺ -Cys-SG(108)-Cys-CB(108)	28.35	110.81	
Zn ²⁺ -Cys-SG(179)-Cys-CB(179)	31.63	113.12	
Zn ²⁺ -His-ND1(111)-His-CE1(111)	54.14	115.37	
Zn ²⁺ -His-ND1(111)-His-CG(111)	54.14	115.37	
Dihedral	<i>V_n</i> (kcal mol ⁻¹)	<i>n</i>	<i>γ</i>
Cys-SG(108)-Zn ²⁺ -His-ND1(111)-His-CE1(111)	1.10	2.00	9.10
Cys-SG(179)-Zn ²⁺ -His-ND1(111)-His-CG(111)	1.40	2.00	15.92
PTPS			
Bonds	<i>K_r</i> (kcal mol ⁻¹ Å ⁻²)	<i>r_{eq}</i> (Å)	
Zn ²⁺ -His-NE2(43)	95.86	2.01	
Zn ²⁺ -His-NE2(29)	95.86	2.01	
Zn ²⁺ -His-NE2(43)	95.86	2.09	
Angles	<i>K_θ</i> (kcal mol ⁻¹ rad ⁻²)	<i>θ_{eq}</i> (degrees)	
His-NE2(29)-Zn ²⁺ -His-NE2(43)	25.53	115.04	
His-NE2(29)-Zn ²⁺ -His-NE2(41)	25.53	115.04	
His-NE2(41)-Zn ²⁺ -His-NE2(43)	26.69	114.46	
Zn ²⁺ -His-NE2(29)-His-CD2(29)	41.87	129.13	
Zn ²⁺ -His-NE2(29)-His-CE1(29)	41.11	124.88	
Zn ²⁺ -His-NE2(29)-His-CD2(41)	34.65	129.19	
Zn ²⁺ -His-NE2(29)-His-CE1(41)	34.01	125.02	
Zn ²⁺ -His-NE2(43)-His-CD2(43)	44.26	131.29	

Zn²⁺-His-NE₂(43)-His-CE₁(43)	44.85	122.56	
Dihedral	<i>V_n</i> (kcal mol⁻¹)	<i>n</i>	<i>γ</i>
His-NE₂(43)-Zn²⁺-His-NE₂(29)-His-CD₂(29)	1.31	2.00	-7.80
His-NE₂(43)-Zn²⁺-His-NE₂(41)-His-CD₂(41)	0.95	2.00	-6.32
His-NE₂(29)-Zn²⁺-His-NE₂(41)-His-CD₂(41)	5.22	1.00	-1.99

3.3.5 Force field parameters evaluation

In order to evaluate the newly generated force field parameters, the derived parameters were incorporated into the GROMACS MD simulation package. Separate MD simulation runs of the complete GCH1 homo-decameric and PTPS homo-hexameric structures were carried out for a total of 100 ns in triplicates. The resulting MD trajectories were analysed through the RMSD, RMSF, and Rg calculations. The 100 ns MD simulations were sufficient to infer the stability of the proteins and the metal ions coordination environments. From the RMSD results, both proteins were shown to sustain a steady deviation (below 2 Å) from their initial structures, which were further maintained throughout the simulation course (stabilizing at an average of 1.5 Å) (Figure 3.7-A and Figure 3.8-A). The Rg results showed that both proteins remained compact throughout the MD simulations (Figure 3.7-C and Figure 3.8-C).

In the RMSF profiles, the active site residues of both GCH1 and PTPS showed minimum fluctuation, indicating their stability and steady coordination to the metal ions (Figure 3.7-B and Figure 3.8-B). Furthermore, the RMSF values were compared to the experimental *B*-factor values and were shown to agree, in which the proteins terminal regions exhibited the highest fluctuation, and the protein core maintained a relatively steady fluctuation profile (Appendix 1.6).

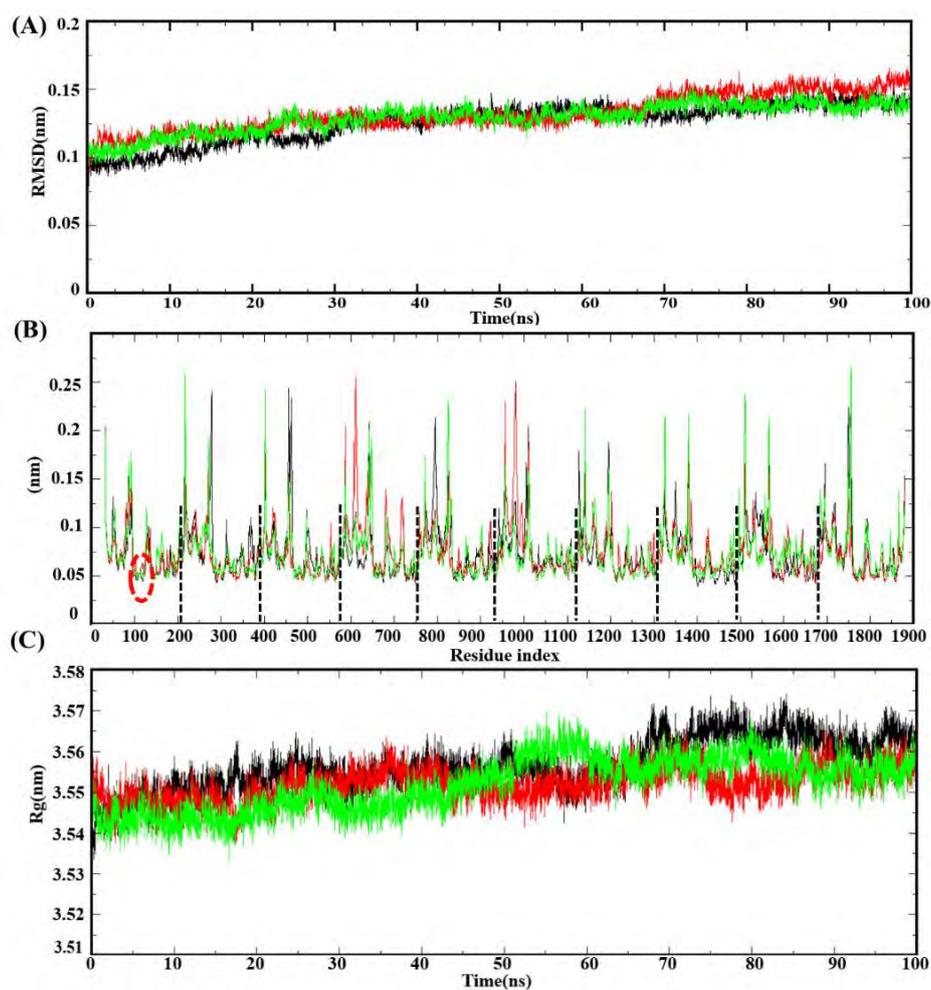


Figure 3.7. Assessment of the GCH1 stability by (A) RMSD, (B) RMSF, and (C) Rg. The black dashed lines separate each chain of the protein. The dashed red circles highlight the location of the active site residues responsible for the metal's coordination. (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020)).

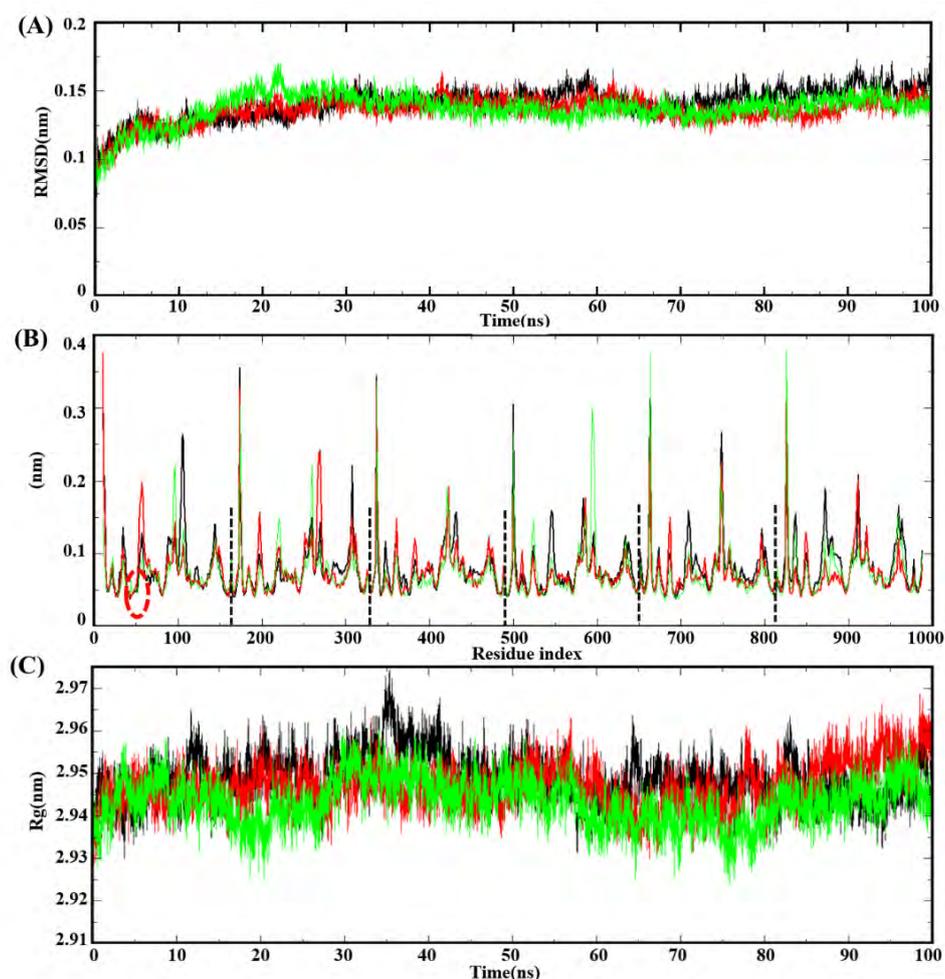


Figure 3.8. Assessment of the PTPS stability by (A) RMSD, (B) RMSF, and (C) Rg. The black dashed lines separate each chain of the protein. The dashed red circles highlight the location of the active site residues responsible for metal coordination. (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020)).

3.3.6 Metal ions RMSD

In order to investigate the overall changes of the metal sites during the 100 ns MD simulations, the metal ions RMSD was monitored. The RMSD values of the metal sites (below 2 Å) confirmed the steadiness of the metal ions, as shown in Figure 3.9, which further indicates that both metals remained stable at their positions and did not escape the active site pocket. The obtained result confirms that the newly generated force field parameters adequately described the metal centres and preserved their coordination environments.

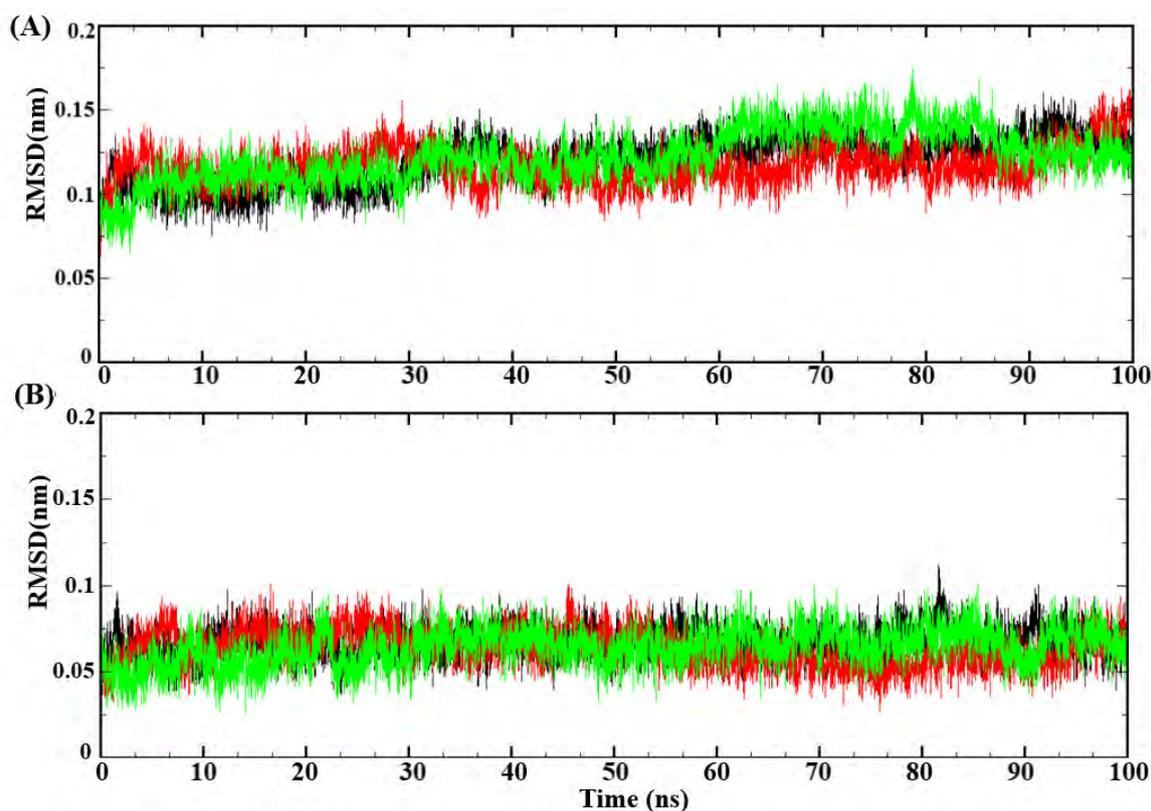


Figure 3.9. RMSD values of the Zn^{2+} ions of (A) GCH1 (PDB ID: 1WUR) and (B) PTPS (PDB ID: 1Y13) over the 100 ns MD simulations. (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020)).

The bonded terms of the GCH1 and PTPS Zn^{2+} ions were monitored during the 100 ns MD simulations and shown to be steady throughout (Figure 3.10). The average coordination bond length throughout the MD simulations for both GCH1 and PTPS Zn^{2+} ion was calculated and presented in Table 3.1. The calculated average coordination bond length was reported as ± 1 standard deviation. The obtained MD values showed better agreement with the experimental crystallography data.

A histogram showing the distribution and occurrence of the metal ions coordination bond length over the 100 ns MD simulations are presented in Figure 3.11. From the histogram data, the metal coordination distances were shown to be normally distributed, peaking around the calculated averages. The result further indicates that the coordination bond lengths near the mean values are more frequent in occurrence. The coordination distances were stable and maintained throughout the simulations; this reflects that the metal coordination geometry was governed by these steady coordination distances.

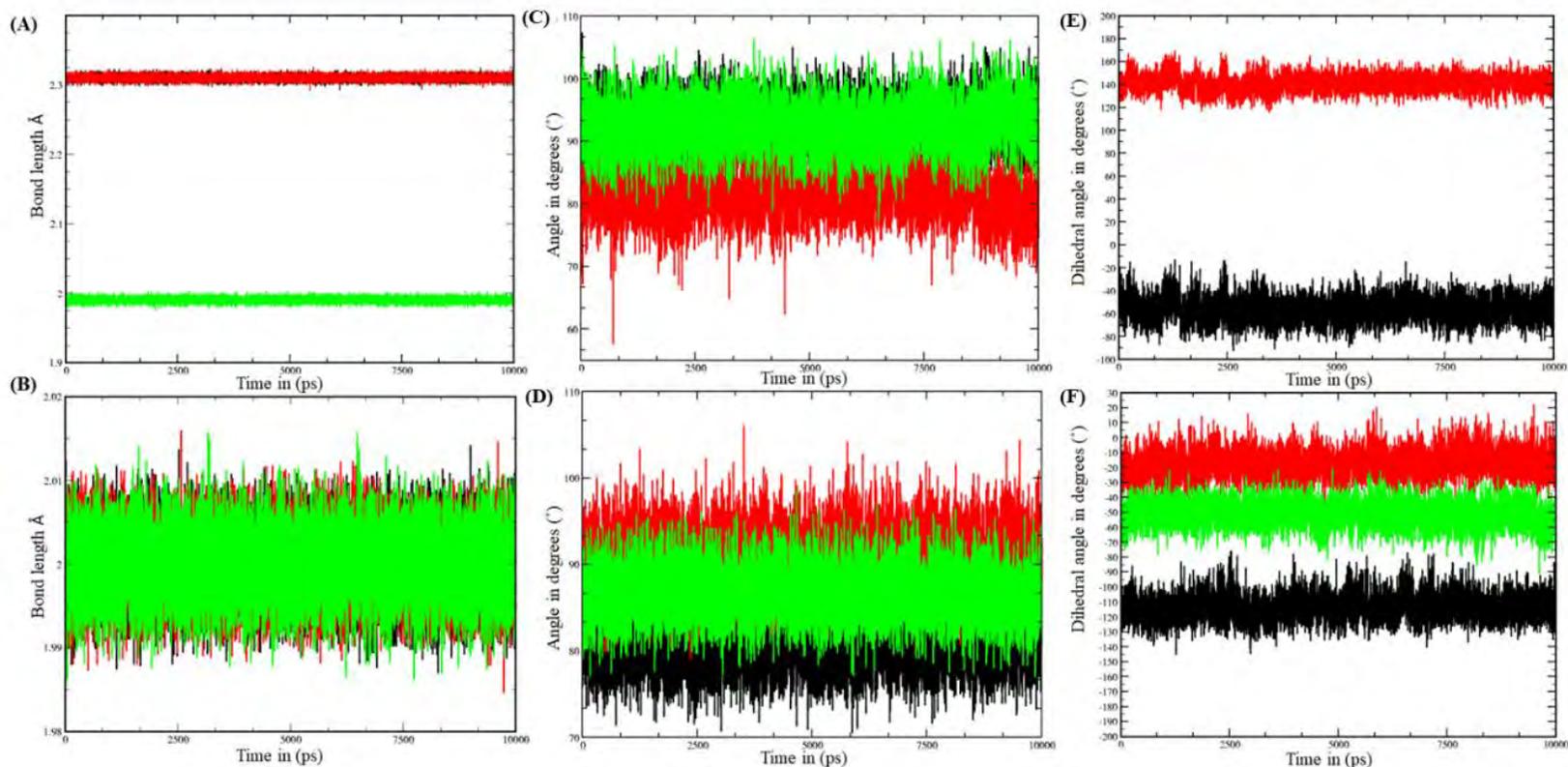


Figure 3.10. (A) GCH1 Zn^{2+} ion coordination distances during the 100 ns MD simulations. The black line represents the coordination distance of Zn^{2+} -His-ND1(111), the red line represents the coordination distance of Zn^{2+} -Cys-SG(108), and the green line represents the coordination distance of Zn^{2+} -Cys-SG(179). (B) PTPS Zn^{2+} ion coordination distances. The black line represents the coordination distance of Zn^{2+} -His-NE2(29), the red line represents the coordination distance of Zn^{2+} -His-NE2(41), and the green line represents the bond distance of Zn^{2+} -His-NE2(43). (C) GCH1 Zn^{2+} ion coordination angles during the 100 ns MD simulations. The black line represents the coordination angle of Cys-SG(108)- Zn^{2+} -His-ND1(111); the green line represents the coordination angle of Cys-SG(197)- Zn^{2+} -His-ND1(111), and the red line represents the coordination angle of Cys-SG(108)- Zn^{2+} -Cys-SG(179). (D) PTPS Zn^{2+} ion coordination angles. The red line represents the coordination angle of His-NE2(41)- Zn^{2+} -His-NE2(29), the green line represents the coordination angle of His-NE2(41)- Zn^{2+} -His-NE2(43),

and the black line represents the coordination angle of His-NE2(29)-Zn²⁺-His-NE2(43). (E) GCH1 Zn²⁺ ion coordination dihedral angles during the 100 ns MD simulations. The red line represents the dihedral angle of Cys-SG(108)-Zn²⁺-His-ND1(111)-His-CE1(111); the black line represents the dihedral angle of Cys-SG(108)-Zn²⁺-His-ND1(111)-His-CG(111). (F) PTPS Zn²⁺ ion coordination angles. The red line represents the dihedral angle of His-NE2(43)-Zn²⁺-His-NE2(41)-His-CD2(41), the green line represents the coordination angle of His-NE2(43)-Zn²⁺-His-NE2(29)-His-CD2(29), and the black line represents the dihedral angle of His-NE2(29)-Zn²⁺-His-NE2(41)-His-CD2(41). (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020)).

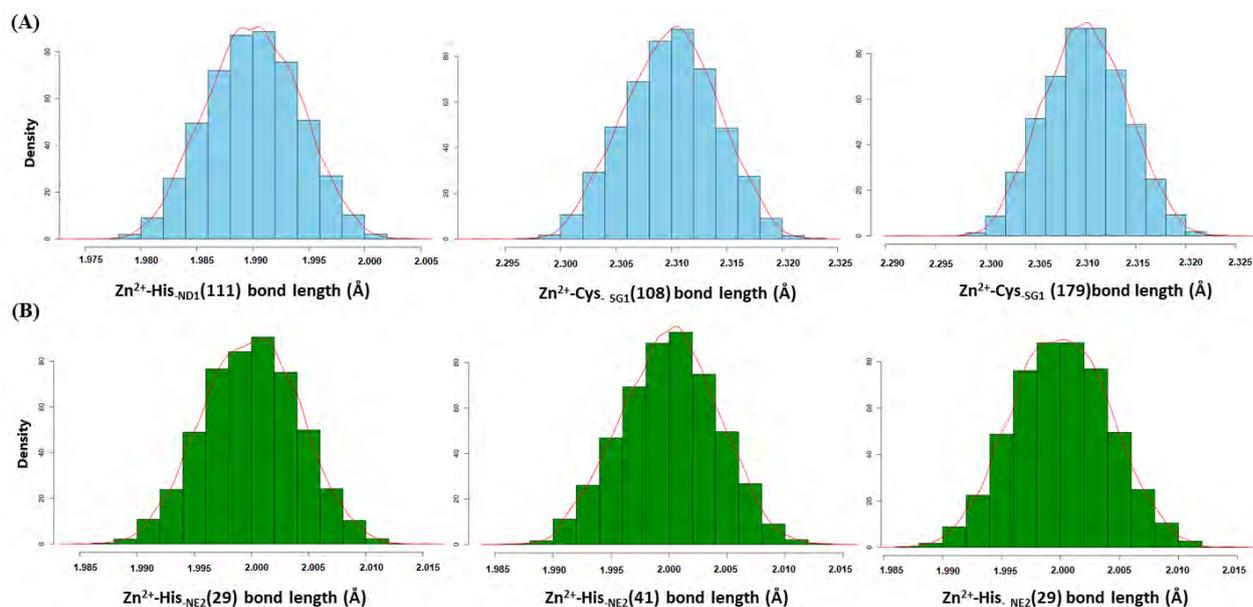


Figure 3.11. Density histogram illustrating the distribution and occurrence of the Zn^{2+} ion coordination distances during the 100 ns MD simulations of (A) GCH1 and (B) PTPS. The red line is an overlaying density line drawn to show the kernel density of the histograms. (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020)).

3.3.7 GCH1 and PTPS reaction mechanisms

As mentioned earlier, the fourth coordination of the GCH1 and PTPS metal ions is occupied by a water molecule. In GCH1, the water molecule has been described to play a key role during catalysis, in which it acts as the nucleophile that initiates the hydrolysis of the substrate ring, affording zinc bound 8R hydrate of the GTP substrate (Rebelo *et al.*, 2003; Tanaka *et al.*, 2005; Gräwert, Fischer and Bacher, 2013). During MD simulations, a water molecule was found to arrive at proximity to the Zn^{2+} ions of both GCH1 and PTPS metal centres, thereby completing the Zn^{2+} ions tetrahedral coordination environment (Figure 3.12-A, Figure 3.12-B). The water molecules maintained a mean coordination distance of 3.2 Å and 2.2 Å to the GCH1 and PTPS active site Zn^{2+} ions, respectively. During the MD simulations of the GCH1 enzyme, the imidazole ring of His_{ND1}(110) showed a distinct rotation and arrived at proximity to the Zn^{2+} ion. The His_{ND1}(110) coordination bond distance was also monitored over the 100 ns MD simulations presented in Appendix 1.7. His_{ND1}(110) is known to interact with water molecules by forming hydrogen bonds; this mechanism was reported in previous studies to

initiate the hydroxyl nucleophile attack for the substrate ring opening during catalysis (Tanaka *et al.*, 2005). H177 is another essential catalytic residue of GCH1 known to be involved in the hydration of the substrate formyl group via an activated water molecule (Tanaka *et al.*, 2005). The His-ND1(177) coordination bond distance was monitored over the 100 ns MD simulations presented in Appendix 2.7.

In the PTPS active sites, two key residues Glu38 and Glu161, are located close to the Zn²⁺ ion at a distance of 3.5 and 4.5 Å, respectively. These residues were described not to coordinate the Zn²⁺ ion but essential for the catalysis substrate recognition (Bürgisser *et al.*, 1995). The two residues were described as located far from the active site and become closer in the presence of the substrate to initiate the catalysis. During the MD simulations, Glu-OE1(38) and Glu-OE1(161) maintained an average distance of 5.05 ± 0.1472 Å and 3.00 ± 0.0488 Å from the metal centre (Appendix 1.8). The observed distance was shown to agree with previously reported experimental data (Bürgisser *et al.*, 1995; Ploom *et al.*, 1999).

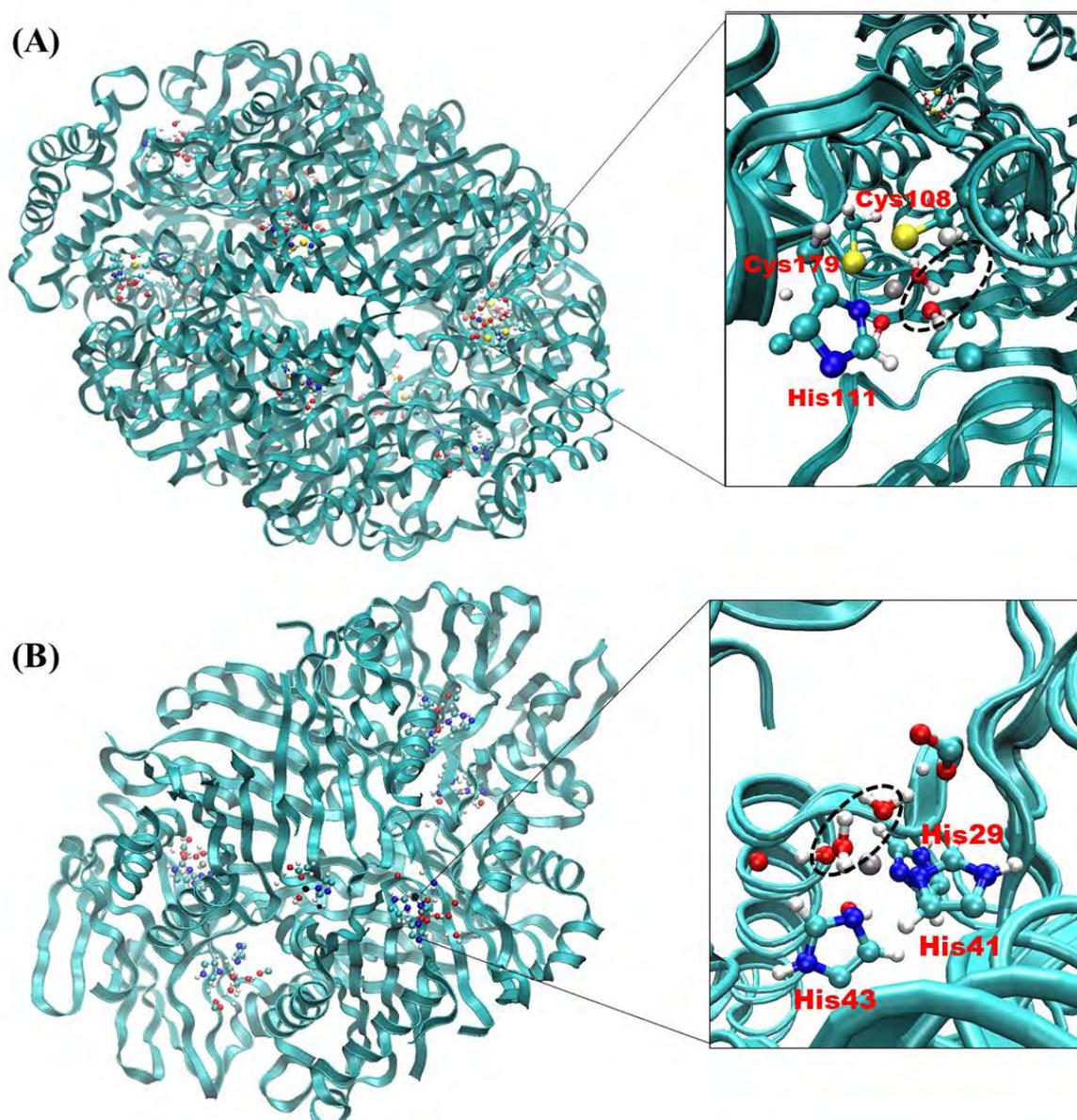


Figure 3.12. MD snapshots of the last frame from both active sites. The black dashed circles show the water molecules arriving at the active sites and completing the tetrahedral geometry of the (A) GCH1 and (B) PTPS Zn²⁺ ions. (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020)).

3.3.8 Validation of the force field parameters in the presence of an inhibitor compound

MD simulations were carried out for a total of 100 ns of an Apo protein and ligand-bound complex, followed by post-MD trajectory analysis. This step was necessary to infer the accuracy and reliability of the newly generated force field parameters in elucidating crucial protein/metal-ligand interaction. Post-MD trajectory analysis, including the RMSD, RMSF, and radius of gyration, showed the effect of the potential inhibitor compound on the protein dynamics (Figure 3.13). Furthermore, the derived

force field parameters were able to disclose critical interactions between the protein and ligands. The metal ion was shown to interact and stabilize the inhibitor compound by interacting with its hydroxyl group. The metal was also found to abstract a water molecule and produce a free hydroxyl group, resulting in the attack of the substrate ring (Figure 3.14). Overall, the derived parameters could maintain the metal ion in the active site and further describe its interactions with a ligand compound.

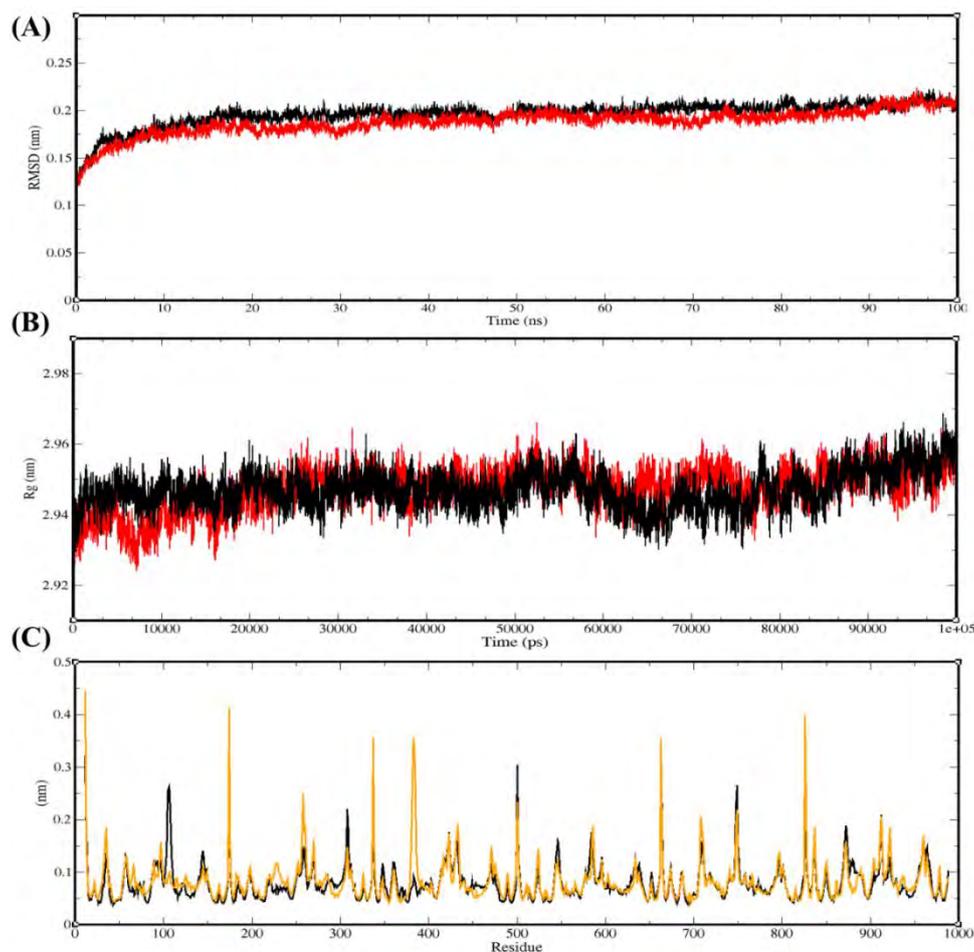


Figure 3.13. Dynamic properties of both the apo form (Black) and ligand-bound complexes over a simulation period of 100 ns. (A) The global stability of the different systems as determined by RMSD (B) The compactness of the apo system (Black) and ligand-bound complexes by Rg. (C) local fluctuation profile of the individual residues by RMSF. The Apo system is shown in black. (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020)).

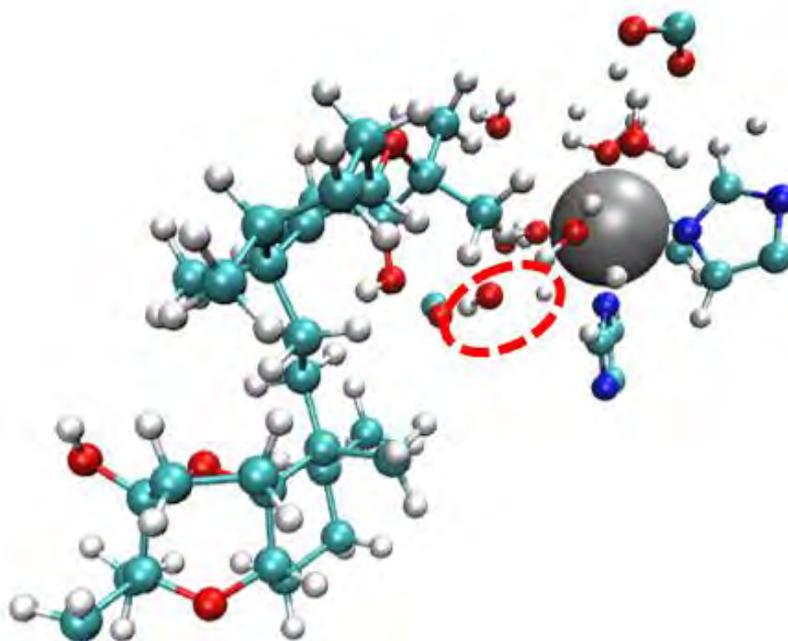


Figure 3.14. Snapshot of the last MD frame showing a docked ligand interaction with the Zn^{2+} . The red circle highlights the ion-induced perturbation of the water hydrogen bonds resulting in a free hydroxyl that we believe is responsible for attacking the substrate C atoms. The Zn^{2+} interaction with the ligand hydroxyl group was the main reason for stabilizing the substrate throughout the MD simulations. (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020)).

3.3.9 Validation of the force field parameters on the *P. falciparum* GCH1 homology model structure

The derived force field parameters were further tested on the malaria parasite *P. falciparum* GCH1 homology model. This was important to ensure their validity for use in future *in-silico* drug discovery studies. The Zn^{2+} ion coordinating residues of the *P. falciparum* GCH1 homology model corresponded to Cys-SG(277), Cys-SG(348), and His-ND1(280). The MD simulations of the *P. falciparum* GCH1 were carried out for 100 ns in triplicates. The *P. falciparum* GCH1 RMSD showed a sustained deviation from the initial protein structure (Appendix 1.9-A). The RMSF values showed the high fluctuation of the protein N-terminal domain and stability of the active site residues by exhibiting minimum fluctuation, indicating their steady/ongoing metal ion coordination (Appendix 1.9-B). The Rg was shown to be steady, demonstrating the compactness of the protein structure throughout the MD simulations (Appendix 1.9-C).

The coordination bond length of the *P. falciparum* Zn²⁺ ion was measured throughout the MD simulation and shown as steady (Appendix 1.10). Angles and dihedral angles were also monitored throughout the MD simulations and were also steady and maintained during simulations (Appendix 1.11 and Appendix 1.12). His-ND1(280) maintained an average coordination bond length of 1.99 Å ± 0.0041, whereas Cys-SG(277) and Cys-SG(348) maintained an average of 2.31 Å ± 0.0042. The distribution and occurrence of the coordination bond lengths were plotted as histograms and shown to follow a normal distribution, indicating that the coordination bond lengths near the mean values are more frequent in occurrence than bond lengths values far from the mean (Appendix 1.13). Lastly, the metal ion RMSD over the 100 ns MD simulations confirmed the steadiness of the metal ion, as shown in Appendix 1.14.

Lastly, given that the metal ions coordinating residues are highly conserved among different species (Nar, 2011; Gräwert, Fischer and Bacher, 2013; Kümpornsin, Kotanan, *et al.*, 2014); the derived force field parameters can be applied to GCH1 and PTPS enzymes from different organisms. The above results support this statement as the newly derived force field parameters precisely described the metal coordination environment of the *P. falciparum* GCH1 homology structure.

3.4 Conclusions

Computer-aided drug discovery studies of metalloproteins can be hampered by the lack of force field parameters that accurately describe and maintain the metal centres during MD simulations. Both GCH1 and PTPS contain catalytically essential metal ions; failure to employ appropriate force field parameters to describe the metals interactions accurately can affect the modelling of the active site and its dynamics. In this chapter, force field parameters describing both GCH1 and PTPS active sites Zn^{2+} ions were derived from QM scans of the PES of the metals bonded terms. The QM calculations were performed at the DFT/B3LYP level of theory using the LanL2DZ basis set for the Zn^{2+} ion and the 6-31G* for all atoms. The PES energy values were fitted via the least square method to terms in AMBER potential then incorporated into the GROMACS MD simulation package for evaluation. Over the 100 ns MD simulations in triplicates, the generated force field parameters were shown to be adequate in describing and maintaining both the GCH1 and PTPS Zn^{2+} ions coordination environment. We regard the newly developed parameters as reliable and would result in the production of an accurate representation of both GCH1 and PTPS metal centres. The reported force field parameters will be of value for MD simulations in future *in-silico* drug discovery studies involving the two enzymes as well as other enzymes with the same Zn^{2+} binding motifs and coordination environments.

Chapter 4

Normal Mode Analysis of the Tunnelling Fold enzyme GTP Cyclohydrolase I Using Elastic Network Models

This chapter describes the normal mode analysis, calculated on the anisotropic network model of the guanosine-5'-triphosphate (GTP) cyclohydrolase I (GCH1) enzyme to elucidate the enzyme large-scale conformational changes that can modulate its function and further identify potential allosteric drug targeting sites.

Chapter 4 is reproduced in part with permission from the following publications:

GTP Cyclohydrolase I as a Potential Drug Target: New Insights into Its Allosteric Modulation via Normal Mode Analysis. [Afrak Khairallah](#), Caroline J. Ross, and Özlem Tastan Bishop. *Journal of Chemical Information and Modeling* 61 (9), 4701-4719. DOI: 10.1021/acs.jcim.1c00898.

Authors Contributions: Ö.T.B. conceived the project. A.K. performed the calculations and analysis of data under the guidance of C.R. and Ö.T.B.

Chapter overview

This chapter involves normal mode analysis (NMA) based on the elastic network model (ENM) of the guanosine-5'-triphosphate (GTP) cyclohydrolase I (GCH1) enzyme to elucidate the structural dynamics that govern its large-scale collective motions as well as identifying regions responsible for modulating conformational changes. This chapter also invites attention to studying the structural dynamics of the mammalian GCH1 enzyme and its regulatory protein, the GFRP. Three sets of NMA calculations were carried out; first, NMA of the *Plasmodium falciparum* GCH1 homology model, followed by NMA of the *Rattus norvegicus* GCH1 structure alone and in complex with its regulatory partner in both inhibitory and stimulatory forms. Several low-frequency non-degenerate modes of functional relevance were characterized, which disclosed key sites exerting allosteric control over the GCH1 global dynamics. Kinetically hot residues, which are crucial for the stability of the protein, were also identified within the high-frequency normal modes. The obtained results can provide a solid starting point to design novel-antimalarial against the malaria parasite *Plasmodium falciparum* GCH1 enzyme and anti-neuropathic treatment for humans.

4.1 Introduction

4.1.1 Protein conformational dynamics and function

Proteins structures are not static; they undergo large-scale domain movement and rearrangement induced by external events such as substrate binding/recognition and physical association of partner proteins (Bahar, Chennubhotla and Tobi, 2007; Henzler-Wildman and Kern, 2007; Teilum, Olsen and Kragelund, 2009; Ma and Nussinov, 2010; Wako and Endo, 2011; Fenwick, Esteban-Martín and Salvatella, 2011; Nussinov and Ma, 2012; Orozco, 2014; Han *et al.*, 2018; Petrovic *et al.*, 2018; Di Cera, 2020; Zhang *et al.*, 2020; Loutchko and Flechsig, 2020). Also, the ability of proteins to carry their biological function is highly dependent on their ability to adopt certain conformations. Thus, it has long been established that the structural dynamics of proteins are linked to function (Eisenmesser *et al.*, 2005; Agarwal, 2006; Henzler-Wildman and Kern, 2007; Teilum, Olsen and Kragelund, 2009; Orozco, 2014). Studying the structural dynamics of protein is therefore imperative and can shed light on many fundamental biological functions.

In a study by Daniels and colleagues, the kinetics of ligand-induced conformational changes was monitored for an RNase P protein with six different ligands of phosphate derivatives. The study reported that the ligands showed different binding affinities depending on the protein conformational state, highlighting the importance of studying the other conformational states and their significance for function (Daniels, Suo and Oas, 2015). Furthermore, the role of protein structures should not be limited to the function only but rather to the evolution of function. In that way, the protein structure-function paradigm, describing the close relationship between the structure and function, can be extended to the protein structure-function–dynamics paradigm, which can allow the understanding of function more comprehensively (Srivastava *et al.*, 2018).

4.1.2 Protein conformational dynamics and allostery

Many protein functions are regulated through allostery, in which the binding of an effector molecule at a distal site induces changes in the functional activity (Gunasekaran, Ma and Nussinov, 2004; Nussinov and Tsai, 2013; Amamuddy *et al.*, 2020). Allosteric drugs can be particular to sites other

than the active sites; hence, they can induce desired effects of activating or inhibiting a protein function via a mechanism that does not target the highly conserved active sites. For such reason, they are regarded as safe and far less toxic to the host. Allosteric effectors can mediate key changes of proteins' thermodynamic and dynamic properties, inducing different conformational states (Guo and Zhou, 2016). A detailed assessment of the potential roles of protein conformational dynamics in allosteric regulation is provided by Guo and Zho (Guo and Zhou, 2016). Among several findings of their study, they deduced that understanding the conformational dynamics can shed light and provide a better understanding of proteins' allosteric regulation. In other words, studying the structural dynamics of proteins can disclose key sites responsible for modulating/driving the conformational transition and thus pinpoint potential allosteric sites.

Discovering allosteric sites with a central effect on a protein conformation can be challenging (Amamuddy *et al.*, 2020). As in many cases, the location of allosteric sites is often unknown, and the effects on the intrinsic motion of the protein are difficult to determine experimentally (Suplatov and Švedas, 2015). As a result, several approaches may be used to study protein structural dynamics and different conformational states, which allows for identifying allosteric sites within proteins (Lu, Shen and Zhang, 2019; Amamuddy *et al.*, 2020).

4.1.3 Approaches for studying proteins conformational dynamics

Studying the structural dynamics of proteins depends significantly on our ability to model/express large-scale conformational changes. The characterization of large-scale conformational changes can pinpoint key sites that modulate the functional mechanisms (Guarnera and Berezovsky, 2020; D. Penkler *et al.*, 2017; D. L. Penkler *et al.*, 2018; Shrivastava *et al.*, 2020). Different approaches were developed to probe the structural dynamics of proteins over different timescales from nanoseconds to milliseconds, which can, in turn, provide the means to extract different conformations of biological/functional relevance. The approaches used for studying the structural dynamics of protein are broadly divided into experimental and computational. The experimental approaches include X-ray, cryo-electron microscopy (Shoemaker and Ando, 2018), nuclear magnetic resonance (NMR) (Becker *et al.*, 2018; Chan-Yao-Chong, Durand, and Ha-Duong 2019), single-molecule fluorescence

spectroscopy (SMFS) (Gall *et al.*, 2015) and mass spectrometry (MS) (Kahsai *et al.*, 2014), fluorescence resonance energy transfer (FRET) (Wang and Geva, 2005; Mazal and Haran, 2019) and atomic force microscopy (AFM) (Visser and Pretorius, 2019), to mention a few. A detailed review of the experimental approaches and advancement made available on the experimental approaches is provided by Orozco, Modesto (Orozco, 2014), Palamini and colleagues (Palamini, Canciani and Forneris, 2016), Maximova and colleagues (Maximova *et al.*, 2016), and Srivastava and colleagues (Srivastava *et al.*, 2018).

4.1.4 Computational approaches

Despite the immense advancement in the experimental approaches for studying protein structure and dynamics, each method presents some shortcomings (Srivastava *et al.*, 2018). Hence, several computational approaches were developed and became essential in mitigating the limitations and integrating the results obtained from the different experimental sources. In the next section, the common key computational methods used for studying protein structural dynamics are reviewed.

MD simulations

Molecular Dynamic (MD) simulations are used widely to study the structural and dynamical aspects and the conformational space of biomolecules on a time scale up to milliseconds (Shaw *et al.*, 2009; Grant, Gorfe and McCammon, 2010; Liu *et al.*, 2017; Gaalswyk, Muniyat and MacCallum, 2018; Harpole and Delemotte, 2018; Jiang *et al.*, 2019; Kawasaki, Soma and Kretsinger, 2019; Bhardwaj and Purohit, 2020) (the theoretical description of MD is detailed in section 3.1.2 of [Chapter 3](#)). MD can provide detailed information on protein dynamics at the atomic level and the global dynamics with relatively low computational cost and time. However, it is still considered computationally inefficient and, in many cases, not feasible, especially with the increasing size of biomolecule systems and that the timescales of biologically key conformational changes are on the order of microseconds (Orellana, 2019). The accessible time scale of all-atom MD beyond the timespan of seconds is often and currently not feasible as the approach scales linearly with the number of simulated atoms and requires several trials of simulations for statistical sampling and inference of the studied system global dynamics (Togashi and Flechsig, 2018; Páll *et al.*, 2020). Bridging the gap between the timescales of

protein conformations that can be studied experimentally and those that MD can study is one of the main challenges, and the need to reduce the cost and time associated with MD simulations even further is needed. One way to reduce MD's associated computational time and the cost is via time-independent approaches such as normal mode analysis (NMA).

Normal modes analysis

Normal mode analysis (NMA) is a powerful and efficient computational approach used to describe the essential vibrational dynamics encoded in protein structures (Bauer, Pavlović and Bauerová-Hlinková, 2019). The approach provides a detailed insight on intrinsic structural dynamics based on the physics of small oscillations (Wako and Endo, 2017; Bauer, Pavlović and Bauerová-Hlinková, 2019). As a result, various flexible states accessible to a protein around an equilibrium position can be deduced.

The theory of NMA was first introduced in the 1950s by Goldstein in classical mechanics (Rosenhead and Goldstein, 1951); since then, the approach has gained popularity and has found applications across different fields. NMA was first applied to study a small peptide (Levy and Karplus, 1979), then in the 1980s to study the bovine pancreatic trypsin inhibitor, a full protein of small size (Noguti and Go, 1982; Go, Noguti and Nishikawa, 1983), followed by relatively larger structures such as the crambin and ribonuclease (Levitt, Sander and Stern, 1985), human lysozyme (Brooks and Karplus, 1985; Levitt, Sander and Stern, 1985), and myoglobin (Seno and Go, 1990), to mention a few.

In NMA, the studied structure is modelled as a set of nodes (atoms) connected by a simple spring with a spring constant; thus, the motion of a simple spring can be analytically solved/described. As the method's name indicates, the resultant vibrational modes are normal to one another, meaning that they move independently. The average oscillation amplitude of a given mode scales as the inverse of its frequency; in that way, modes with higher frequencies describe rapid but small amplitude local motions involving fewer atoms, while those with lower frequencies describe slower displacements involving larger numbers of atoms and thus describe large-scale conformational changes (Bauer, Pavlović and Bauerová-Hlinková, 2019). Notably, the high-frequency vibrating residues correspond

to kinetically hot residues that are important for the stability of the protein (Bahar *et al.*, 1998; Demirel *et al.*, 1998).

Initially, the NMA calculation involved all atoms in a studied structure, including the hydrogen atoms; for large proteins with thousands of atoms, this comes with a computational expense. For that reason, the number of possible residues for the analysis was limited, and the computational cost hindered the extension of NMA to even larger systems. Furthermore, with the development of X-ray crystallography, larger biological systems became more available, limiting the use of the NMA and establishing the need for new methods of NMA to study these larger systems efficiently. Several coarse-grained approximation methods have been developed to overcome such limitations. The most common and widely used of these approaches is the elastic network model (ENM) developed by Tirion (Tirion, 1996). The framework proposed by Tirion further replaced the complicated semi-empirical potentials used in NMA by a simple harmonic potential (a simple pairwise Hookean potential) between atoms within a specified connectivity cutoff of each other. In that way, the protein carbon atoms are modelled as a collection of nodes that are sufficiently close to each other and are connected by a spring that obeys Hooke's law, thus allows solving proteins' intrinsic motions subject to a harmonic potential.

Although the ENM approach does not account for all atoms in the complex and further employs the simple harmonic potential, several large-scale studies have shown the success of this approach in describing the vibrational dynamics of biomolecules and large complexes, and further found agreement between features of the large-scale motions predicted by NMA and the ones characterized experimentally (Bahar and Jernigan, 1998; Bahar *et al.*, 1998; Demirel *et al.*, 1998; Keskin, Jernigan and Bahar, 2000; Tama and Sanejouand, 2001; Temiz and Bahar, 2002; Isin, Doruker and Bahar, 2002; Liu *et al.*, 2008; Lezon, Sali and Bahar, 2009; Isin *et al.*, 2012; Fuglebakk, Tiwari and Reuter, 2015; Lee *et al.*, 2017; Choi *et al.*, 2017; C. J. Ross *et al.*, 2018; Khairallah, Ross and Tastan Bishop, 2021, 2020). Finally, with the advancement in computer processing power and data storage capacity, the performance of NMA calculations has been further improved, allowing NMA on an entire virus capsid (Hsieh *et al.*, 2016).

Anisotropic network model and Gaussian network model

Different ways of implementing the ENM have been made available. The two commonly used ENM methods are the gaussian network model (GNM) (Bahar, Atilgan and Erman, 1997; Haliloglu, Bahar and Erman, 1997) and the anisotropic network model (Hinsen, 1998; Doruker, Atilgan and Bahar, 2000; Tama and Sanejouand, 2001). The ANM method evaluates the directional preferences of the protein vibrational dynamics, whereas GNM assumes that fluctuations are isotropic following a Gaussian distribution around a mean position; thus, it can only provide information about the displacements and correlations between atoms (Bahar, Lezon, Yang, *et al.*, 2010). The GNM only requires solving a singular value decomposition problem of the Hessian matrix, and therefore it makes the calculations computationally inexpensive relative to the ANM (Bauer, Pavlović and Bauerová-Hlinková, 2019).

4.2 Methods

The study presented in this chapter was divided into two main parts; the first part describes the NMA of the malaria parasite *Plasmodium falciparum* (*P. falciparum*) GCH1 homology model to study the structural dynamics of the protein and pinpoint potential allosteric drug targeting sites. In the second part, NMA was applied to the *Rattus norvegicus* (*R. norvegicus*) GCH1 in the absence and presence of its regulatory protein. This was done to understand the dynamics of the mammalian GCH1 alone and further decipher the effect induced on GCH1 upon its interaction with GFRP to reveal regions that allosterically modulate conformational shifts. Notably, GFRP has not been identified in the malarial parasite; hence, the information obtained from the NMA of *R. norvegicus* GCH1 can be of value to investigate potential differences in the regulatory mechanisms compared to the *P. falciparum* GCH1.

4.2.1 Structures retrieval

Due to the absence of a crystal structure of the *P. falciparum* GCH1, a homology model structure was constructed and used in the NMA analysis. The *P. falciparum* homology modelling process is detailed in section 3.2.2 of [Chapter 3](#). Based on the availability of structural data at the time the study was conducted, NMA was performed on the *R. norvegicus* GCH1 crystal structures (extracted from the inhibitory complex of the *R. norvegicus* (PDB ID: 1WPL), the GCH1-GFRP inhibitory complex of *R. norvegicus* (PDB ID: 1WPL) (Maita *et al.*, 2004), and the GCH1-GFRP stimulatory complex of *R. norvegicus* (PDB ID: 1IS7) (Maita *et al.*, 2002). Both complexes consist of the GCH1 homodecameric structure (1940 residues) and two GFRP pentamers (840 residues).

4.2.2 Multiple sequence alignment

Multiple sequence alignment was performed on selected GCH1 sequences from different organisms, including *P. falciparum*, *Drosophila Melanogaster* (*D. Melanogaster*), *Homo sapiens* (*H. sapiens*), *R. norvegicus*, *E. coli*, and *T. thermophilus*. Given the reported difference in the GCH1 regulatory components between higher eukaryotes, lower eukaryotes, and bacteria, such sequences were selected to understand the regulatory elements further and link the sequence information to the NMA results.

The sequences were first retrieved from the Universal Protein Resource (UniProt) (<http://www.uniprot.org>) online database (Apweiler, 2008). The Tree-based Consistency Objective Function Evaluation (T-COFFEE) webserver (Di Tommaso *et al.*, 2011) under the Espresso mode for structure-based alignment was used for the multiple sequence alignment (MSA) with default alignment parameters. Structural information from the 3D structures of the *T. thermophilus* (PDB ID:1WUR) was included in the alignment process. Lastly, the ESPript webserver (Robert and Gouet, 2014) was used to render sequence similarities and secondary structure information and visualise the alignment output.

4.2.3 Calculation of the normal modes and atomic fluctuation

The normal modes of the proteins were calculated based on the ANM (Atilgan *et al.*, 2001). Specifically, elastic network models were constructed on the atomic coordinates of all the C_β atoms given in the selected PDB files. The C_α atoms were used to represent the glycine residues, as they lack the C_β atoms. The C_β atoms were selected following previous ENM studies showing that when a single atom represents a whole residue, the C_β atoms better represent the side-chain orientation (Atilgan, Turgut and Atilgan, 2007). The NMA was performed using the MODE-TASK software suite (C. Ross *et al.*, 2018). The software suite employs the ANM (Atilgan *et al.*, 2001) to construct an ENM of the protein structure. Essential outputs from the MODE-TASK ANM toolset include a coarse-grained PDB structure and a complete set of eigenvalues and eigenvectors.

In the ANM, each residue of the protein structure was reduced to one node placed at its C_β atom coordinate, and all interactions between each pair of residues separated within a defined cutoff distance (R_c) = 15 Å, were modelled as a set of springs with a uniform force constant γ . This yielded a network that contained N nodes and M springs (N residues connected by M springs) representing the total number of interactions defined within the network, such that any given pair of nodes within R_c of each other will interact in accord with a conventional harmonic potential. The interacting cut-off distance, R_c , was defined as 15 Å. Following a series of initial analyses performed at increasing cut-off distances from 7 to 30 Å, it was found that 15 Å was the minimum R_c , which yielded six trivial

modes, therefore allowing for sufficient connection between nodes across the structure. The normal modes of this system could then be solved by applying the following:

In the absence of an external force, the equilibrium condition for each residue can be invoked. This condition requires that the respective sums of the x-, y- and z- components of the internal forces (f) that act on the residue be equal to zero.

Accordingly, the following equations can be derived:

$$\begin{aligned}\sum_j f_{ij} \cos \alpha_{ij}^X &= \sum_j f_{ij} (X_i - X_j)/s_{ij} = 0 \\ \sum_j f_{ij} \cos \alpha_{ij}^Y &= \sum_j f_{ij} (Y_i - Y_j)/s_{ij} = 0 \\ \sum_j f_{ij} \cos \alpha_{ij}^Z &= \sum_j f_{ij} (Z_i - Z_j)/s_{ij} = 0 \quad \text{Equation 4.1}\end{aligned}$$

Where f_{ij} is the force acting on node i from its interaction with node j , s_{ij} equals the separation vectors between the nodes i and j and α_{ij}^X is the angle between the X -axis and the force f_{ij} . The summations are performed over all near neighbours of i . As the system contains multiple interacting residues, this force balance must be extended to the complete set of N nodes and M interactions, which leads to the derivation of the $3N \times M$ direction cosine matrix \mathbf{B} :

$$[\mathbf{B}]_{3N \times M} [\mathbf{f}]_{M \times 1} = [\mathbf{0}]_{3N \times 1} \quad \text{Equation 4.2}$$

For a defined cut-off distance R_C , each node will have multiple interacting neighbours, each of which exerts a force on the node such that the $3N$ equalities in equation 4.2 are fewer than the number of unknown forces (M). Therefore, the system of equations is underdetermined. In order to solve the system for all unknowns, two additional properties can be invoked. Firstly, the forces can be related to the deformations $\Delta \mathbf{s}$ by the linear approximation:

$$[\mathbf{f}]_{M \times 1} = [\mathbf{K}]_{M \times M} [\Delta \mathbf{s}]_{M \times 1} \quad \text{Equation 4.3}$$

Where $[\mathbf{K}]_{M \times M}$ is a diagonal matrix whose i th element is the force constant γ of the i th residue pair. Secondly, the kinematics relationship between the deformation and the residue positions is given by:

$$[\mathbf{B}]_{M \times 3N}^T [\Delta \mathbf{R}]_{3N \times 1} = [\Delta \mathbf{s}]_{M \times 1} \quad \text{Equation 4.4}$$

By substituting equation 4.3 and equation 4.4 into equation 4.2 the following relationship is obtained:

$$[\mathbf{B}]_{3N \times M} [\mathbf{K}]_{M \times M} [\mathbf{B}]_{M \times 3N}^T [\Delta \mathbf{R}]_{3N \times 1} = [\mathbf{0}]_{3N \times 1} \quad \text{Equation 4.5}$$

If all M bonds in the network are assumed to be harmonic interactions with uniform force constants, then $\gamma[\mathbf{B}][\mathbf{B}]^T$ is equivalent to the Hessian matrix \mathbf{H} . In the general case of N residues connected by M springs, \mathbf{H} is a $3N \times 3N$ super-matrix that may be derived from the second derivatives of the overall potential V , with respect to the components of \mathbf{R}_i , where \mathbf{R}_j are the fluctuation vectors of the individual residues. Therefore, the Hessian matrix describes the force constant of the system. \mathbf{H} is composed of $N \times N$ super-elements, i.e.

$$\mathbf{H} = \begin{bmatrix} H_{11} & H_{12} & \cdots & H_{1N} \\ H_{21} & & \cdots & H_{2N} \\ \vdots & & \cdots & \vdots \\ H_{N1} & & \cdots & H_{NN} \end{bmatrix} \quad \text{Equation 4.6}$$

Where each super-element \mathbf{H}_{ij} is a 3×3 matrix that holds the anisotropic information regarding the orientation of nodes i, j , the ij th super-element ($i \neq j$) of \mathbf{H} is defined as:

$$\mathbf{H}_{ij} = \begin{bmatrix} \partial^2 V / \partial X_i \partial X_j & \partial^2 V / \partial X_i \partial Y_j & \partial^2 V / \partial X_i \partial Z_j \\ \partial^2 V / \partial Y_i \partial X_j & \partial^2 V / \partial Y_i \partial Y_j & \partial^2 V / \partial Y_i \partial Z_j \\ \partial^2 V / \partial Z_i \partial X_j & \partial^2 V / \partial Z_i \partial Y_j & \partial^2 V / \partial Z_i \partial Z_j \end{bmatrix} \quad \text{Equation 4.7}$$

At equilibrium, the second derivatives may be calculated for the ANM using the C_β position vectors of PDB structures such that the equation gives the elements of the off-diagonal \mathbf{H}_{ij} :

$$\partial^2 V / \partial X_i \partial Y_j = -\gamma (X_j - X_i)(Y_j - Y_i) / S_{ij}^2 \quad \text{Equation 4.8}$$

And the elements of the diagonal super-elements \mathbf{H}_{ii} are given by the equations:

$$\partial^2 V / \partial X_i^2 = \gamma \sum_j (X_j - X_i)^2 / S_{ij}^2 \quad \text{Equation 4.9}$$

For the diagonal elements of \mathbf{H}_{ii} and

$$\partial^2 V / \partial X_i \partial Y_j = \gamma \sum_j (X_j - X_i)(Y_j - Y_i) / S_{ij}^2 \quad \text{Equation 4.10}$$

For the off-diagonal elements of \mathbf{H}_{ii}

Note that \mathbf{H} and $[\mathbf{B}][\mathbf{B}]^T$ are equivalent to each other; thus, all elements of $[\mathbf{B}][\mathbf{B}]^T$ and \mathbf{H} are equal.

As an example, consider the first diagonal element of \mathbf{H}_{11} of \mathbf{H} ,

$$[\gamma \mathbf{B} \mathbf{B}^T]_{11} = \gamma \sum_j \cos^2 \alpha_{1j}^X = \gamma \sum_j (X_j - X_1)^2 / S_{1j}^2 \quad \text{Equation 4.11}$$

\mathbf{H} can be decomposed into $3N-6$ eigenvalues and $3N-6$ eigenvectors corresponding to the respective frequencies and directions of the individual non-trivial modes. Thus, for each normal mode, a set of eigenvectors is obtained for each eigenvalue; the eigenvectors represent the direction/displacement at every C_β atom in the protein for a given mode. Modes with the lowest frequencies are termed the slowest modes and define the most collective, or global, motions of the protein, whereas the highest-frequency modes describe the more localised and rapid motions within the studied system. The inverse of \mathbf{H} is equivalent to the covariance matrix \mathbf{C} that is composed of $N \times N$ super-elements. Each off-diagonal, ij th, super-element of \mathbf{H}^{-1} contains the 3×3 matrix of correlations between the x-, y- and z-components of fluctuation vectors of residues i and j . In contrast, the i th super-element of \mathbf{H}^{-1} describes the self-correlation between the components of fluctuation vectors of residue i . The mean square fluctuation (MSF) of the C_β atoms of each residue was then obtained by summing the self-correlations in the x-, y- and z- components of the atomic fluctuations in the respective modes.

The MSF of the normal modes were calculated using the MODE-TASK software suite (C. Ross *et al.*, 2018), and the collective MSF profiles were obtained by summing the fluctuations over a specific mode range. The algorithm employed within the MODE-TASK software suite builds the \mathbf{H} matrix on the C_β atoms of each residue by incorporating the ALGLIB library to calculate the pseudoinverse of \mathbf{H}

by singular value decomposition (C. Ross *et al.*, 2018), consequently produce the eigenvalues and eigenvectors of the normal modes.

4.2.4 Visualization of the normal modes

The displacement vectors of the selected low-frequency non-degenerate modes were studied by projecting the respective eigenvectors of each of the identified modes onto the protein structure. Furthermore, movies were constructed to visualise global motions by projecting the eigenvectors onto the structure as a set of frames; the vectors were added to the original atomic coordinates in increasing steps, and corresponding arrows were given in a Tcl script. The visual molecular dynamics (VMD) program (Humphrey, Dalke and Schulten, 1996) was used to visualize the vector arrows (eigenvectors) describing the structural change in each mode.

4.2.5 Residue cross-correlation analysis

The magnitude of the pairwise cross-correlation coefficients was calculated using the BIO3D R package for the exploratory analysis of the structure and sequence data (Grant *et al.*, 2006). The BIO3D pairwise cross-correlation is given by the equation below:

$$S(i, j) = (\Delta r_i \cdot \Delta r_j) / (\Delta r_i^2)^{\frac{1}{2}} (\Delta r_j^2)^{\frac{1}{2}} \quad \text{Equation 4.12}$$

Where Δr_i and Δr_j are the displacement vectors for atoms i and j , respectively. The elements $S(i, j)$ are stored in matrix form and displayed as a three-dimensional dynamical cross-correlation map. If $S_{ij} = 1$, the fluctuations of atoms i and j are completely correlated; if $S_{ij} = -1$, the fluctuations of atoms i and j are completely anti-correlated; and if $S_{ij} = 0$, the fluctuations of i and j are not correlated. The BIO3D cross-correlation $S(i, j)$ is given by:

$$S(i, j) = (\Delta r_i \cdot \Delta r_j) / (\Delta r_i^2)^{\frac{1}{2}} (\Delta r_j^2)^{\frac{1}{2}} \quad \text{Equation 4.13}$$

4.2.6 Deformation energy analysis

The deformation energy given in equation 4.12 is defined as the energy density due to deformation as a function of atomic position (Hinsen, 1998). In that way, the analysis provides a measure of local

flexibility within the protein structure; As a result, it can disclose hinge sites (Demerdash, Daily and Mitchell, 2009). The deformation energy was calculated over the 20 lowest non-trivial normal modes from each input structure using the BIO3D R package for the exploratory analysis of the structure and sequence data (Grant *et al.*, 2006).

$$E = \frac{1}{2} \sum_1^N K \left(R_{ij}^{(0)} \right) \frac{\left| (r_i - r_j) \cdot R_{ij}^{(0)} \right|^2}{\left| R_{ij}^{(0)} \right|^2} \quad \text{Equation 4.14}$$

Where r_i , r_j donate the displacement of atom i and j in the mode to be analysed, $R_{ij}^{(0)}$ is the pair distance vector ($R_i - R_j$) in the input arrangement, and K is the pair force constant.

4.2.7 Mechanical stiffness analysis

The mechanical stiffness analysis evaluates the protein mechanical resistance by calculating effective force constants responding to uniaxial extensional forces exerted at each residue pair (Eyal and Bahar, 2008; Bakan, Meireles and Bahar, 2011; Mikulska-Ruminska *et al.*, 2017). The analysis is essential to disclose rigid/stiff regions that can be essential for maintaining the protein stability and folding thus function, or it can reveal flexible sites responsible for modulating the protein's global conformational shifts. The ProDy Python package for protein structural dynamics analysis (Bakan, Meireles, and Bahar, 2011; Bakan *et al.*, 2014), in particular, the ProDy MechStiff function was used to construct the mechanical resistance matrix of the protein structure (Bakan, Meireles and Bahar, 2011; Bakan *et al.*, 2014) as well as a 2D profile of effective stiffness for each residue averaged over all pairs it forms. The obtained results were visualized using Matplotlib library and VMD.

4.2.8 Modes overlap analysis between the inhibitory and stimulatory GCH1-GFRP complexes

The mode overlap analysis measures how much of the NMA predicted displacement has contributed to the transition towards an experimentally determined conformation (C. Ross *et al.*, 2018). In this case, the overlap between the predicted displacement (ΔR_j), obtained from the NMA of the GCH1-GFRP inhibitory complex (PDB ID: 1WPL), and the experimentally known conformation change observed during the shift from the inhibitory complex to the GCH1-GFRP stimulatory complex (PDB ID: 1IS7). The modes overlap was calculated over each residue i , $\Delta S_i = S_{Ii} - S_{Si}$, and is defined as:

$$O_j = (\Delta R_j)^i \cdot (\Delta S)^i / \left| (\Delta R_j)^i \right| \left| (\Delta S)^i \right| \quad \text{Equation 4.15}$$

In order to measure the similarity between an individual normal mode with eigenvector A_j and the experimentally known conformational change, the above equation was used with $\Delta R_j = A_j$ to calculate the overlap between the two vectors A_j and ΔS (C. Ross *et al.*, 2018). An overlap of 1 indicates that the predicted displacement of the normal mode acts in the same direction of the conformational change ΔS . In contrast, a value of zero indicates that the mode and the experimental change vectors are orthogonal.

4.2.9 Root mean square inner product analysis

The root mean square inner product (RMSIP) measures a cumulative overlap between sets of principal eigenvectors obtained from normal modes analysis to quantify similarity in the directionality of residue motions (Fuglebakk, Tiwari and Reuter, 2015; Yazhini and Srinivasan, 2020). The RMSIP was calculated for the *T. thermophilus* GCH1 crystal structure and the *P. falciparum* GCH1 homology model. It was also calculated for the stand-alone *R. norvegicus* GCH1 and *P. falciparum* GCH1 homology. Lastly, for the GCH1-GFRP inhibitory and stimulatory complexes. The RMSIP analysis was essential to illustrate the overlapping principal subspaces/set of modes between the structures. The RMSIP score can range from 0 (independent/orthogonal) subspaces to 1 (identical) subspaces. The RMSIP was computed for the essential subspaces (the first 20 non-trivial modes) of each input structure using the BIO3D function (Leo-Macias *et al.*, 2005).

$$\text{RMSIP}(I, J) = \left(\frac{1}{I} \sum_{i=1}^I \sum_{j=1}^J (u_i \cdot v_j)^2 \right)^{\frac{1}{2}} \quad \text{Equation 4.16}$$

Where I and J represent the modes/vectors to be compared, and u_i and v_j are eigenvectors of the selected two subspaces.

4.3 Results and Discussion

PART I:

In this part, NMA was applied to the malaria parasite *P. falciparum* GCH1 homology model. NMA of the *P. falciparum* GCH1 was carried out for clinical implications involving this enzyme, such as the understanding the allosteric modulation of its activity to reveal novel allosteric drug targeting sites for antimalarial drug discovery efforts.

4.3.1 NMA of the *P. falciparum* GCH1 homology model

NMA was first carried out on the available GCH1 crystal structure of *T. thermophilus* (PDB ID: 1WUL) as a control, given that this structure is the template used to model the *P. falciparum* GCH1 homology model. NMA was also performed for the *P. falciparum* GCH1 homology model, and the results were compared to the ones obtained from the *T. thermophilus* GCH1 crystal structure. Information about the template structure, sequence alignment and homology modelling process is provided as supplementary material (Appendix 1.1-Appendix 1.5).

NMA of the *T. thermophilus* GCH1 crystal structure (PDB ID: 1WUR) yielded a total of 5538 non-trivial modes and 5478 of the *P. falciparum* GCH1 homology model. The normal modes were obtained from the decomposition of the $3N \times 3N$ **H** matrix constructed on all C β (and C α in Gly) atoms.

The similarity between the two sets of modes from the *T. thermophilus* GCH1 crystal structure and the *P. falciparum* GCH1 homology model was quantified through the RMSIP. The RMSIP value over the 20 lowest-frequency modes was 0.95 (Figure 4.1-A); this is considered as excellent, suggesting a high similarity between the two sets of modes (Amadei, Ceruso and Di Nola, 1999). The MSF profiles of the 20 lowest-frequency modes from both structures were also captured and compared; both proteins exhibited similar fluctuation profiles, and the results were comparable (Figure 4.1-B). The N-terminal regions displayed high flexibility, and the core was stable. The *P. falciparum* GCH1 global

motions were guided by the protein terminal helices (legs), resulting in the tunnel and core deformation. Overall, the obtained results validate the NMA of the homology model.

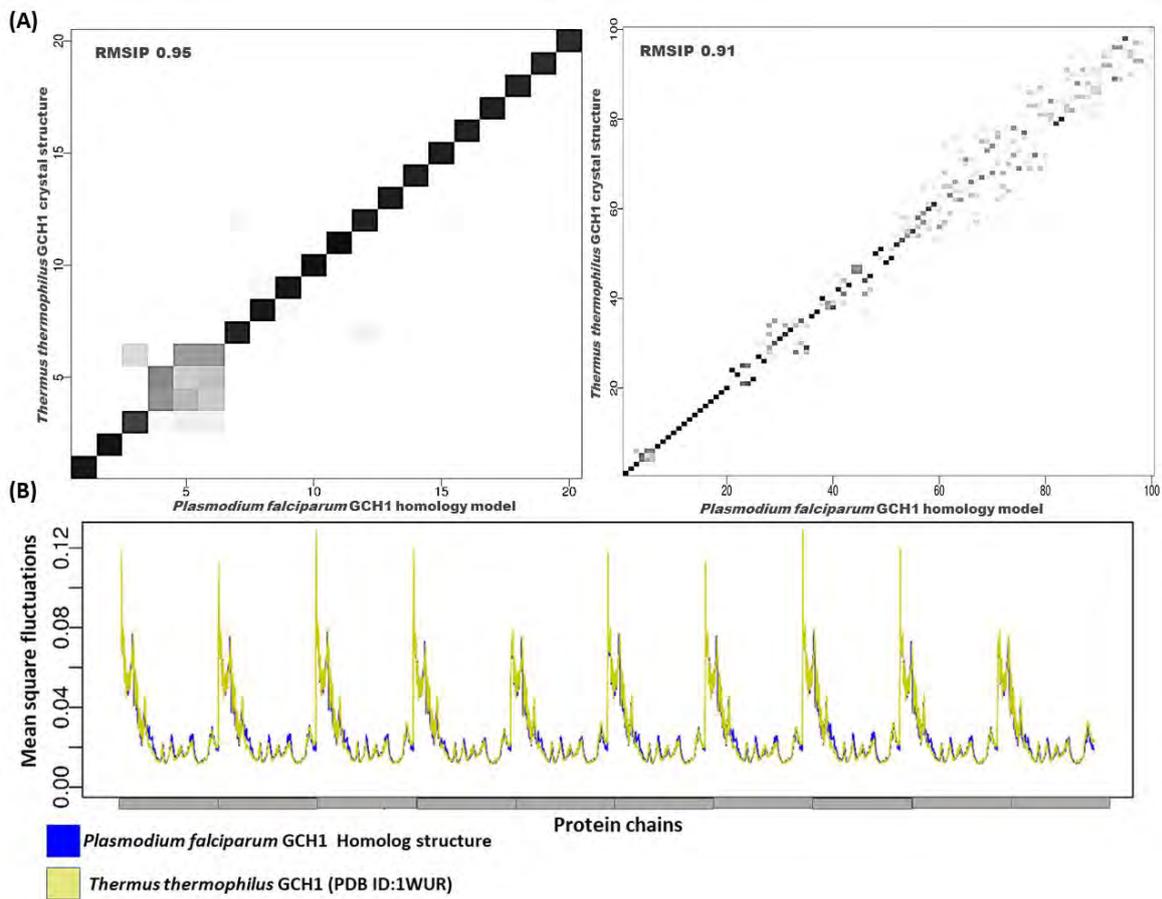


Figure 4.1. (A) The RMSIP of the first 20 non-trivial modes subspaces (left) and the first 100 non-trivial modes subspaces from both *T. thermophilus* GCH1 crystal structure (PDB ID:1WUR) and *P. falciparum* GCH1 homology model structure (B) MSF profiles of the 20 lowest-frequency normal modes of *T. thermophilus* GCH1 crystal structure (PDB ID:1WUR) and *P. falciparum* GCH1 homology model. The grey blocks partition each chain of the GCH1 homo-decameric structure.

Given that highly collective protein motions are captured within the low-frequency modes (Bahar, Lezon, Bakan, *et al.*, 2010), the first 20 non-trivial low-frequency modes of the *P. falciparum* GCH1 homology model were extracted and presented in Table 4.1. Within the selected low-frequency modes, both degenerate and non-degenerate modes were identified. The degenerate modes share identical eigenvalues and comprise modes for which any spatial rotation results in a valid representation of the motion. In contrast, non-degenerate modes appear with unique eigenvalues and act equally across the entire structure. Non-degenerate modes present the simultaneous displacement

of all monomers within a complex and are instrumental in showing the transformational transitions of large multimeric structures (Bahar, Lezon, Bakan, *et al.*, 2010). This concept has been proved in previous studies, in which dominant conformational changes were captured in one or two non-degenerate modes (Atilgan *et al.*, 2001; Chennubhotla *et al.*, 2005; Shrivastava and Bahar, 2006; Wako and Endo, 2011; Isin *et al.*, 2012; Lee *et al.*, 2017; C. Ross *et al.*, 2018). When summed, the degenerate modes can present symmetric global motions (Zheng, Brooks and Thirumalai, 2007). The linear combinations of the *P. falciparum* GCH1 degenerate modes over the 20 lowest-frequency modes was computed and showed symmetric global motions (Appendix 2.1).

4.3.2 Visual characterization of the *P. falciparum* GCH1 low-frequency non-degenerate normal modes

Large collective protein motions comprise a network of residues that modulate the protein conformational transitions hence function. Identifying such key residues can be instrumental for understanding the mechanism of function and pinpointing drug targeting sites (Yang and Bahar, 2005). Within the first 20 non-trivial low-frequency modes, four non-degenerate modes were identified (Mode 11, 12, 13, and 18) (Figure 4.2-A). The identified four non-degenerate modes presented collective/large-scale motions of the *P. falciparum* GCH1, including an anti-correlated movement of the protein two pentameric units resulting in the narrowing of the protein tunnel in **Mode 11** ([Movie 1](#)) as well as a common outward movement of the central five-helix bundles in **Mode 12** ([Movie 2](#)). **Mode 13**, however, featured a global breathing motion ([Movie 3](#)). Lastly, **Mode 18** presented a global wringing movement ([Movie 4](#)).

The MSF profile of the C_β atoms from each residue was calculated to gain a detailed insight into the atomic protein flexibility within each of the four slowest non-degenerate modes and deduce the atomic contribution to the collective dynamics (Figure 4.2-B). One chain representation of the MSF profiles from the selected non-degenerate modes is shown in Figure 4.2-B; residues of considerable mobility are illustrated on each profile. The MSF profiles of the complete homo-decameric structure are provided shown in Appendix 2.2.

The MSF profiles of the *P. falciparum* GCH1 individual normal modes showed the high fluctuation of the N-terminal residues E200, K221, N244 and D246, corresponding to L68, Q87, Q110, T112 in the human GCH1 sequence and the neighbouring residues of the BH4 binding site such as N259, Y332 which correspond to N159 and A196 in the human GCH1 sequence. The BH4 binding site residues of the human GCH1 E370, N373, and V376 corresponding to F234, D237, and T240 also showed some fluctuation in **Mode 12**. Notably, the *R. norvegicus* BH4 binding site residues F225 and T231 corresponded to the *P. falciparum* GCH1 residues E370 and V376 and were later identified among the residues of notable fluctuation in **Mode 12** of the *R. norvegicus* stand-alone GCH1 (Figure 4.6). The identified residues of notable fluctuation within the non-degenerate slow modes thus demonstrate their importance for the protein conformational changes and functionally essential.

Table 4.1. Low-frequency normal modes of the *P. falciparum* GCH1 homology model, associated eigenvalues, and degeneracy level. The first six trivial modes were excluded, and therefore Mode 1 represents the computed Mode 7.

Mode	Eigenvalue	Degeneracy
Mode 1	0.48	2
Mode 2	0.49	2
Mode 3	0.57	2
Mode 4	0.57	2
Mode 5	0.58	2
Mode 6	0.58	2
Mode 7	0.71	2
Mode 8	0.71	2
Mode 9	0.72	2
Mode 10	0.72	2
Mode 11	0.73	1
Mode 12	0.80	1
Mode 13	1.13	1
Mode 14	1.18	2
Mode 15	1.20	2
Mode 16	1.21	2
Mode 17	1.23	2
Mode 18	1.51	1
Mode 19	1.52	2
Mode 20	1.52	2

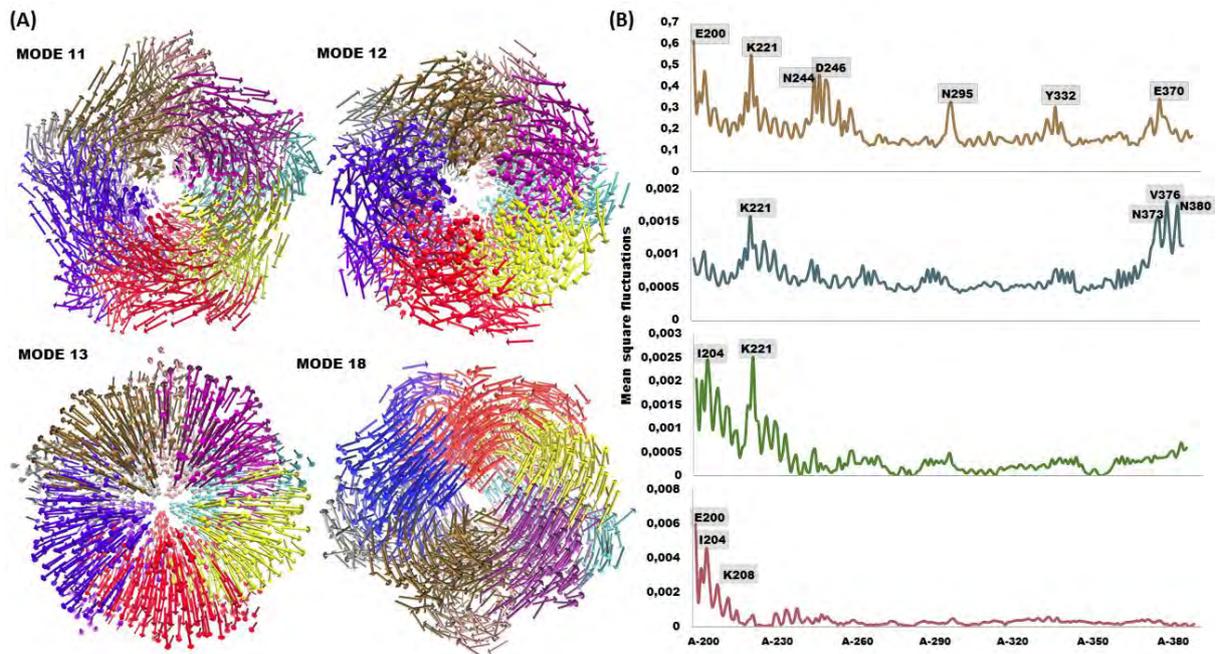


Figure 4.2. (A) The characterized global motions of the *P. falciparum* GCH1 homology model structure within the four non-degenerate low-frequency modes. Each arrow denotes the protein C β atoms vector of movement. Each of the protein chains is coloured differently for illustration purposes. (B) Individual MSF profiles of the four non-degenerate low-frequency modes, from top to bottom Mode 7, 12, 15, and 18. Residues of a notable fluctuation are labelled on the MSF plot of the one GCH1 chain. (Reproduced with permission from (Khairallah, Ross and Tasthan Bishop, 2021)).

4.3.3 Multiple sequence alignment and NMA

MSA of GCH1 sequences from *P. falciparum*, *D. melanogaster*, *H. sapiens*, *R. norvegicus*, *E. coli*, and *T. thermophilus* was carried out to evaluate the sequence variation within key regions identified by the NMA and further connect it to the GCH1 intrinsic fluctuation profile. Information about the retrieved GCH1 sequences is presented in Appendix 2.3. The MSA showed the high sequence variation at key sites, including the N-terminal regions, in which the *P. falciparum* and *D. melanogaster* GCH1 had the most heterogeneous extended N-terminal sequences (Figure 4.3-A). The BH4 binding site and the central five-helix bundles also varied among the different species. Notably, the BH4 binding site was conserved in the *H. sapiens*, and *R. norvegicus* GCH1 sequences, then varied to the *P. falciparum* GCH1 sequence (Figure 4.3-A).

Most importantly, the two key residues of the human GCH1 BH4 binding site R235 and R241 were substituted in the *P. falciparum* GCH1 by K371 and H377, respectively. The active site surrounding

residues and the GFRP binding sites also varied across the different species (Figure 4.3-A). The GCH1 active site catalytic residues remained highly conserved among all the species; hence, the GCH1 catalytic mechanism is well conserved across species. The sequence variation in the N-terminal region and allosteric BH4 sites can suggest different regulatory components of GCH1 across species, primarily that the GCH1 regulatory protein GFRP has not been found in both *P. falciparum*, *D. melanogaster* organisms. The difference of the GCH1 regulatory component can also be reasoned to its distinct functional role in prokaryotes. In this next section, the fluctuation of residues within the identified sites of notable sequence variability is studied and evaluated.

4.3.4 Collective MSF analysis of the *P. falciparum* GCH1 normal modes

Modes of low-frequency describe collective (large-amplitude) global motions that can provide relevant information about the protein global dynamics and function (Rosca *et al.*, 2002; Tama *et al.*, 2003; Delarue and Dumas, 2004; Ma and Nussinov, 2010; Wako and Endo, 2011). On the other hand, high-frequency modes usually denote kinetically hot residues “hotspot” regions of rapid local motions (Bahar *et al.*, 1998; Emekli *et al.*, 2008; Dutta and Bahar, 2010; Liu and Bahar, 2010). A study by Haliloglu and colleagues further confirmed that highly conserved residues/hotspots often vibrate with high frequencies and contribute significantly to protein function, folding, and stability (Haliloglu *et al.*, 2005).

The collective MSF profile of the 20 lowest-frequency modes of the *P. falciparum* GCH1 from one chain displayed the high fluctuation of the N-terminal helices (residues: E202-D246) (Figure 4.3-B). In contrast, the collective MSF profile of the 20 highest-frequency modes showed a high fluctuation of residues within the protein core, such as the active site neighbouring residues I290, L301 and V340. Other residues of notable fluctuation were located at other key sites, including the BH4 and GFRP binding sites (Figure 4.3-C); this includes K267 and Y364 within the BH4 binding site and T322 within the GFRP binding site. Notably, I290 and V340 were conserved among the studied sequences, whereas I131 substituted K267 in the human GCH1 sequence. Furthermore, K267, I290, and V340 corresponded to I122, I145 and V195 in *R. norvegicus* GCH1; these residues were also identified with a notable fluctuation in the 20 highest-frequency normal modes MSF profile of *R. norvegicus* GCH1

(Figure 4.7-D). Overall, the NMA of *P. falciparum* disclosed hotspot residues unique to the *P. falciparum*; such residues were located within the BH4, GCH1-GFRP binding sites, active site surrounding region, the central five-helix bundle, and the unique extended N-terminal helices sites. Lastly, despite the GCH1 sequence disparities within the different species regulatory components, the NMA from three representative structures showed similar fluctuation profiles and structural dynamics (Appendix 2.4). This finding further confirms that the intrinsic flexibility of proteins is far more conserved than the structure and sequence.

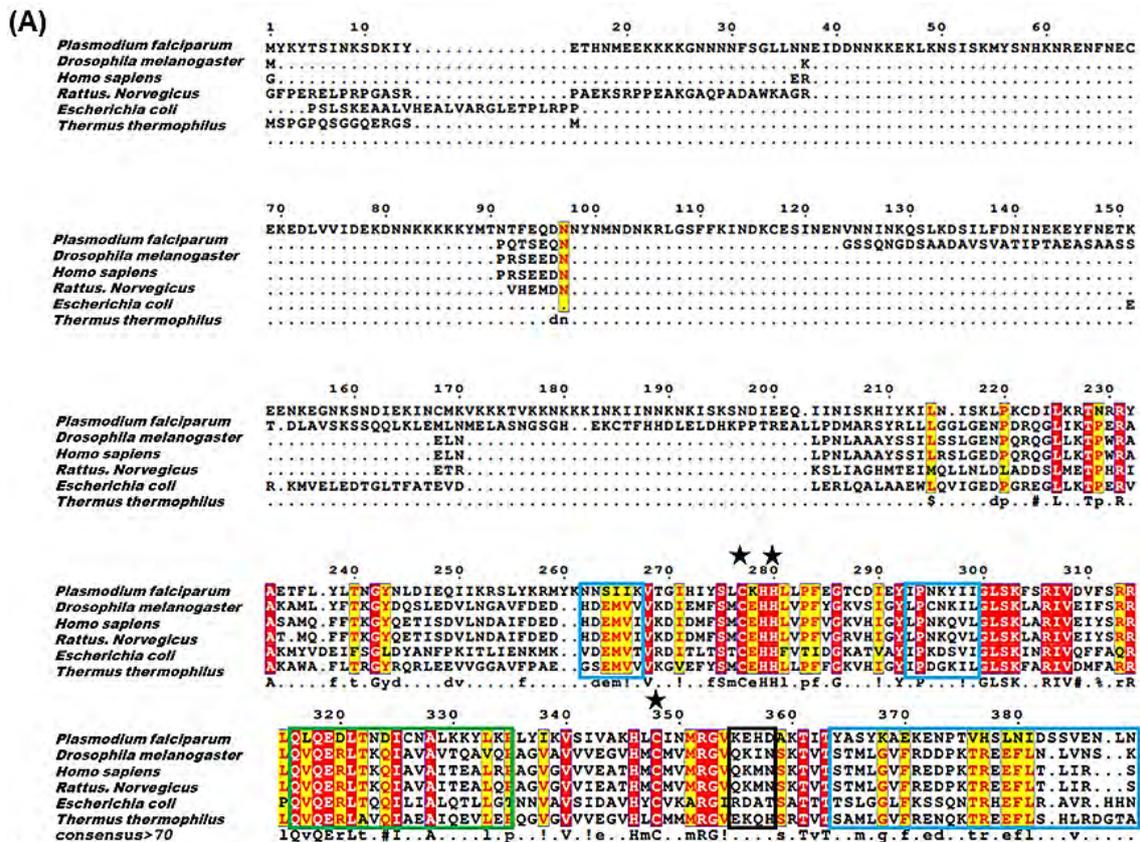


Figure 4.3. (A) Multiple sequence alignment of the *P. falciparum*, *D. melanogaster*, *H. sapiens*, *R. norvegicus*, *E. coli*, and *T. thermophilus* GCH1 sequences. The black stars show the catalytic and metal coordinating residues. The black block shows the loop region surrounding the active site pocket. The blue blocks show the BH4 binding site, and the green block shows the GFRP binding site. One chain representation of the *P. falciparum* GCH1 (B) 20 Lowest-frequency and (C) 20 highest-

frequency normal modes MSF profiles. The solid red circle highlights the active site region. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2021)).

4.3.5 *P. falciparum* GCH1 hinge residues

Hinge residues are tightly restricted in terms of mobility, thus act as hinges or anchors for collective motions (Su *et al.*, 2011). Such sites are essential for modulating different conformational changes of functional significance and can be identified within the lowest-frequency modes obtained from NMA, in which they appear as minima of fluctuations (Su *et al.*, 2011). The *P. falciparum* GCH1 MSF profile of the 20 lowest-frequency modes disclosed the location of hinge sites within the protein core (Figure 4.3-B). Notably, the key residues of functional importance, such as the active site and neighbouring residues: I267, I290, L301, E319, V340, S364 and F380, to mention a few, are located within the protein core. Hinge residues are established as highly conserved residues of functional importance (Ung *et al.*, 2013). Here, the identified hinge residues I290, I301, V340 and F380 were also shown to be conserved among the different species (Figure 4.3-A), which further confirm their crucial role for the dynamics and function.

4.3.6 Deformation energy analysis of the *P. falciparum* GCH1

Deformation is defined as any structural change due to an external force or a change in temperature (Bao, 2002). Structural deformation is essential for adopting different conformational states that modulate protein activity (Mitchell, Tlustý and Leibler, 2016). Studying the protein deformation energy provides a rigidity measure to disclose hinge regions (Hinsen, Thomas and Field, 1999; Kovacs, Chacón and Abagyan, 2004; Demerdash, Daily and Mitchell, 2009). Rigid areas are hardly deformed; thus, deformation energy analysis was performed to study the distribution of deformation energy across the GCH1 structure and pinpoint rigid and high deformed regions.

The deformation energy analysis of the *P. falciparum* GCH1 disclosed a notable build-up of deformation energy in the protein core, which signifies the existence of an energy barrier that reduced its deformation, demonstrating the high rigidity (local flexibility) of the protein core (Figure 4.4-B). This result also agreed with the MSF profile of the 20 lowest-frequency modes in which rigidity/low atoms mobility was observed in the protein core (Figure 4.4-A). On the other hand, lower

deformation energy was observed in the N-terminal helices, indicating their low rigidity (global flexibility), thus their high degree of collective motion.

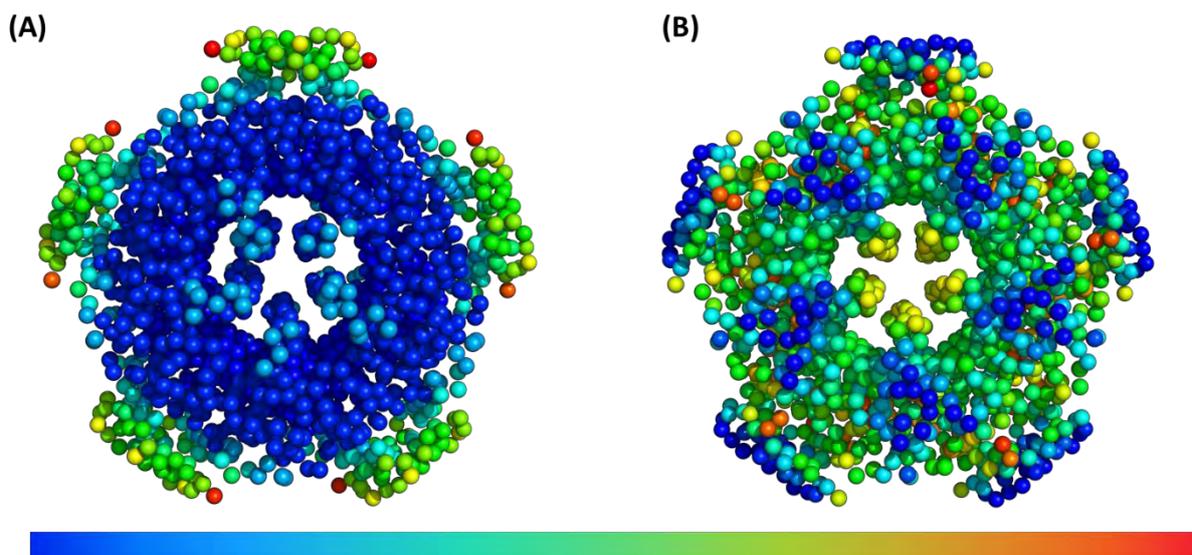


Figure 4.4. (A) MSF values over the 20 lowest-frequency modes mapped onto the *P. falciparum* GCH1 structure. The fluctuations colour scale is blue to red (low to high) atomic fluctuation (B) Atomic deformation energy values mapped onto the *P. falciparum* GCH1 structure. The deformation energy colour scale is blue to red (low to high) deformation energy.

4.3.6 Residue cross-correlation of the *P. falciparum* GCH1

Characterizing correlated motions and the directions in which major change occurs can provide key information about protein allosteric mechanisms. Correlated motions can occur among residues of proximity, which are more likely to move together. Nevertheless, the long-term correlation/anti-correlation is also important as it can show the interdependence of two or more isolated structural regions to attain a desired conformational state. The residue cross-correlation of the *P. falciparum* GCH1 over the first 20 non-trivial modes was calculated and plotted in a heatmap (Figure 4.5). A strong cross-correlation was identified along the diagonal, demonstrating the correlation of neighbouring atoms within the same chain. Off-diagonal positive and negative correlations were also identified. A positive correlation was observed between atoms of chain A and E within the same pentameric unit and across the two pentameric units, specifically, chain A of one pentamer and chain E of the second pentamer. The identified anti-correlation can be explained by the wringing movement

of the N-terminus, in which the wringing in one pentamer is accompanied by an opposing wring movement of the second pentamer ([Movie 1](#)).

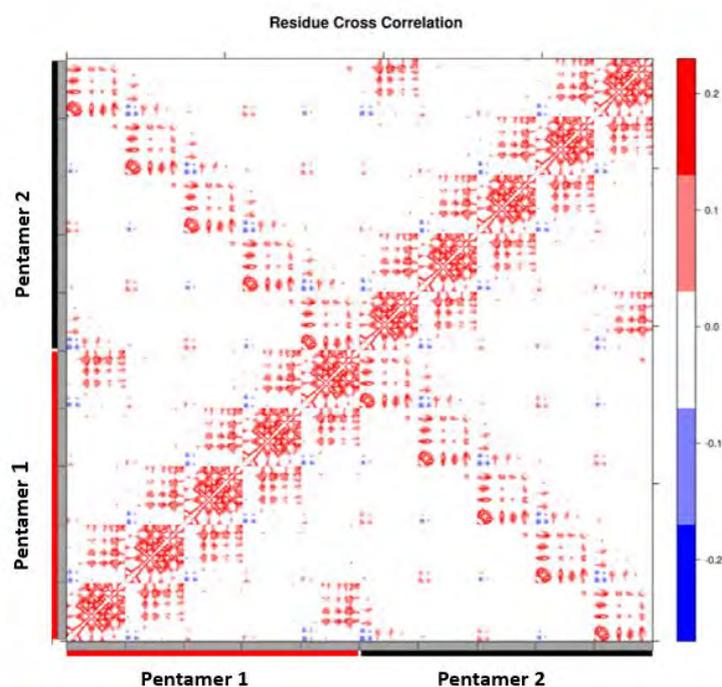


Figure 4.5. A heatmap showing GCH1 residue cross correlation. Correlated regions (red) show the residues moving in the same direction, e.g., upward and upward, and anti-correlated (blue) residues moving in directions that are linearly opposite to each other, e.g., upward and downward. The high diagonal values occur where $i = j$, in this case, the same atom.

Overall, the NMA MSF and deformation energy results of the *P. falciparum* GCH1 revealed unique hotspot residues of high-frequency vibration located in key structural sites within the protein core, including the effector molecule BH4 and GFRP-binding sites, as well as the central five-helix bundle. Furthermore, hinges of collective motions were identified, and the rigidity of the protein core along with the N-terminus high flexibility was confirmed. The identified sites can be of great value for antimalarial drug design studies, in which such locations can be targeted to regulate the functional dynamics of the enzyme and restrict its conformational states, thus, induce allosteric effects.

PART II:

The regulation of the GCH1 activity is imperative for maintaining BH4 biosynthesis levels in higher eukaryotes. Hence, the function of GCH1 is regulated by the partner protein GFRP (details about the GCH1 feedback regulatory protein is presented in [Chapter 1](#)). This part of the study invites special attention to the stand-alone GCH1 and its complex with the GFRP. This was done to understand further the structural dynamics of the mammalian GCH1 and its regulatory protein, the GFRP. Furthermore, disclose allosteric drug targeting sites that can aid in developing novel anti-neuropathy treatments for humans.

4.3.7 NMA of *R. norvegicus* stand-alone GCH1 structure

NMA of the GCH1 structure from *R. norvegicus* (excluding the GFRP of the inhibitory complex 1WPL) yielded a total of 5514 non-trivial modes. Within the 20 lowest-frequency modes (Table 4.2), four non-degenerate modes were identified (7, 12, 15, and 18), the nature and direction of motions within these modes were further studied. The nature of motion in each mode is shown in Figure 4.6-A.

Table 4.2. Low-frequency normal modes of the stand-alone GCH1, associated eigenvalues, and degeneracy level. The first six trivial modes were excluded, and therefore Mode 1 signifies Mode 7.

Mode	Eigenvalue	Degeneracy
Mode 1	0.49	2
Mode 2	0.50	2
Mode 3	0.50	2
Mode 4	0.51	2
Mode 5	0.59	2
Mode 6	0.60	2
Mode 7	0.66	1
Mode 8	0.75	2
Mode 9	0.76	2
Mode 10	0.79	2
Mode 11	0.79	2
Mode 12	0.80	1
Mode 13	1.02	2
Mode 14	1.02	2
Mode 15	1.03	1

Mode 16	1.07	2
Mode 17	1.09	2
Mode 18	1.27	1
Mode 19	1.44	2
Mode 20	1.44	2

4.3.8 The linear combination of *R. norvegicus* GCH1 degenerate modes

The linear combination of *R. norvegicus* GCH1 degenerate modes of similar eigenvalues within the first 20 non-trivial low-frequency modes produced symmetric global motions such as global wringing, shearing, and shrinking motions (Appendix 2.5). Such motions can be of functional importance. However, these combined degenerate modes had almost zero overlaps to the experimental change observed between the stimulatory and inhibitory conformations of GCH1 (Appendix 2.6). As a result, only the non-degenerate modes for which non-zero overlaps were obtained with the stimulatory conformational state were studied.

4.3.9 Visual characterization of the *R. norvegicus* GCH1 low-frequency non-degenerate normal modes

Mode 7 featured an anti-correlated motion of the GCH1 two pentameric units, driven by the protein N-terminal helices high mobility ([Movie 5](#)). A similar motion to **Mode 7** was observed in **Mode 12**, accompanied by an outward movement of the protein central five-helix bundle ([Movie 6](#)). Modes of notable overlap with the direction of experimental conformational changes can disclose motions of functional importance. The predicted displacement in both modes 7 and 12 compared to the experimentally conformational state of the stand-alone GCH1 from the stimulatory form (PDB ID 1IS7) showed a notable non-zero overlap of 0.1 and 0.4, respectively. In contrast, **Mode 15** and **Mode 18** exhibited a zero overlap (Appendix 2.7). Although the experimental change from the stimulatory to inhibitory conformation appears minor between the GCH1 crystal structures of the stimulatory and inhibitory complex (RMSD of 1 Å), close inspection of the two structures revealed small local changes involving an inward contraction of the central tunnel (Appendix 2.8). We hypothesize that this local contraction of the tunnel, which was captured by **Mode 12** (Figure 4.6-A and [Movie 6](#)),

essentially shifts the state of the complex such that it is poised to function in its inhibitory form and may open dynamic pathways that modulate transition between the two functional forms.

An overall outward expansion involving all toms was observed in **Mode 15**, thus presented a breathing-like motion ([Movie 7](#)). Lastly, **Mode 18** captured a global contraction/narrowing of the protein barrel ([Movie 8](#)). Notably, in their cryo-electron microscopy (cryo-EM) structure determination of an inhibited GCH1, Ebenhoch and colleagues reported the shrinking of the β -barrel radius as well as the narrowing of its central helices (Ebenhoch *et al.*, 2020). In the characterized modes, the GCH1 tunnel gating mechanism presented as the narrowing and widening of its central tunnel was accompanied by the high mobility of the N-terminal helices (bending and wringing motions). The observed high mobility of the N-terminals further demonstrates the active role of the N-terminal helices in modulating the GCH1 tunnel gating mechanism, while the protein core plays an essential role in maintaining the stability of the folded complex. The twist and shear motions of the two pentameric units can describe the pushing or passing of the reactants.

The breathing, twisting, and outward movement of the GCH1 central five-helix bundle can make the tunnel more accessible to the substrate and permit possible physical interactions with its regulatory protein or subsequent proteins in the same pathway to pass the unstable intermediate substrate DHNP. GCH1 is already known to have physical protein partners such as the GFRP, tyrosine hydroxylase, and catecholamines-up proteins (Swick and Kapatos, 2006). Further, the identified global motions of the GCH1 enzyme were found to be similar to those probed in the PTPS enzyme (Khairallah, Ross and Tastan Bishop, 2020) ([Chapter 5](#)). The shared global motions mainly involved regulating the tunnel opening and closing by the N-terminal helices and the shear or breathing motions across the dimeric units that resulted in the upward and inwards movement of the tunnels to enhance accessibility. This can further suggest a shared structural mechanism of function and possible physical interaction of the two enzymes; it also demonstrates that the intrinsic flexibility of proteins is far more conserved than structures and sequences.

4.3.10 MSF analysis of the *R. norvegicus* GCH1 individual non-degenerate slow modes

Due to the homo multimeric nature of GCH1, the MSF profiles of the identified non-degenerate modes were presented from one GCH1 chain (Figure 4.6-B). The MSF of the complete homodecameric structure in each mode was symmetric across all chains except for **Mode 15** (Appendix 2.9). The MSF profiles of **Mode 7** and **Mode 18** displayed the high mobility of N-terminal domain residues R48 to D106, followed by residues surrounding the BH4 binding site and the active site (residues F113 and H134). In **Mode 12**, however, a notable fluctuation was observed in the BH4 binding site residues T220 to R232, the central five-helix bundle, followed by W87, M120 and L148 of the BH4 site. Notably, R232 corresponding to R241 in the human GCH1 was previously reported to form a direct interaction with the effector molecule BH4.

In **Mode 15**, residues located in the N-terminus R48-W87, and E115 of the BH4 binding site loop region, showed the highest fluctuation. In summary, the obtained MSF profile of the selected low-frequency non-degenerate modes identified regions of high flexibility which corresponded to the N-terminal, BH4 binding sites, and C-terminal regions. The identified residues of notable mobility within the non-degenerate slow modes demonstrate their essential contribution toward the global dynamics of the protein and function.

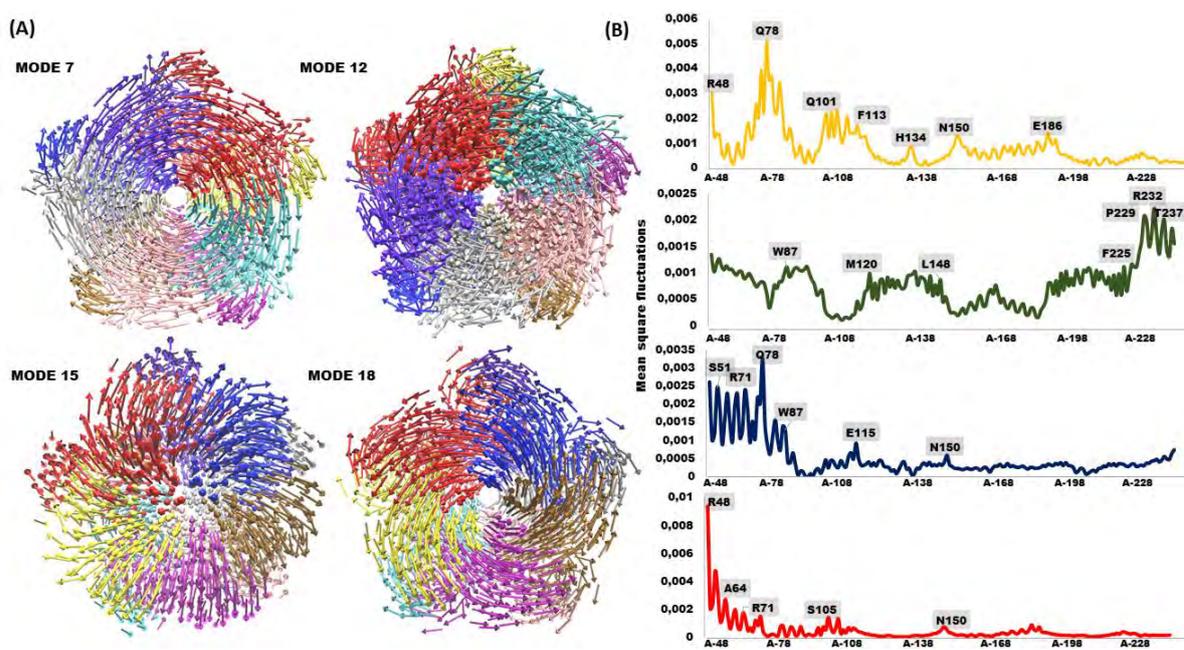


Figure 4.6. (A) The characterized global motions of the *R. norvegicus* stand-alone GCH1 structure within the four non-degenerate low-frequency modes. The obtained eigenvectors were projected as a set of arrows that denote the direction of motion of the protein C β atoms. Each of the protein chains is coloured differently for illustration purposes. (B) Individual MSF profiles of the four non-degenerate low-frequency modes, from top to bottom Mode 7, 12, 15, and 18. Residues of a notable fluctuation are labelled on the MSF plot of the one GCH1 chain. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2021)).

4.3.7 Collective MSF analysis of the *R. norvegicus* GCH1 lowest and highest-frequency normal modes

The collective MSF profile of the 20 highest-frequency modes was compared to the collective MSF profile calculated over the 20 lowest-frequency modes to identify key residues responsible for modulating the protein conformational transition; and high-frequency vibrating residues that underlie the protein stability. The collective MSF profile of the 20 lowest-frequency modes was shown to signify the high mobility of the N-terminal helices (residues R48-Q101) (Figure 4.7-A and Figure 4.7-C). The core of the protein, which consists of anti-parallel beta-sheets, exhibited the least fluctuation.

In contrast, the MSF profile of the 20 highest-frequency modes showed that high-frequency vibrating residues were mostly located within the GCH1 core, more specifically residues involved in the binding of the effector molecule BH4 within the protein core, such as M120, I122, and S219 (Figure 4.7-B and Figure 4.7-D). Residues belonging to the central five-helix bundle helices also showed some fluctuation, such as E234 and L238 (Figure 4.7-D). Overall, the MSF of the 20 highest-frequency normal modes disclosed hotspot residues that can be crucial for the stability and function of the protein, whereas the MSF of the 20 lowest-frequency normal modes disclosed key residues involved in driving the protein conformational transition. Overall, the collective MSF profiles of *R. norvegicus* GCH1 were similar to those obtained from the *P. falciparum* GCH1 structure, in which the high flexibility of the N-terminal regions and rigidity of the protein core was confirmed in both cases.

In an attempt to evaluate the precision of the characterized atomic fluctuation by the NMA and further confirm that the high mobility of the N-terminus is not a result of tip effects (given their location at the edge of the protein structure), the MSF profile of all normal modes was plotted against the crystal structure experimentally determined fluctuation, the so-called *B*-factor. The MSF values agreed with the *B*-factor (Appendix 2.10); thus, we can deduce the reliability of the ENM model atomic fluctuation.

The calculated collective MSF of the 20 lowest-frequency normal modes were also mapped onto the *R. norvegicus* GCH1 structure and qualitatively compared to the experimental *B*-factor (Figure 4.7-E and 4.7-F). The fluctuations colour scale is from blue to red (High/ flexible to low/rigid) atomic fluctuation. The absolute atomic motion over the 20 lowest-frequency modes recognised the N-terminal helices as the most mobile (red) regions of the GCH1 structure, which was also confirmed by the experimental *B*-factor (Appendix 2.10). Hinge residues were identified within the MSF profile of all modes. The identified hinge residues were located within the protein core (Appendix 2.10).

4.3.8 Deformation energy analysis of the *R. norvegicus* stand-alone GCH1 structure

The deformation energy analysis of the *R. norvegicus* stand-alone GCH1 structure revealed a build-up of deformation energy in the protein central region/core during its large global motions, indicating the existence of an energy barrier that reduced its deformation. On the other hand, lower deformation energy was observed in the N-terminal helices, BH4, and GFRP binding regions, demonstrating their high degree of collective motion (Figure 4.7-G).

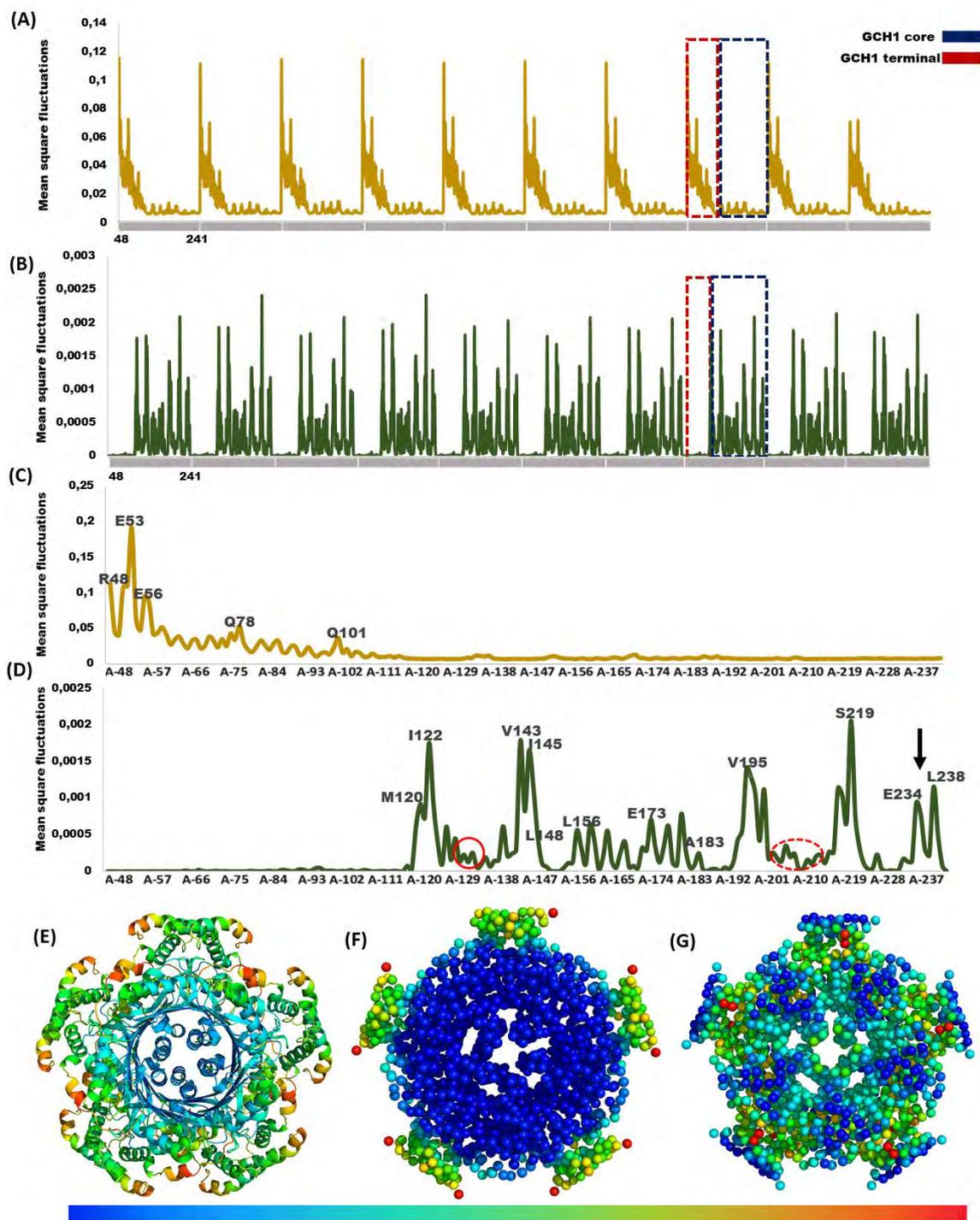


Figure 4.7. Collective MSF profile of the (A) 20 lowest-frequency and (B) 20 highest-frequency normal modes. The grey blocks partition each chain of the GCH1 homo-decameric 3D structure. One chain representation of the *R. norvegicus* stand-alone GCH1 of the (C) 20 lowest-frequency and (D) 20 highest-frequency MSF profiles. The solid red circle highlights the active site catalytic residues, and the dashed red circle highlights neighbouring residues that form the active site pocket. The black arrow points towards the central five-helix bundle region. (E) Experimental *B*-factor of the *R. norvegicus* GCH1 crystal structure (PDB ID:1WPL). (F) MSF values over the 20 lowest-frequency modes mapped onto the *R. norvegicus* GCH1 structure. The fluctuations colour scale is from blue to

red (low to high) atomic fluctuation. (G) Atomic deformation energies: the structure is coloured such that atoms of the lowest deformation energy are in blue, and atoms of highest deformation energy are in red. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2021)).

4.3.9 Mechanical stiffness analysis

The anisotropic response of the GCH1 structure to external perturbations was evaluated via the mechanical stiffness analysis, which has allowed for identifying weak/strong pairs of interactions and sites susceptible to deformation. From the mechanical resistance map, we can deduce that residues belonging to the GCH1 N-terminal helices (residues R48 to Q80) as well as residues involved in the regulatory protein binding (residues E173-L187) were prone to deformation, as they exhibited weak pairs of interactions with the remaining residues of the structure. The mechanical stiffness analysis also disclosed the weak pairs of interactions between the active site region, BH4 binding site, and the N-terminal regions (Figure 4.8-A). This is further supported by previous studies in which notable structural changes in regions forming the active site pocket, including (residues M204–N213), were linked to changes in the BH4 binding site and the N-terminus (Maita *et al.*, 2004). As expected, and further supported by previous deformation energy analysis, residues within the protein core were more resistant to deformation. Figure 4.8-B shows the mean resistance of each residue averaged over all pairs, demonstrating a strong mechanical resistance within the protein core and flexibility of its terminus.

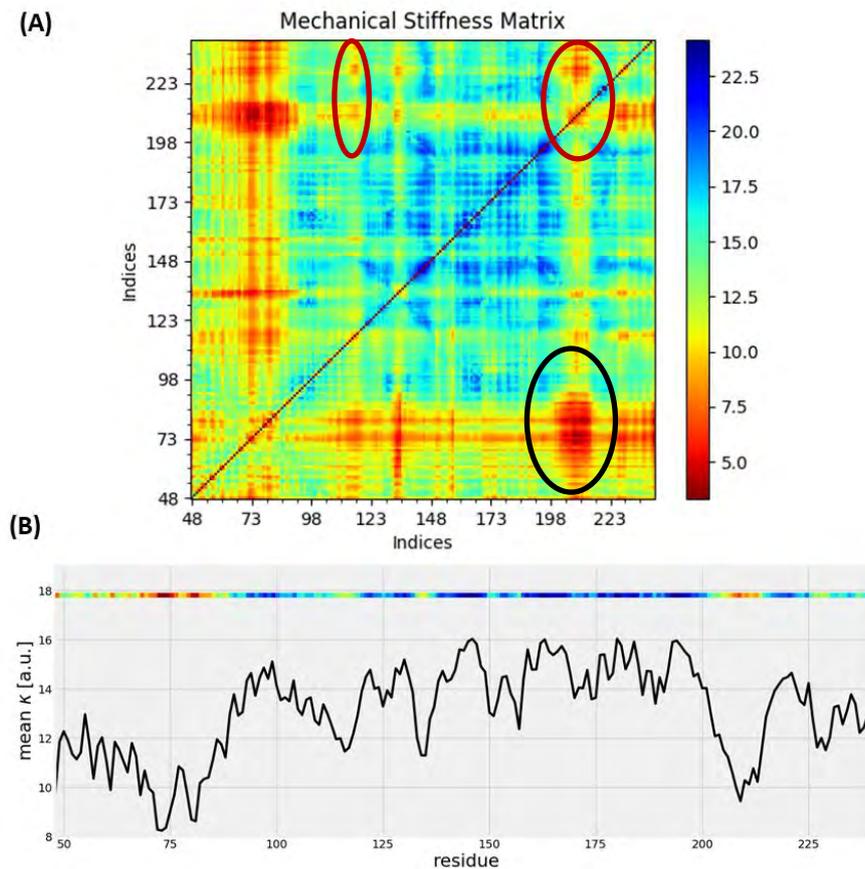


Figure 4.8. (A) ProDy mechanical stiffness heatmap of the *R. norvegicus* stand-alone GCH1. The colour bar illustrates the effective force constant values (blue as a strong /stiff pair of interactions and red as a weak/flexible pair of interactions). (B) The mean value of the effective force constant of residue pairs, the GCH1 secondary structure, is shown along the upper abscissa. The mechanically stiff/rigid regions are shown in blue, representing the GCH1 20 stranded β -barrel. The black circle highlights the residues participating in forming the active site pocket, whereas the red circles indicate the residues of the BH4 binding site. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2021)).

4.3.10 NMA of GCH1 in complex with its feedback regulatory protein

NMA analysis was performed for the GCH1-GFRP complex in both inhibitory and stimulatory forms, 1WPL and 1IS7, respectively. A total of 8334 non-trivial modes were obtained from each complex. Due to the large size of the protein complex and the extensive degeneracy of the normal modes, the 1% lowest-frequency modes were studied. Within the 1% lowest-frequency modes, six non-degenerate normal modes were characterized in the inhibitory complex, 1WPL, and four non-degenerates in the stimulatory complex, 1IS7 (Table 4.3). The large-scale motions within the selected low-frequency non-degenerate modes of the GCH1 inhibitory complex, 1WPL, and stimulatory complex, 1IS7, were shown to be similar; this was further supported by an RMSIP value of 0.7 (Appendix 2.11). Hence, the non-degenerate low-frequency modes of the GCH1 inhibitory complex were only studied and presented here.

Table 4.3. Low-frequency normal modes of the stand-alone GCH1, GCH1-GFRP inhibitory complex, 1WPL and GCH1-GFRP stimulatory, 1IS7 and their associated eigenvalues, and degeneracy level. The first six trivial modes were excluded, and therefore Mode 1 represents Mode 7. The subscripted IC and SC stand for GCH1-GFRP inhibitory complex and stimulatory complex, respectively.

GCH1-GFRP (Inhibitory, 1WPL)			GCH1-GFRP (Stimulatory, 1IS7)		
Mode	Eigenvalue	Degeneracy	Mode	Eigenvalue	Degeneracy
Mode _{IC} 1	0.17	2	Mode _{SC} 1	0.16	2
Mode _{IC} 2	0.17	2	Mode _{SC} 2	0.17	2
Mode _{IC} 3	0.23	1	Mode _{SC} 3	0.20	1
Mode _{IC} 4	0.31	1	Mode _{SC} 4	0.26	1
Mode _{IC} 5	0.39	2	Mode _{SC} 5	0.37	2
Mode _{IC} 6	0.39	2	Mode _{SC} 6	0.37	2
Mode _{IC} 7	0.42	1	Mode _{SC} 7	0.40	1
Mode _{IC} 8	0.48	2	Mode _{SC} 8	0.50	2
Mode _{IC} 9	0.48	2	Mode _{SC} 9	0.51	2
Mode _{IC} 10	0.51	2	Mode _{SC} 10	0.54	2
Mode _{IC} 11	0.51	2	Mode _{SC} 11	0.56	2
Mode _{IC} 12	0.58	2	Mode _{SC} 12	0.58	2
Mode _{IC} 13	0.58	2	Mode _{SC} 13	0.58	2
Mode _{IC} 14	0.60	1	Mode _{SC} 14	0.61	2
Mode _{IC} 15	0.63	2	Mode _{SC} 15	0.62	2

Mode_{IC} 16	0.63	2	Modesc 16	0.64	1
Mode_{IC} 17	0.67	2	Modesc 17	0.66	2
Mode_{IC} 18	0.67	2	Modesc 18	0.67	2
Mode_{IC} 19	0.72	2	Modesc 19	0.75	2
Mode_{IC} 20	0.72	2	Modesc 20	0.77	2
Mode_{IC} 26	1.22	1	Modesc 26	1.14	2
Mode_{IC} 77	2.59	1	Modesc 77	2.50	2

The nature, direction and fluctuation profile of the GCH1-GFRP complex motions in each mode are illustrated in Figure 4.9. **Mode_{IC} 3** featured a prominent anti-correlated rotation of the GCH1 and GFRP proteins ([Movie 9](#)). **Mode_{IC} 4** captured a stretch-like motion, in which the two proteins were shown to pull away from each other ([Movie 10](#)). The identified stretch motion also resulted in the shrinking of the GCH1 tunnel and may explain the dissociation of the two proteins to afford the inhibitory form. In **Mode_{IC} 7**, a double plane rotation involving the contra-rotation of the GFRP and GCH1 counter-clockwise rotation (in a different plane) was observed. The rotation was accompanied by an expansion of the GCH1 tunnel that appeared to be driven by the wringing of the GCH1 N-terminal helices ([Movie 11](#)). Notably, a rotation of the GCH1 in the presence of the GFRP was reported by Maita and colleagues (Maita *et al.*, 2004). **Mode_{IC} 12** captured a tilt-and-twist motion of the GFRP, accompanied by the twisting of GCH1 N-terminal helices, allowing the GCH1 to contract into closer proximity to GRRP ([Movie 12](#)). **Mode_{IC} 26** and **Mode_{IC} 77** captured engulf-like motions in which the GCH1 undergoes conformational changes accompanied by the GFRP inward movements ([Movie 13](#) and [Movie 14](#)). These two modes showed a notable overlap with the stimulatory complex crystal structure in the modes overlap analysis; hence they were visualized and found to present distinctive global motions.

Overall, the six non-degenerate modes described the coupling of the two proteins and further explained their structural dynamics, in which changes around the GCH1 tunnel were mostly driven by the protein N-terminal helices movements and accompanied by the arrival of the GFRP at proximity. Furthermore, the GCH1 central five-helix bundle exhibited high fluctuation in an upward direction, making the tunnel more accessible.

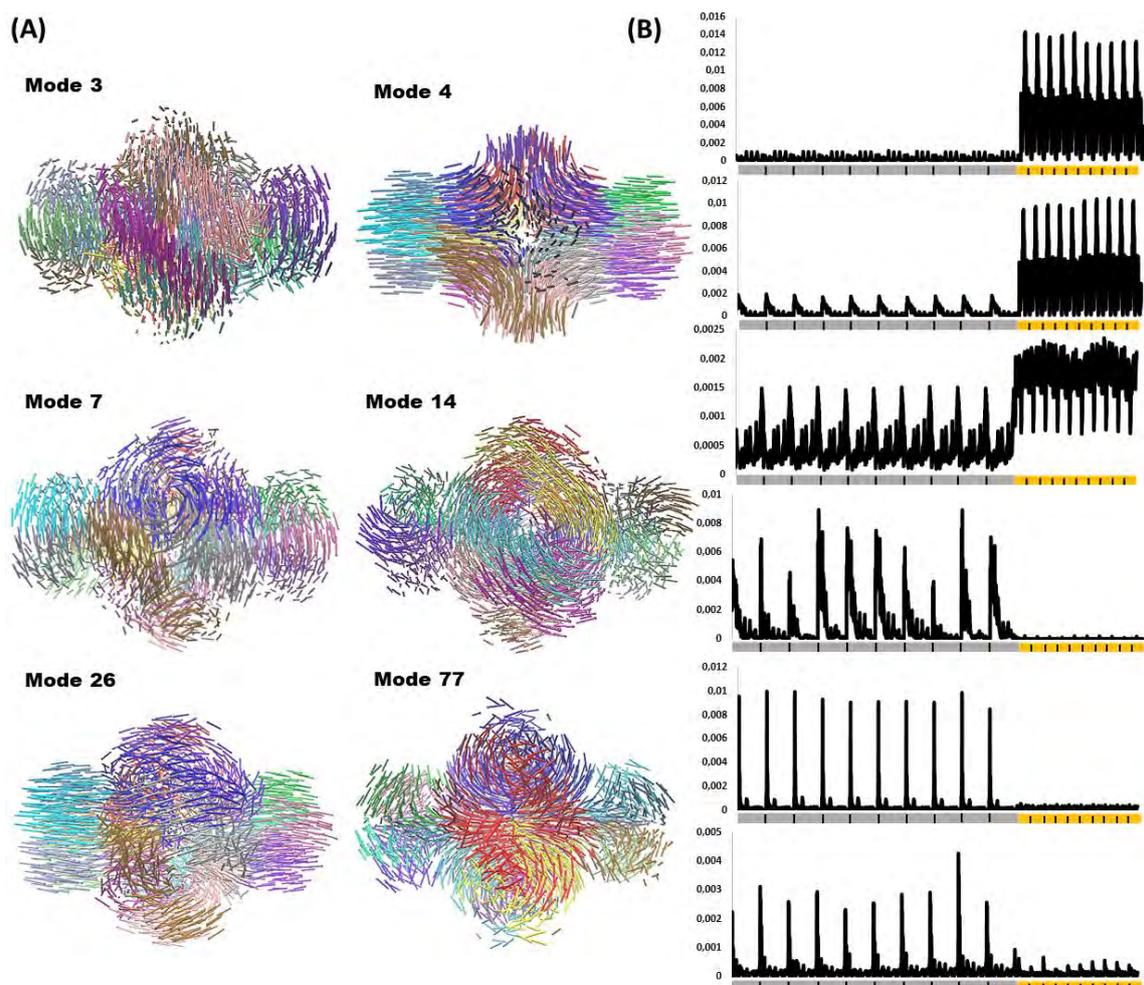


Figure 4.9. (A) The characterized global motions of the *R. norvegicus* GCH1-GFRP structure (PDB ID: 1WPL). Each arrow denotes the protein C_{β} atoms vector of motion. The obtained eigenvectors were projected as a set of arrows that indicate the direction of movement of the protein C_{β} atoms. Each of the protein chains is coloured differently for illustration purposes. (B) Individual MSF profiles of the six non-degenerate low-frequency modes, from top to bottom Mode 3, 4, 7, 12, 26, and 77. The grey blocks represent each chain of the GCH1 structure, and the yellow blocks represent each chain of the GFRP. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2021)).

4.3.11 Mode overlap analysis

First, the crystal structures of the GCH1-GFRP inhibitory and stimulatory complexes were superimposed to study the experimental structural change between the two complexes. The result showed a slight inward shift (from the stimulatory towards the inhibitory form), more evident at the GFRP loop regions, the tip of the central helices and the last β strand of the GCH1 (Figure 4.10 and Appendix 2.8). However, the observed structural changes were relatively minute and further quantified by a root mean square deviation (RMSD) of 1.32 Å. Based on static crystal structures, one

may suggest slight or no structural changes between the two complexes. Thus, it was essential to consider the structural dynamics whilst studying these complexes.

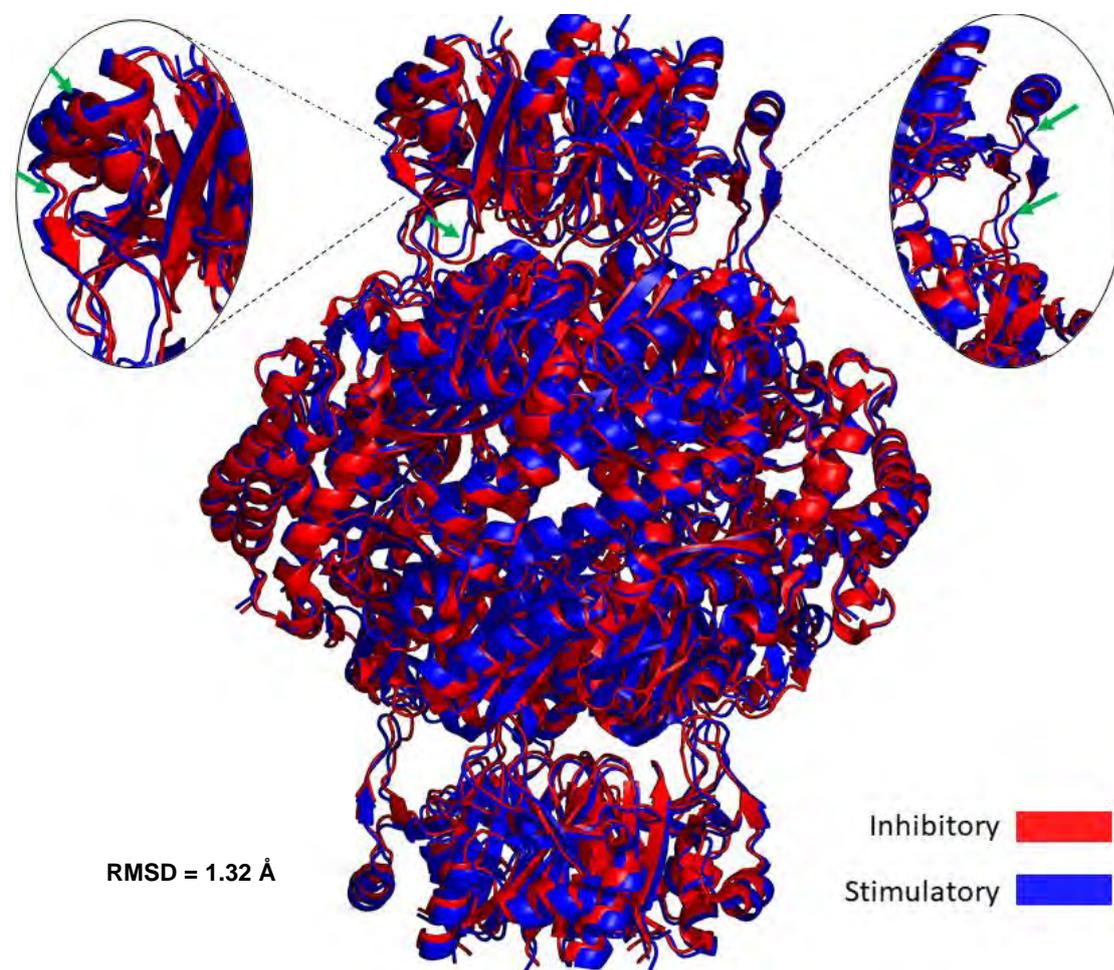


Figure 4.10. Superimposition of the *R. norvegicus* GCH1-GFRP inhibitory complex (PDB ID:1WPL) and the *R. norvegicus* GCH1-GFRP stimulatory complex (PDB ID: 1IS7). The green arrows highlight structural differences, which are noted by an inward contraction of the protein complex as it transitions from stimulatory to inhibitory form. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2021)).

The modes overlap analysis was then carried out to quantify the overlap between the predicted displacement of all modes in the inhibitory complex, 1WPL, and the experimental conformational change, in this case, GCH1-GFRP stimulatory complex crystal structure 1IS7. The overlapping mode analysis captured a shift from the inhibitory to stimulatory form presented by a non-zero overlap from four non-degenerate modes (modes 3, 7, 14, and 77) (Table 4.4). Thus, the identified non-degenerate modes captured an overlap towards an experimental conformational shift of the stimulatory form

(PDB ID: 1IIS7). It is important to mention that even though the structural changes between the two complexes were small, these normal modes still captured an overlap toward the experimental conformational state and can be of functional relevance by driving conformational shifts. Furthermore, the overlapping non-degenerate such as Mode_{IC} 3 and Mode_{IC} 77 captured a local change involving the loop regions between the GFRP secondary structures and the contraction of the GCH1 barrel ([Movie 9](#) and [Movie 14](#)). This further validates the predicted motions in these modes.

Table 4.4. The overlap between the predicted displacement of the inhibitory complex (PDB ID: 1WPL) in the normal modes and the experimental conformation of the stimulatory complex (PDB ID: 1IIS7) crystal structure. Notably, the overlap values within the identified four modes were rather small and can be reasoned to the large size of the complex and its extensive degeneracy.

Mode	Overlap
Mode 3	0.31
Mode 77	0.30
Mode 7	0.23
Mode 14	0.19

Ebenhoch and colleagues reported structural changes at the GCH1 central five-helix bundle and its barrel (Ebenhoch *et al.*, 2020). The barrel/tunnel was more contracted in the inhibitory form, confirming the inward shift from stimulatory to inhibitory. Although structural differences were illustrated between static structures of the inhibitory and stimulatory complexes, it is still not understood how the physical association of the two proteins occurs or changes as the complex shifts from inhibiting to stimulating the function. In the next section, the normal modes of the inhibitory complex were compared to those obtained from the stimulatory complex in an attempt to explain the structural mechanism in which the GCH1 regulation can occur and reveal key sites that are responsible for inducing structural changes perceived from the NMA MSF profiles and deformation energy analysis of the two individual complexes.

4.3.12 Collective MSF analysis of the *R. norvegicus* GCH1-GFRP complexes

The collective MSF profiles over the 1% lowest and highest frequency normal modes of the inhibitory and stimulatory complexes were individually assessed and compared to each other (Figure 4.11-B).

This was done to disclose the structural difference and highlight sites that can modulate the transitions between the two complexes. The obtained MSF values over the 1% lowest-frequency modes from both complexes were also mapped onto each structure to illustrate regions of notable mobility (Figure 4.11-A), then compared to the experimental structural data, which was shown to agree.

Within the 1% lowest-frequency MSF profile, the GCH1 N-terminal helices were slightly higher in the inhibitory complex. Notably, the modal analysis showed that the wringing motions of these N-terminal helices resulted in the contraction of the protein core to expel the GFRP protein and form the inhibitory state of the complex (Figure 4.11-C). Furthermore, previous studies reported that the deletion of the human GCH1 N-terminal domain caused a modest increase in the enzyme activity relative to the full-length protein; thus, it was suggested that the N-terminal helices function as an autoinhibitory control element (Higgins and Gross, 2011). The GFRP residues were found to fluctuate more in the stimulatory form (Figure 4.11-B), more specifically, for residues located within the GFRP phenylalanine binding sites and loop regions between its secondary structures.

The deformation energy analysis revealed the large deformation of the GCH1 core involving the active site pocket and BH4 binding site. The GCH1 N-terminal helices exhibited less deformation demonstrating their high mobility and collective motion. Furthermore, the interface region formed between the GCH1 and GFRP exhibited high deformation, which was also shown to vary among the two complexes. The GFRP loop regions positioned toward the GCH1 and reported to make contact with the protein were also identified among the most deformed regions (Figure 4.11-A).

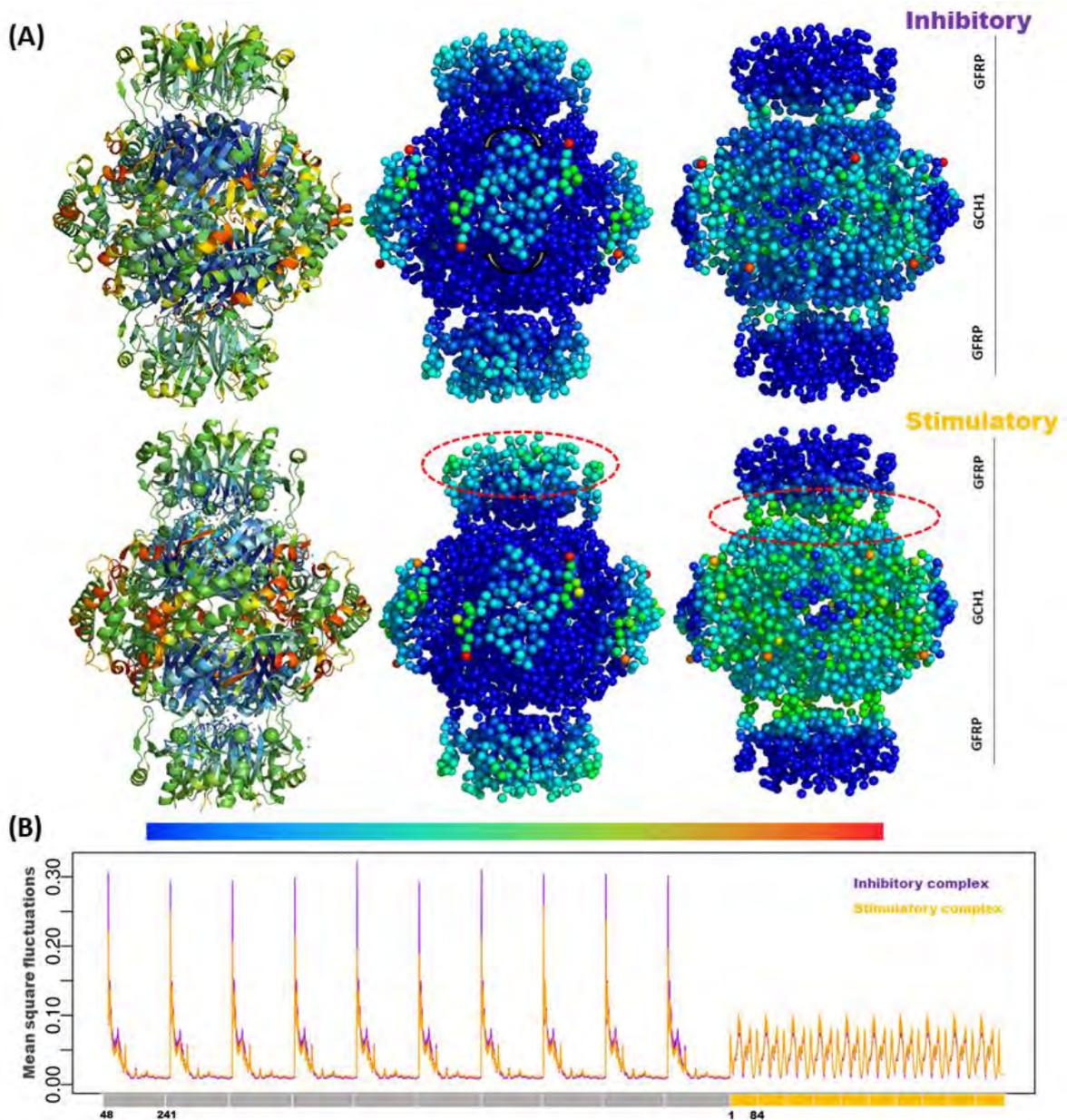


Figure 4.11. (A) From left to right: the experimental B -factors of the GCH1-GFRP crystal structure of the *R. norvegicus*, atomic fluctuation within the 1% lowest-frequency modes: regions of the lowest MSF are shown in blue. Higher MSF in green, yellow, and red and the atomic deformation energies mapped onto the structure and coloured such that the most deformed regions are in red, and the least deformed are in blue. The arrows show the wringing motion of the terminal helices. (B) MSF profiles over the 1% lowest-frequency modes of both complexes. The grey blocks partition each chain of the GCH1 homo-decameric 3D structure, and the orange blocks partition each chain of the GFRP protein. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2021)).

Due to the homomultimeric nature of both GCH1 and GFRP proteins, the protein's monomeric units exhibited similar fluctuation. Thus, the MSF profiles over the 1% lowest and highest-frequency modes from one chain of the two proteins were extracted to comprehensively study the residues fluctuation (Figure 4.12-B and Figure 4.12-C). In the MSF profile of the 1% lowest-frequency modes, the main differences were identified within the GCH1 N-terminal helices (residues R48 to Q10), which was presented by a notable fluctuation and slight variation among the two complexes (Figure 4.12-B). The GCH1 loop regions highlighted in a green box (Figure 4.12-B) showed restricted fluctuation in both complexes, suggesting their role as hinges of collective motions. Figure 4.12-A shows the location of BH4 and Phe binding sites and their surrounding residues; most of these residues were found among the highly fluctuating residues within the 1% highest-frequency modes, which will be discussed next.

The collective MSF profile within the highest-frequency 1% modes disclosed high fluctuating residues within the GCH1 core; this includes the active site, BH4, and GFRP binding sites in addition to the GFRP phenylalanine binding site (Figure 4.12-C). In the inhibitory form, residues within the BH4 binding region, including (residues T220–E234) and its flanking regions of (residues D118–I122) and (residues V143–N150), exhibited higher fluctuation in the inhibitory complex (Figure 4.12-C). This further suggests the importance of these residues in inducing the GCH1 inhibitory form. On the other hand, the GCH1 catalytic active site residues and its flanking residues, including H134, H135, C132, Q171, and K215, GFRP binding site (residues E173- L187), and GFRP loop β 1- β 2 (residues Q9–P16) exhibited more fluctuation in the stimulatory complex (Figure 4.12-C). These residues were reported for their crucial role in stabilizing the protein and substrate binding (Maita *et al.*, 2004).

V182 of the GCH1 GFRP binding site was also reported to form hydrophobic interactions with the GFRP L40 and P190 to interact with the GFRP N43; these two residues were also found among the high-frequency vibrating residues as shown in Figure 4.12-C. Mutagenesis study involving V182 and P190 was shown to disturb the GFRP mediated feedback mechanisms of GCH1 (Maita *et al.*, 2004). Maita and collages also reported local displacements in the GFRP loop regions β 1- β 2 (residue Q9-

P16) and β 3- β 4 (residues L73–Q75) of the stimulatory complex (Maita *et al.*, 2002), these regions relatively exhibited more fluctuation in the stimulatory form (Figure 4.12-C).

The two residues of the human GCH1 R235 and R241 were reported as key residues that change position to enable the binding of BH4, thus induce a shrinking of the protein central five-helix bundle and a conformational rearrangement of the β -barrel (Ebenhoch *et al.*, 2020). The two residues correspond to R226 and R232 of the *R. norvegicus* GCH1 (Figure 4.12-C). Furthermore, site-directed mutagenesis involving BH4 inhibitor binding residues H126, D127, E128, corresponding to H117, D118, E119 in the *R. norvegicus* structure as well as R235, and R241, showed that all GCH1 mutants except for R235 and R241 were still functional or exhibited a slight reduction in activity relative to the wild type (Ebenhoch *et al.*, 2020). In the MSF profile of both complexes, R226 and R232 showed minimal fluctuations; however, their surrounding residues, such as S219 and E234, presented notable high fluctuation in the inhibitory complex (Figure 4.12-C). Lastly, H117, D118, E119 were reported to trigger structural changes of residues located on the path to the active site (Ebenhoch *et al.*, 2020). Within the MSF profile of the 20 highest-frequency modes, these residues showed a relatively higher fluctuation as well as I122 in the inhibitory form. Overall, using the NMA approach, high-frequency vibrating residues of the GCH-GFRP complexes were successfully identified. Such residues can be regarded as kinetically hot residues/hotspots essential for protein stability and function. Furthermore, local differences between the two complexes were illustrated, highlighting which residues were necessary for the modulation/ formation of a particular complex form.

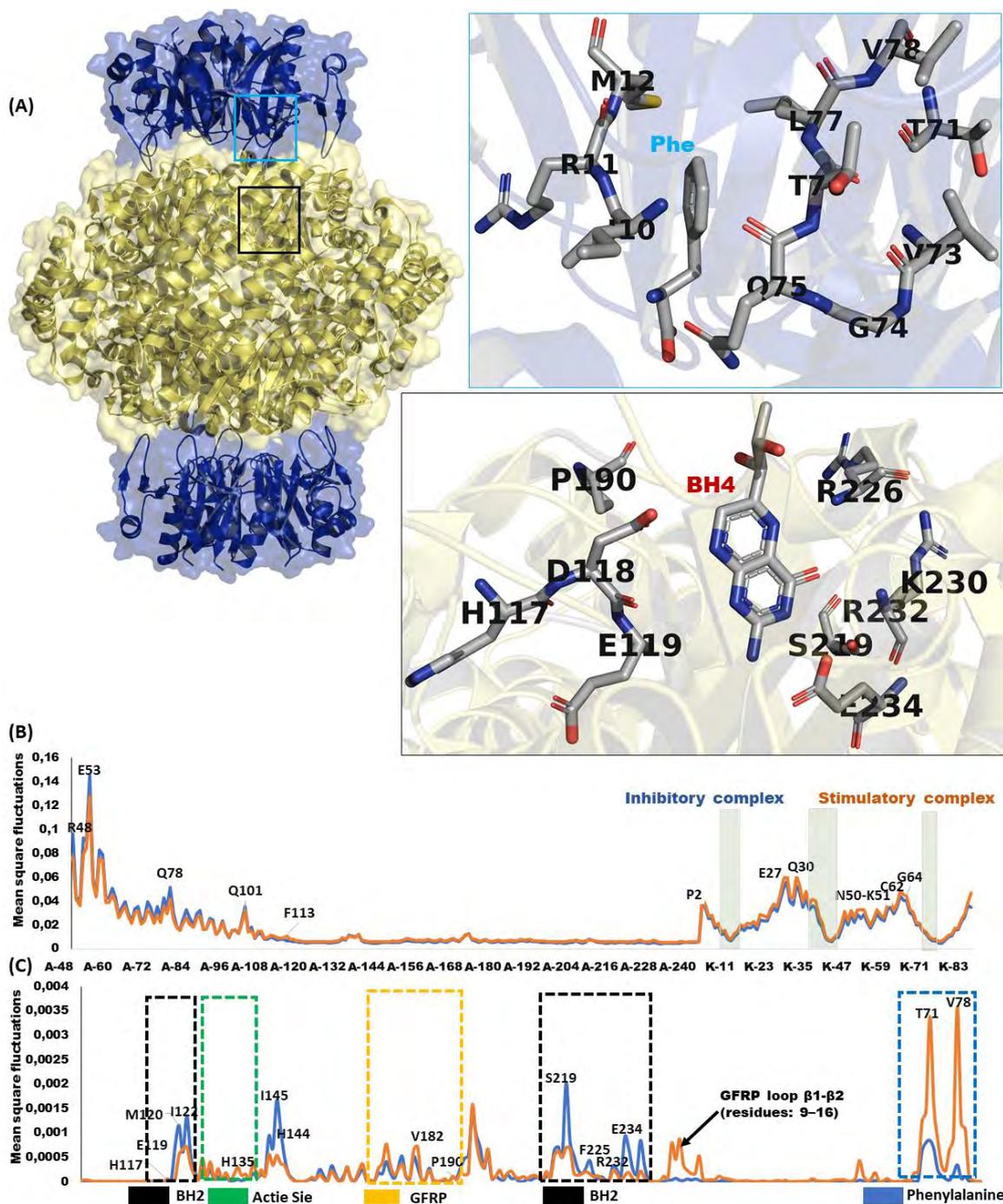


Figure 4.12. (A) The structure of the GCH1-GFRP stimulatory complex from *R. norvegicus* (PDB ID: 1IS7), the blue zoomed-out box, shows the GFRP phenylalanine (Phe) binding site residues, and the zoomed out black box shows the GCH1 BH4 binding site residues. (B) MSF profiles over the 1% lowest-frequency normal modes for GCH1-GFRP inhibitory (blue) and stimulatory (orange) complexes. (C) MSF profiles over the 1% highest-frequency normal modes for GCH1-GFRP inhibitory (blue) and stimulatory (orange) complexes. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2021)).

4.3.13 GCH1-GFRP residues cross-correlation analysis

The residue cross-correlation of the *R. norvegicus* GCH1-GFRP inhibitory complex over the 1% non-trivial lowest-frequency modes was calculated and plotted as a heat map (Figure 4.13-A). The heatmap showed a strong cross-correlation along the diagonal line, illustrating a correlation of neighbouring atoms within the same chain. A positive correlation was observed between atoms of chain A and E within the same pentameric unit and across the two pentameric units, specifically, chain A of one pentamer and chain E of the second pentamer. The observed residue cross-correlation of the *R. norvegicus* GCH1 was similar to the *P. falciparum* GCH1 (Figure 4.5). However, in Figure 4.13, particularly interesting are the blue regions illustrating that the GCH1 and GFRP proteins move in an opposing direction (Figure 4.13-A). The observed anti-correlation between the two proteins can demonstrate the rotation of GCH1 around its 5-fold axes relative to the GFRP while moving closer to it in screw-like motion (Figure 4.13-B). The insertion of the GFRP into GCH1 can be related to GCH1 tunnel expansion. Thus, the dissociation of the two proteins can be explained by the GCH1 tunnel contraction upon binding to BH4 to disengage the GFRP, hence forming the inhibitory complex (Figure 4.13-C). This finding is further supported by a recent experimental study involving a comparison of cryo-electron microscopy of both inhibitory and stimulatory GCH1–GFRP complexes; the results showed an increase in the density of the electron microscopy map at the core of the protein core involving the β -barrel and the central five-helix bundle of the inhibitory complex (Ebenhoch *et al.*, 2020).

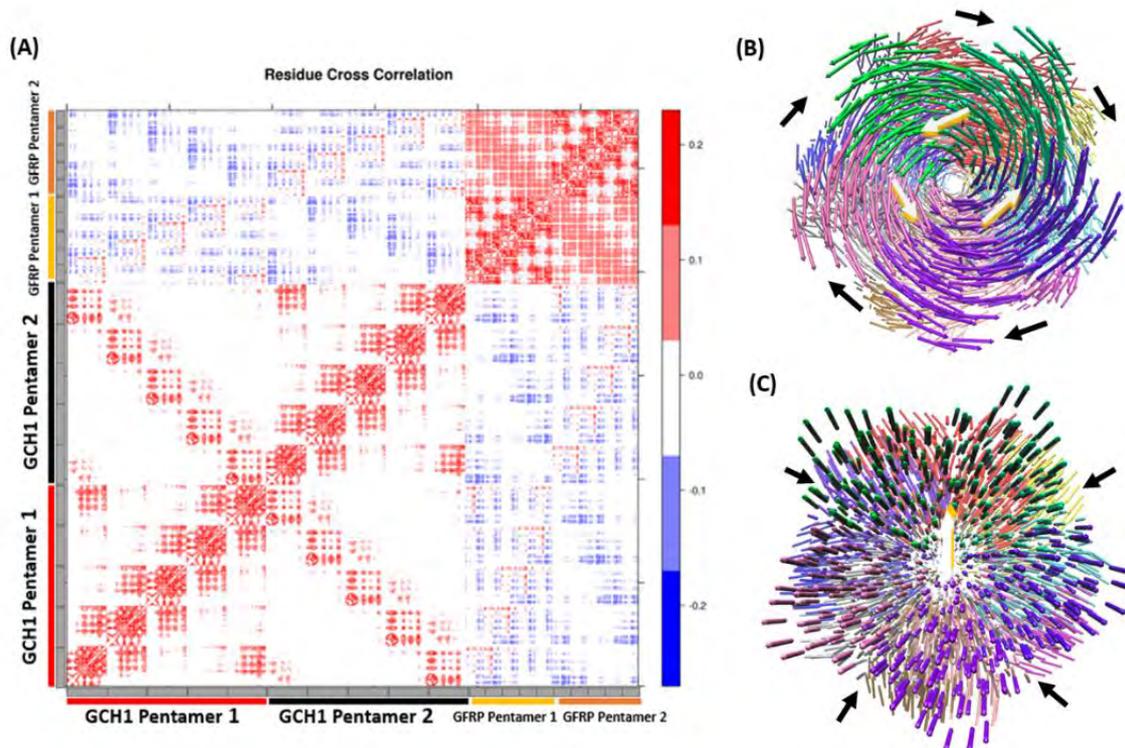


Figure 4.13. (A) A heatmap representation of the GCH1-GFRP complex residue pair-wise cross-correlation heatmap. The scale colour bar on the right indicates the extent of the correlation in which the red colour highlights correlated motions (residue pairs moving together in the same direction), while the blue colour highlights the anti-correlated movements (residue pairs moving in the opposite direction). The high diagonal values are occurring where $i = j$ shows the correlation within the same atom. The grey blocks partition each chain of the proteins. (B) The rotation of GCH1 in an opposing direction to the GFRP as identified in **Mode_{IC} 3**. (C) The inward contraction of GCH1 and the outward expansion/expelling of the GFRP as identified in **Mode_{IC} 4**. The black arrows show the GCH1 direction of motion, and the white arrows represent the GFRP direction of movement. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2021)).

4.4 Conclusions

GCH1 is an attractive drug target for treating infectious diseases such as malaria and BH4 deficiency-related diseases in humans. The enzyme has a unique structure and sequence features, e.g. its distinct heterogenous N-terminus and central tunnel/cavity, which can be exploited to discover allosteric drugs of a novel mechanism of action. In this chapter, NMA was employed to gain a detailed insight into the GCH1 structural dynamics with low computational cost and time, which could not be feasible experimentally or via conventional simulation approaches. The NMA disclosed essential structural information about the protein intrinsic dynamic and mechanism of allosteric modulation, further highlighted regions that govern the protein conformational changes such as the N-terminal regions, and explained the coupling mechanism between the mammalian GCH1 and its regulatory partner GFRP.

NMA of the *P. falciparum* GCH1 structures elucidated the structural dynamics of the protein and disclosed essential residues for the conformational changes; hotspot residues were identified within the high-frequency profile and were found to be unique to the malaria parasite *P. falciparum* GCH1. The hotspot residues of high-frequency vibration were located within the protein core at key sites such as the effector molecule BH4 and GFRP binding sites and the central five-helix bundle. In a total of four low-frequency non-degenerate modes, large collective motions of the *P. falciparum* GCH1 were characterized, including global breathing, twist, and shear motions that made the GCH1 central cavity/tunnel more accessible. The characterized structural features and hotspot residues can be regarded as potential allosteric drug targeting sites of value in developing allosteric drugs. In the second part of the study, NMA of the *R. norvegicus* stand-alone GCH1 elucidated the structural dynamics of the protein and disclosed hotspot residues within the protein core. A build-up of deformation energy was also observed within the *P. falciparum* GCH1 and *R. norvegicus* stand-alone GCH1, thus illustrating its conserved high rigidity. The N-terminals, however, were less rigid and exhibited high mobility.

The N-terminal helices were found to exert an allosteric control over the GCH1 tunnel gating and play a key role in the association of *R. norvegicus* GCH1 with its regulatory protein, this was characterized within six distinctive low-frequency non-degenerate modes. Local differences between the *R. norvegicus* GCH1 inhibitory and stimulatory complex were also identified within the 20 highest-frequency normal modes. Such sites may be viewed as essential for the formation/transition to each complex.

In summary, based on the NMA, the structural dynamics of the GCH1 enzyme was elucidated, and sites of structural significance for GCH1 allosteric modulation were identified. Targeting the identified key regions either by blocking the tunnel's expansion or inhibiting its terminus's wringing motion may exert allosteric effects and restrict certain conformational states. This finding may provide a solid starting point to design novel novel-antimalarial drugs against the malaria parasite *P. falciparum* GCH1 enzyme to overcome the current antifolate drug resistance and offer a new generation of antifolates as well as anti-neuropathic treatment for humans.

Chapter 5

Normal Mode Analysis of the Tunnelling Fold enzyme 6-Pyruvoyl Tetrahydropterin Synthase Using Elastic Network Models

This chapter describes the normal mode analysis, calculated on the anisotropic network model of the *Plasmodium falciparum* 6-pyruvoyl tetrahydropterin synthase enzyme, to elucidate large-scale collective motions that modulate its function and identify potential allosteric drug targeting sites.

Chapter 5 is reproduced in part with permission from the following publications:

Probing the Structural Dynamics of the *Plasmodium falciparum* Tunnelling-Fold Enzyme 6-Pyruvoyl Tetrahydropterin Synthase to Reveal Allosteric Drug Targeting Sites. [Afrah Khairallah](#), Caroline J. Ross, and Özlem Tastan Bishop. *Frontiers in Molecular Biosciences*. 2020. 7, 1-17. DOI: 10.3389/fmolb.2020.575196.

Authors Contributions: Ö.T.B. conceived the project. A.K. performed the calculations and analysis of data under the guidance of C.R. and Ö.T.B.

Chapter overview

The second enzyme of the malaria parasite *de novo* folate biosynthesis pathway 6-pyruvyl tetrahydropterin synthase (PTPS) presents as a potential alternative target that can be studied to develop new antimalarial antifolate drugs of a novel mechanism of action. This chapter examined the *Plasmodium falciparum* PTPS structural dynamic using normal mode analysis (NMA) based on an elastic network model (ENM) to elucidate its structural dynamics and pinpoint key sites for the allosteric modulation of its binding properties, including hinge sites. The NMA analysis was further combined with motif analysis to identify unique sequence patterns within the *Plasmodium falciparum* PTPS structure and detect their contribution towards the enzyme structural dynamics. Based on the NMA, key sites were identified to play an essential role in the PTPS tunnel gating and substrate binding. The characterized sites include the PTPS N-terminal and central helices, which regulate the large-scale protein motion. Hotspot residues of rapid local vibration were identified within the protein core, specifically residues of the active site and surrounding region. Furthermore, unique sequence motifs were found within regions of structural significance. The findings provide key guidelines towards selective inhibition of the Plasmodium PTPS enzyme by providing a solid starting point for the drug design of allosteric modulators that can be more specific and less toxic to the human host.

5.1 Introduction

Protein structures undergo large-scale conformational changes allowing them to convey functions (Grant, Gorfe and McCammon, 2010; Haspel *et al.*, 2010; Orellana, 2019). Structural changes can be induced by several biological events, including ligands recognition/binding, catalysis, signal transduction, or allostery (Grant, Gorfe and McCammon, 2010; Fenwick, Esteban-Martín and Salvatella, 2011; Sacquin-Mora, 2018). The large-scale conformational changes are therefore linked to function; hence, studying the structural dynamics of proteins is important to disclose key sites that modulate function. Several experimental and computational approaches were developed to study the structure, dynamics, and function of proteins (details about the different approaches are presented in sections 4.1.3 and 4.1.4 of [Chapter 4](#)).

Normal mode analysis (NMA) has become a popular computational approach that facilitates speedy and systematic identification of protein structural dynamics and potential allosteric sites (Bauer, Pavlović and Bauerová-Hlinková, 2019). Various studies have demonstrated that the results obtained from NMA can characterize the dynamical aspects of proteins (details about NMA are presented in section 4.1.4 of [Chapter 4](#)). The slow modes from the NMA (modes of low frequency) account for global collective motions while fast modes (modes of high frequency) describe rapid and local motions that correspond to kinetically hot residues; such residues are instrumental in sustaining the stability and function of proteins (Bahar, Lezon, Yang, *et al.*, 2010). As a result, the information deduced from low and high-frequency normal modes can be of great value in revealing key sites modulating the protein structure and function.

NMA of the *P. falciparum* PTPS

Due to the current antimalarial antifolate drug resistance, identifying new drugs and drug targets has become crucial. 6-pyruvol tetrahydropterin synthase (PTPS) is the second enzyme of the malaria parasite *de novo* folate biosynthesis pathway; the enzyme has never been targeted before and presents as an attractive drug target (details about the PTPS enzyme are given in section 1.1.10 of [Chapter 1](#)). Studying the structural dynamics of this protein can be of great value for identifying new allosteric drug targeting sites for malaria treatment. Allosteric drugs present key advantages over orthosteric

drugs, in which they modulate the protein activity via a mechanism that does not rely on targeting the highly conserved active sites, hence are perceived as safer and less toxic to the host (Amamuddy *et al.*, 2020) (details about protein allostery are presented section 4.1.2 of [Chapter 4](#)). In this chapter, NMA calculated on the Anisotropic Network Model (ANM) were employed to study the structural dynamics of the protein and to disclose potential allosteric sites. The NMA analysis was combined with motif analysis to recognise motifs uniquely conserved in Plasmodium and located in regions of structural importance. Consequently, establish guidelines towards the selective inhibition of the malaria PTPS enzyme.

5.2 Methods

5.2.1 Structure and sequence data retrieval

The crystal structure of the *P. falciparum* PTPS was retrieved from the online Protein Data Bank (PDB ID: 1Y13). The PTPS structure contained six identical chains, each of 163 residues. The protein homo-hexameric functional unit is thus composed of a total of 978 residues. This structure was used as an input to construct an ENM on the C_{β} atomic coordinates. The ENM was built on the C_{β} atoms as it has been established that they provide a better representation of the side chains' orientation. A harmonic potential within a cut-off distance (R_c) of 15 Å was used to account for the pairwise interactions between all of the C_{β} atoms. The *P. falciparum* PTPS sequence was first retrieved from the PlasmoDB database PlasmoDB (Aurrecochea *et al.*, 2009), then used as a query in a BLAST search for other protein sequences of PTPS from the UniProt database (Bateman 2019). 21 PTPS protein sequences were retrieved, including ten Plasmodium species, four mammalian species, four bacteria species, and three fungi species (details about the sequences retrieval and motif discovery methods are presented in [Chapter 2](#)).

5.2.2 Calculation of the normal modes and atomic fluctuation

[Chapter 4](#)

5.2.3 Modes contribution

The percentage contribution of each mode to the overall protein motion was calculated as follow:

$$\frac{1}{\lambda_x} \times 100 \quad \text{Equation 5.1}$$
$$\frac{1}{\sum_i^N \left(\frac{1}{\lambda}\right)}$$

Where λ_x represent the eigenvalue corresponding to a certain mode x , and N is the total number of modes. The contribution is higher when λ is the more minor; thus, lower frequency modes display a higher contribution to the overall motion.

5.2.4 Visualization of the normal modes

[Chapter 4](#)

5.2.5 Residue cross-correlation analysis

[Chapter 4](#)

5.2.6 Deformation energy analysis

[Chapter 4](#)

5.2.7 Mechanical stiffness analysis

[Chapter 4](#)

5.2.8 Gaussian network model calculations

Gaussian Network Model (GNM) calculations were further performed on the *P. falciparum* PTPS structure using the DynOmics portal (Li *et al.*, 2017) for identifying hinge regions within the PTPS structure. The GNM calculations were performed using a cutoff distance of 15 Å, with a spring constant scaling factor of 10 for contact distances ≤ 4.0 Å and a distance scaling exponent of 2 to decrease the strength of interactions with increasing distance. The DynOmics portal was used to validate the ANM high-frequency vibrating residues further and analyse the mechanical properties of the protein.

5.3 Results and Discussion

5.3.1 NMA of the *P. falciparum* PTPS

NMA was performed on the *P. falciparum* PTPS (PDB ID:1Y13) crystal structure to study its structural dynamics and identify key regions responsible for conformational transitions. NMA on the *P. falciparum* PTPS 3-fold symmetry structure yielded a total of 2934 normal modes. The first six modes were excluded from the analysis as they correspond to global translations and rotations; thus, they amount to a simple rigid body motion. The low-frequency normal modes often describe large-amplitude collective motions involving a larger number of atoms (Bahar, Lezon, Yang, *et al.*, 2010; Mahajan and Sanejouand, 2015; Bauer, Pavlović and Bauerová-Hlinková, 2019). Hence, the first 20 non-trivial modes were selected and studied to identify slow and global conformational changes of the *P. falciparum* PTPS (Table 5.1).

Within the selected 20 lowest-frequency modes, both degenerate and non-degenerate modes were identified. The degenerate modes are modes of the same vibrational frequency and consequently show the orthogonal transformation. In contrast, the non-degenerate ones show unique vibrational frequency and present unique motions that often capture meaningful global motions (Na and Song, 2016). Several studies showed that dominant conformational changes of complexes were found within the lowest non-degenerate modes (Atilgan *et al.*, 2001; Chennubhotla *et al.*, 2005; Isin *et al.*, 2012; Lee *et al.*, 2017; Ross *et al.*, 2018; Shrivastava and Bahar, 2006; Wako and Endo, 2011). Within the first 20 lowest-frequency modes of the *P. falciparum* PTPS, eight non-degenerate modes were identified; each mode exhibited unique eigenvalues frequency, illustrating unique global motions (Table 5.1). The first 20 non-trivial normal modes of PTPS, their associated frequencies, and degeneracy levels are shown in Table 5.1.

5.3.2 Modes contribution

The contribution of the individual low-frequency modes towards the overall motion of the protein was calculated (Table 5.1). The contribution was relatively low and did not reveal a single dominant motion. This can be explained by the large protein size and the extensive degeneracy of the normal

modes. Here the 20 lowest-frequency modes were analysed, which only represent 0.68% of the total modes, yet when combined, they account for 11.08% of the total motion of the protein. The eight non-degenerate modes represent only 0.27% of the total modes but also represent 3.64% of total motion.

Table 5.1. PTPS first 20 non-trivial normal modes, associated eigenvalues, and level of degeneracy. The first six trivial modes were excluded; therefore, mode 1 represents mode 7.

Mode	Eigenvalue	Degeneracy	Contribution%
Mode 1	0.33	2	1.38
Mode 2	0.33	2	1.37
Mode 3	0.48	1	0.96
Mode 4	0.55	2	0.84
Mode 5	0.55	2	0.84
Mode 6	0.89	1	0.52
Mode 7	0.92	1	0.50
Mode 8	0.95	2	0.48
Mode 9	0.96	2	0.48
Mode 10	0.96	2	0.48
Mode 11	0.97	1	0.47
Mode 12	1.02	2	0.45
Mode 13	1.03	2	0.44
Mode 14	1.21	1	0.38
Mode 15	1.34	2	0.34
Mode 16	1.35	2	0.34
Mode 17	1.63	1	0.28
Mode 18	1.68	1	0.27
Mode 19	1.79	1	0.26
Mode 20	1.95	2	0.24

5.3.3 Visualisation of the *P. falciparum* PTPS low-frequency non-degenerate normal modes

The displacement vectors of the individual eight non-degenerate modes were studied by projecting the respective eigenvectors of each of the identified eight non-degenerate modes onto the structure of PTPS (Figure 5.1). This was done to characterize the large-scale global motions within the selected modes. The selected eight non-degenerate modes captured a coupling movement between the tunnel, the N-terminal β -strands as well the central helices, thus demonstrating that these regions primarily

regulate the global dynamics of the protein. The PTPS motions of each low-frequency non-degenerate mode are detailed in the next section.

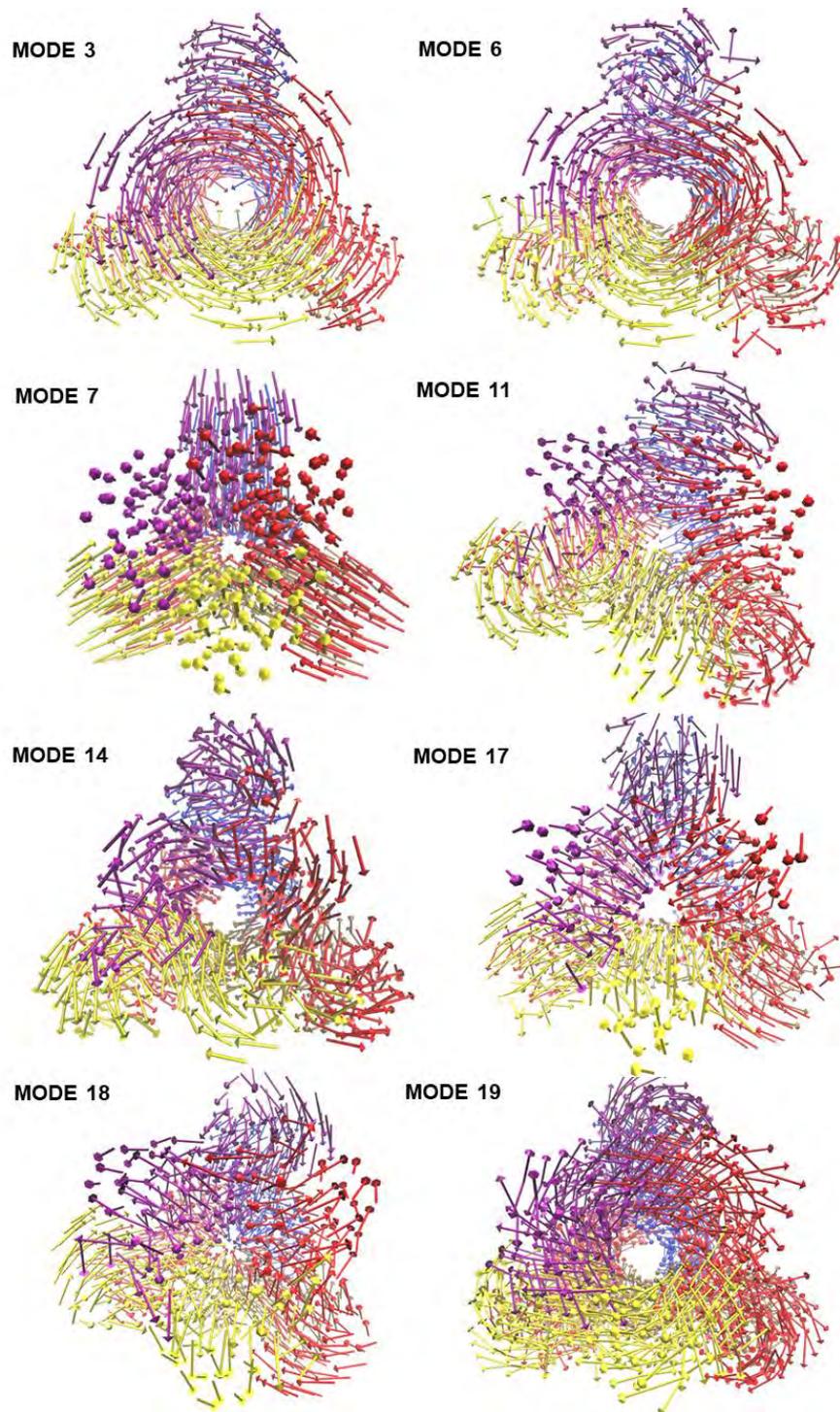


Figure 5.1. The characterized global motions of the *P. falciparum* PTPS structure within the eight slowest non-degenerate modes. The obtained eigenvectors were projected as a set of arrows that denote the direction of motion of the protein C β atoms. Each of the protein chains is coloured

differently for illustration purposes. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2020)).

Mode 3 featured an asymmetric global twist of the PTPS N-terminal regions across the two trimers, which was also accompanied by the rotation of the protein central tunnel in an opposing direction ([Movie 1](#)). **Mode 6** featured a tilt motion of the PTPS N-terminal regions leading to a side-to-side movement of the entire protein structure ([Movie 2](#)). **Mode 7** captured a stretch-like motion resulting in a lateral expansion and contraction of the protein central tunnel. In this mode, the extensive fluctuation of the N-terminal regions was shown to modulate the tunnel expansion and contraction and induce structural changes around the active site region ([Movie 3](#)). **Mode 11** displayed a tilt motion of the two central helices in an opposing direction to the third central helix. ([Movie 4](#)). In all four normal modes: 3,6,7 and 11, the characterized distinctive global motions involving the structural changes around the PTPS tunnel and the active site region were primarily driven and regulated by the protein N-terminal helices. **Mode 14** captured an outward and inward movement of the PTPS tunnel (likened to an engulfing movement) ([Movie 5](#)). **Mode 17** showed a prominent bending of the central helices from side to side, resulting in the protein tightening in the same direction ([Movie 6](#)). **Mode 18** featured a breathing-like motion characterized by the upward and downward movement of the central β -sheets ([Movie 7](#)).

Lastly, **Mode 19** displayed a clockwise rotation of the protein core while the N-terminal regions twisted in an opposing direction ([Movie 8](#)). From the characterized global motion of the eight non-degenerate modes, we can summarize the following: The PTPS N-terminal β -strands (residue 76 to 114) and the central helices (residue 136 to 150) exhibited the highest mobility and were responsible for regulating the global motions of the protein. Motions such as the wringing and tilting of the PTPS N-terminal β -strands modulated the tunnel's surface exposure and buried regions within the tunnel; thus, it can promote the binding of PTPS to substrates or accessory proteins. Lastly, the induced structural changes of the PTPS tunnel can be of importance to the efficiency/exposure of the active site, in which motions such as expansion and contraction of the tunnel can promote the entry of the substrate. Furthermore, the twist and shear motions may propel the substrate through the tunnel.

Overall, the findings disclose key information on the PTPS structural dynamics and its function, which can lay the ground for designing allosteric modulators that target regions of structural significance and relevance to the enzyme's function.

5.3.4 MSF analysis of the individual eight low-frequency non-degenerate modes

The atomic MSF profile of the individual eight low-frequency non-degenerate modes was calculated and plotted (Appendix 3.1). This was done to further show mobile regions that drive the global protein motion in each mode. The N-terminal β -strands and the central helices have shown notable mobility in each profile, whereas the protein core presented minimal fluctuation. Sharp peaks on the MSF profiles designate these regions.

A plot graph of the MSF values from all the normal modes and the B -factor values is shown in Appendix 3.2. In this graph, high peaks corresponded to the N-terminal regions followed by the central helices were found to agree in both MSF and B -factors. Notably, the protein tunnel region showed relatively higher fluctuation in the B -factor graph, suggesting less agreement in this region. Nevertheless, this can be reasoned to the fact that the B -factor only represents a static snapshot of the protein dynamics.

Figure 5.2-B shows the atomic fluctuation over the first 20 non-trivial low-frequency modes mapped onto the PTPS structure. The combination of the lowest 20 modes MSF identifies the PTPS terminal regions as the most mobile (red) regions of the structure (Figure 5.2-B). To further validate the atomic fluctuations of the first 20 non-trivial low-frequency modes, the MSF values were compared to the crystallographic atomic displacement parameters, the B -factors. Figure 5.2-A shows the experimental B -factor values mapped onto the structure; it also identifies the N-terminal β -strands and the central helices as regions of high mobility.

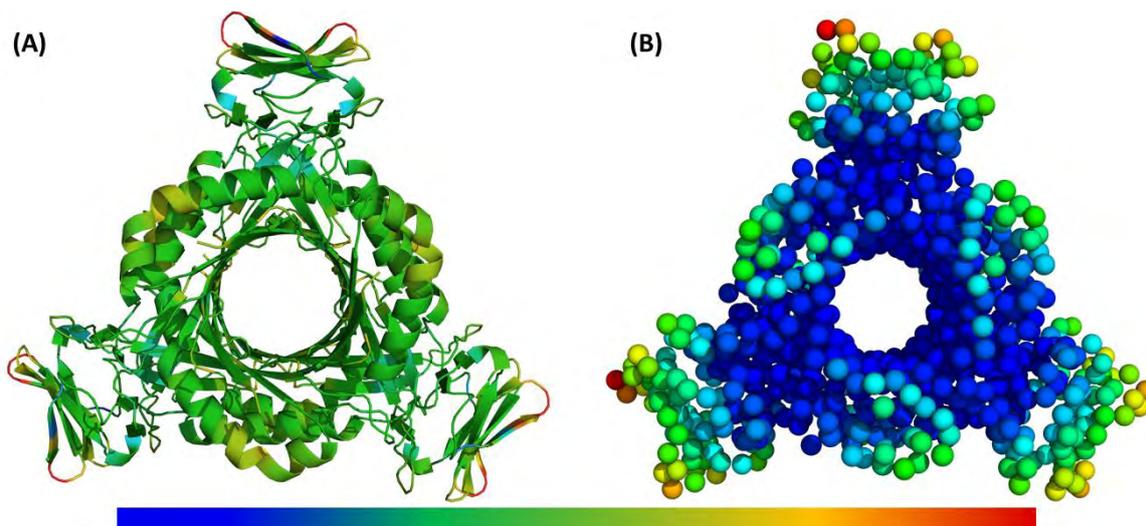


Figure 5.2. (A) The experimental B -factors mapped onto the *P. falciparum* PTPS structure (PDB ID: 1Y13) (B) Atomic mobility over the 20 lowest-frequency normal modes. Regions of the lowest fluctuation are shown in blue, with higher mobility in green, yellow, and red. The figure illustrates the high mobility of the protein terminal regions, central helices, and the rigidity of the protein core, indicating that the NMA has captured most of the experimentally determined atomic fluctuation.

5.3.5 Deformation energy analysis

The deformation energy analysis was carried out to study the deformation energy distribution across the protein structure, hence, identify rigid regions within the PTPS structure. The deformation energy values were calculated as the sum of the contribution from the first 20 non-trivial low-frequency modes. The derived energy values were mapped onto the PTPS structure (Figure 5.3). The obtained results demonstrated the high mobility of the PTPS N-terminal regions (low deformation energy) and a high build-up of deformation energy in the PTPS central tunnel (Figure 5.3). The result further illustrates that the fluctuation of N-terminal regions (located ~ 30 Å away) may have a long-distance effect on the protein tunnel core.

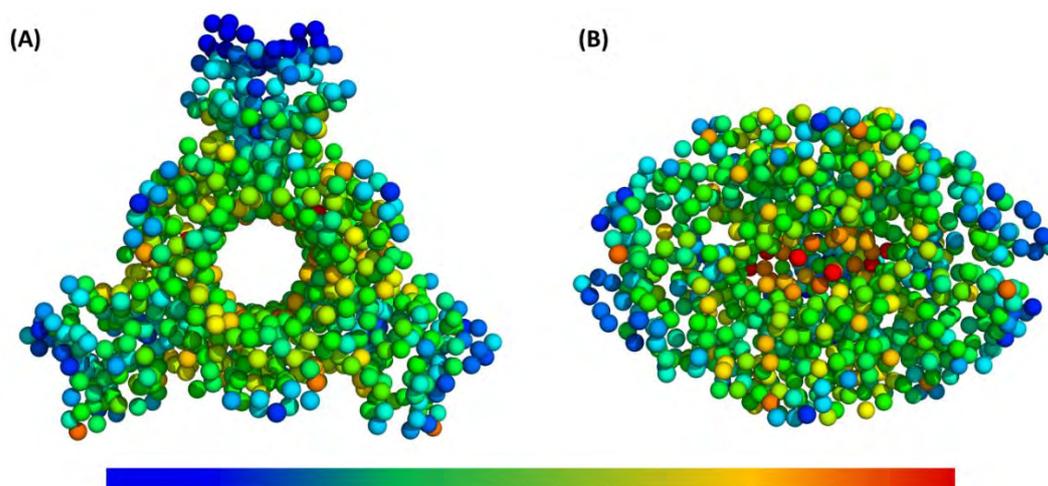


Figure 5.3. Deformation energy values mapped onto the *P. falciparum* PTPS structure (A) top view (B) side view. The colours of the atoms indicate the amount of deformation. The dark blue sites are the least deformed/most mobile, whereas red areas are strongly deformed/ rigid. The highest deformation energy was observed within the protein tunnel region (red).

5.3.6 Residue cross-correlation analysis

A heatmap showing the cross-correlation of the C_{β} atoms within motions of the first 20 non-trivial low-frequency modes is shown in Figure 5.4. The analysis highlighted residues moving in a concerted manner, which further illustrates their involvement in the collective structural dynamics of the enzyme. The off-diagonal elements presented a positive correlation coefficient within the same protein chain; more specifically, the residues of the N-terminal β -strands gave rise to a positive correlation. Given that PTPS is a multimeric protein, chains within one trimer exhibited a similar direction of motion. Anticorrelated motions were identified across the two dimers, which can be understood due to the twisting or wringing of the helices across the trimers. The observed anticorrelation across the two trimers designates the opposing twisting or wringing of the terminals, allowing the tunnel to open and close accordingly.

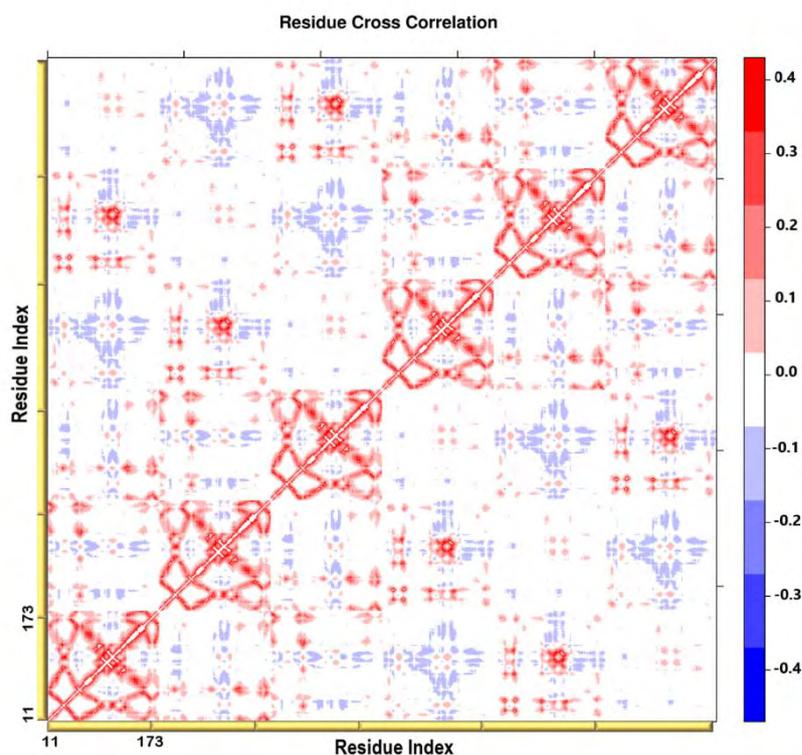


Figure 5.4. A heatmap representation of the *P. falciparum* PTPS residues pairwise cross-correlation. The scale colour bar on the right indicates the extent of the correlation in which the red colour highlights correlated motions (residue pairs moving together in the same direction), while the blue colour highlights the anti-correlated movements (residue pairs moving in the opposite direction). The yellow blocks partition each chain of the protein. (Reproduced with permission from (Khairallah, Ross and Tasthan Bishop, 2020)).

5.3.7 Collective MSF analysis of the lowest and highest-frequency normal modes

Given that low-frequency, normal modes capture large-scale/dominant protein movements that can be of functional significance (Bahar, Lezon, Bakan, *et al.*, 2010; Mahajan and Sanejouand, 2015), a collective MSF profile of the 20 lowest-frequency modes from one protein chain was studied (Figure 5.5-B). This was important to identify key residues modulating the protein conformational transitions. Within the MSF profile, residues of large contributions toward the global protein dynamics were identified and presented as sharp peaks of maximal fluctuation.

As shown in Figure 5.5-B, the most substantial contributions to the protein atomic fluctuations emerged from the PTPS terminal regions and central helices, demonstrating their key role and involvement in driving the global protein dynamics, e.g. controlling the open-closed conformational

transitions of the PTPS protein. G35, K97, N107, and S145 were among the high fluctuating residues and located within the *P. falciparum* PTPS unique motifs. N107 was found in Motif 7 (position 9), and S145 was found in Motif 8 (position 10) (Figure 5.5-A).

As the frequency increases, the modes become more localized and accompanied by the fast-rapid local vibration of individual residues. High-frequency vibrating residues are highly conserved and essential for maintaining the stability and functions of proteins (Haliloglu *et al.*, 2005). As a result, the collective MSF profile of the PTPS 20 highest-frequency modes was also studied (Figure 5.5-C). Within this profile, high-frequency vibrating residues were identified and presented by the sharp peaks. Some of the identified residues of high-frequency vibration were also located in motifs sites, including the active site residues H29, H41, H43, and other catalytically important residues, including L78, I83, V159, E161, and A167. The identified residues further overlapped with motif sites, with L78 and I83 found in Motif 6 at (positions 3 and 8), V159, E161, and A167 in Motif 5 (position 6, 8, and 14), respectively (Figure 5.5-A). Overall, the identified key structural sites were located within motifs sites. These motifs were already identified as unique to the Plasmodium sequences and displayed high sequence variation compared to the mammalian sequences (section 2.3.2 of [Chapter 1](#)). From this result, we can establish that the lowest and highest-frequency normal modes revealed residues of notable mobility that were also located in the conserved structural motifs.

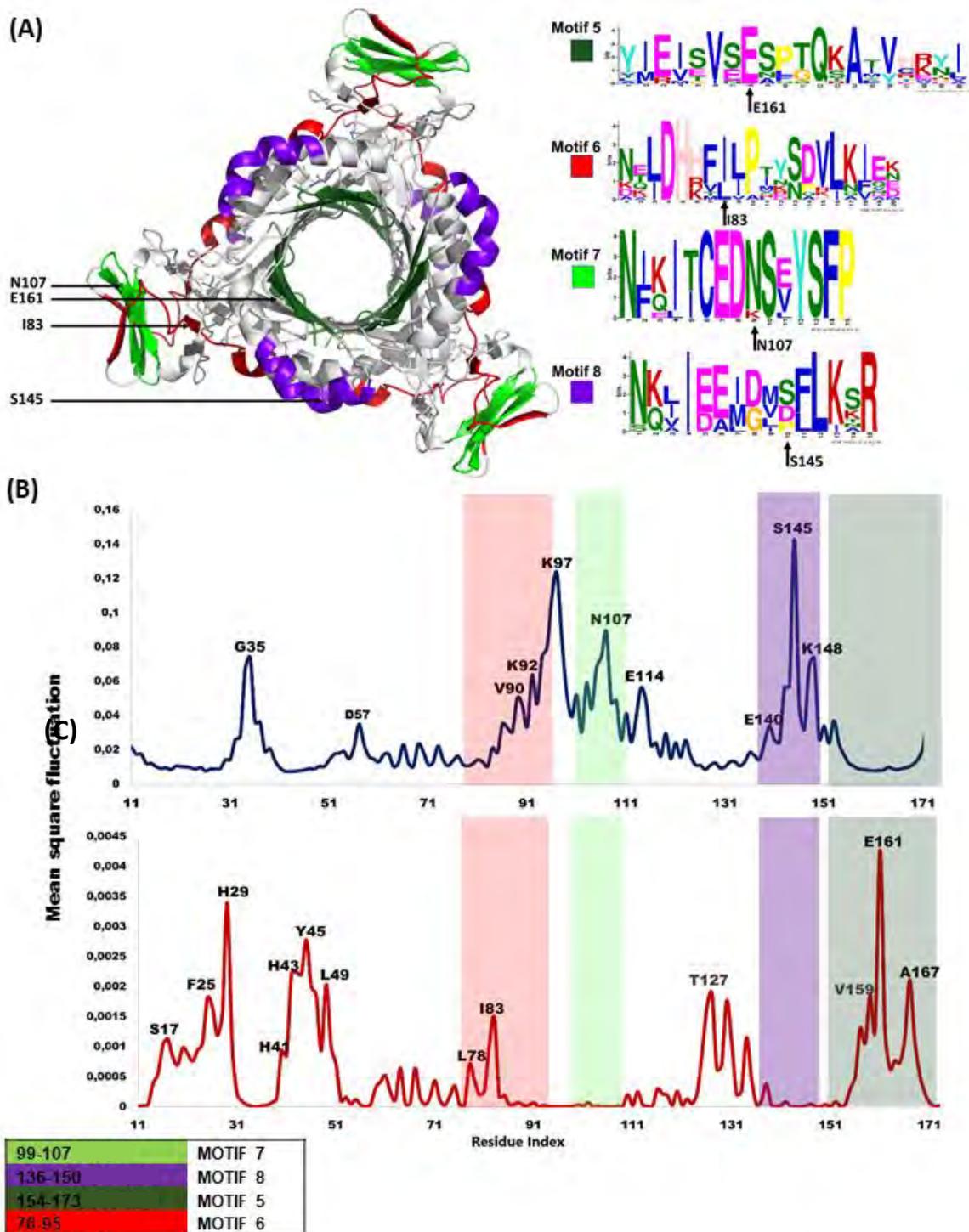


Figure 5.5. (A) The *P. falciparum* PTPS sequence motifs are mapped onto its structure (right), the location of high-frequency vibrating residues within the structure is illustrated by the black arrows. The MEME sequence logo contains a stack of amino acids letters at every position in the motif (left). The black arrows indicate the location of the high-frequency vibrating residues within the motifs. MSF profiles of (B) 20 lowest-frequency and (C) 20 highest-frequency normal modes. Residues of notable fluctuation within the slowest and fastest modes are labelled on each protein chain. The dark

green, purple, red, and lime green bars highlight the location of motifs 5, 6, 7, and 8, respectively. (Reproduced with permission from (Khairallah, Ross and Tasthan Bishop, 2020)).

The PTPS active site pocket is surrounded by a cluster of hydrophobic residues that were reported to interact with the substrate ring (Bürgisser *et al.*, 1995; Colloc'h, Poupon and Moron, 2000). In the NMA analysis, more specifically the MSF profile of the 20 highest-frequency modes, several of these hydrophobic residues presented notable fluctuation in the high-frequency modes (Figure 5.5-C), including the non-polar and hydrophobic residues L78, I83, V159, A167 as well as the two aromatic residues F25 and Y45. The residues and their location within the active site region are shown in figure 5.6.

Additionally, E161 and T127, located at the bottom of the active site pocket (Figure 5.6-B), showed notable fluctuation within the fast frequency modes, with E161 displaying the highest residue fluctuation (Figure 5.5-B). The two residues were previously reported for their key role in substrate recognition and binding, in which they both act as proton donors and acceptors during catalysis (Nar *et al.*, 1994; Bürgisser *et al.*, 1995; Ploom *et al.*, 1999; Nar, 2011). Nar and colleagues also reported that T105, T106, and E107 constitute an acceptor site for the substrate ring during catalysis in the *Rattus norvegicus* PTPS structure (PDB ID: 1B6Z). In the *P. falciparum* PTPS structure, these residues correspond to S126, T127, E128; all three presented a notable fluctuation within the MSF profile of the 20 highest-frequency modes. Overall, the identified residues of high-frequency vibration also overlapped with previously reported functionally essential residues, further validating our NMA results.

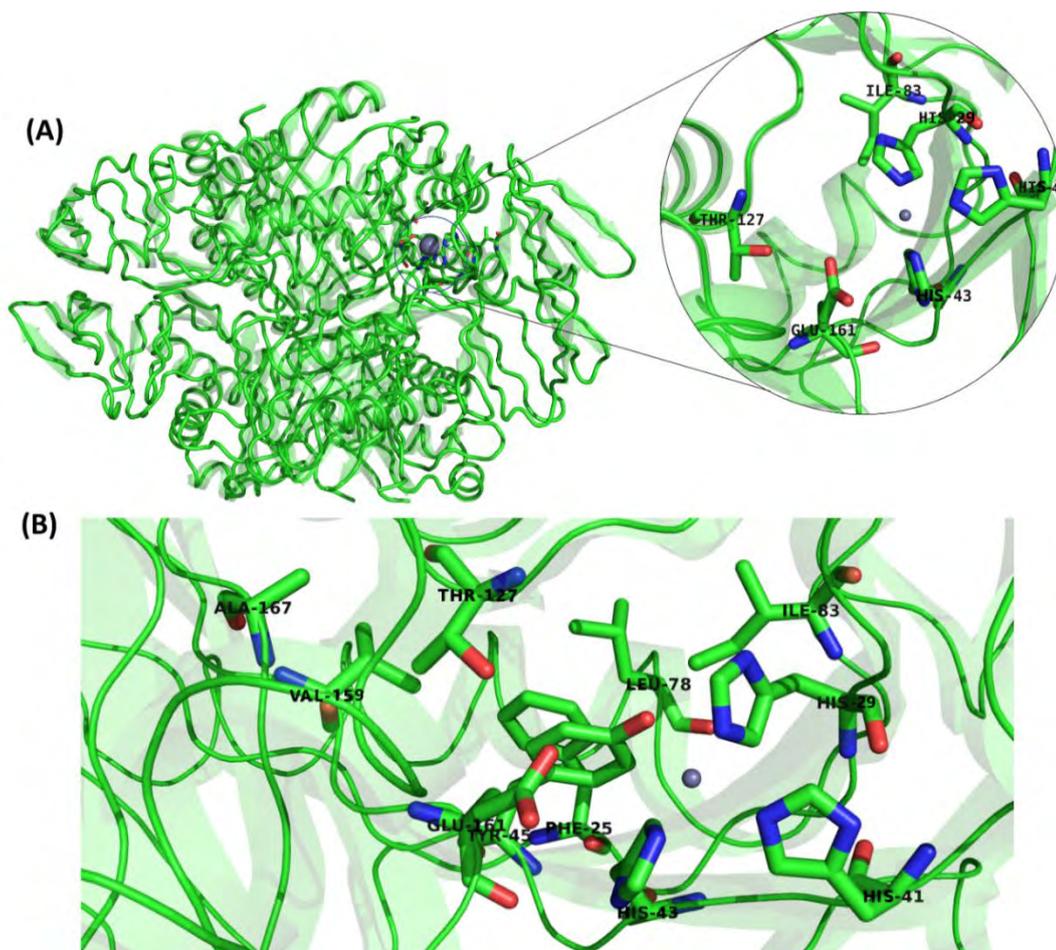


Figure 5.6. (A) The structure of *P. falciparum* PTPS and the metal centre (active site) in the circled zoom (B) boxed zoom showing the location of the identified high-frequency vibrating residues. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2020)).

5.3.8 *P. falciparum* PTPS hinge residues

Hinge regions are often found between domains in a protein to allow flexibility and permit the domains to move relative to one another or clamp down on a substrate (Yang and Bahar, 2005; Towler et al., 2004; Amusengeri and Tastan Bishop, 2019). Hinge regions are involved in mediating conformational changes of functional significance (Flores *et al.*, 2007), such residues can be identified within the low-frequency normal modes of the ANM or modes obtained by GNM (Emekli *et al.*, 2008; Dutta and Bahar, 2010; Liu and Bahar, 2010; Su *et al.*, 2011). GNM allows for predicting the magnitude of motions at the cost of losing directions (Atilgan *et al.*, 2001) and thus easily enables the identification of hinge sites. Here the DynOmics portal was used to disclose hinge regions within the PTPS structure, specifically the domain separation analysis based on the GNM. Residues acting as

hinges were identified and found to well correspond with the ANM high-frequency vibrating residues, including the PTPS active site and neighbouring residues H29, H41, and H43 as well as E16 (Figure 5.7). The high-frequency vibrating residues Y45, L49, T127, S148, V159, E161 (as well as their neighbouring residues) were also identified as hinges (Figure 5.7). Given that high-frequency vibration residues donate kinetically hot residues that underlie protein's fold and stability (Demirel *et al.*, 1998), we can perceive these residues as functionally necessary and act as hinges/rigid points to maintain the protein fold and stability during the large-scale domain movement.

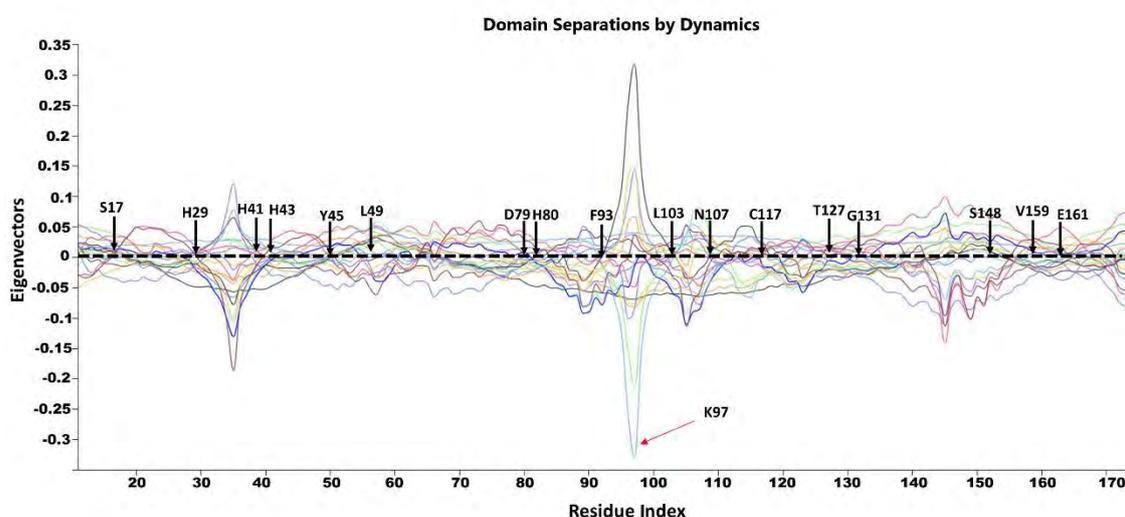


Figure 5.7. GNM based identification of global hinge sites over the 20 lowest-frequency modes. Hinge residues are located at the crossover line where the eigenvectors' values are equal to zero. Residues surrounding the hinge residues fluctuated in the positive and negative direction of motion. The black arrows illustrate the identified hinge residues. K97 (indicated by a red arrow) is an example of a residue identified of notable fluctuation within the MSF profile of the 20-lowest-frequency normal modes. Thus, it contributes to driving the global protein motion and does not act as a hinge site. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2020)).

GNM analyses were carried out using the DynOmics portal (Li *et al.*, 2017) to further validate the high-frequency vibrating residues that were identified in the ANM calculations. The results obtained from the GNM analysis agreed with the ANM results, in which the N-terminal and central helices were identified as the most mobile regions that contribute primarily toward the large conformational changes (Figure 5.8-A). The GNM calculations located high energy hotspots in the active site region that also overlapped with the high-frequency vibrating residues identified in the ANM (Figure 12-B).

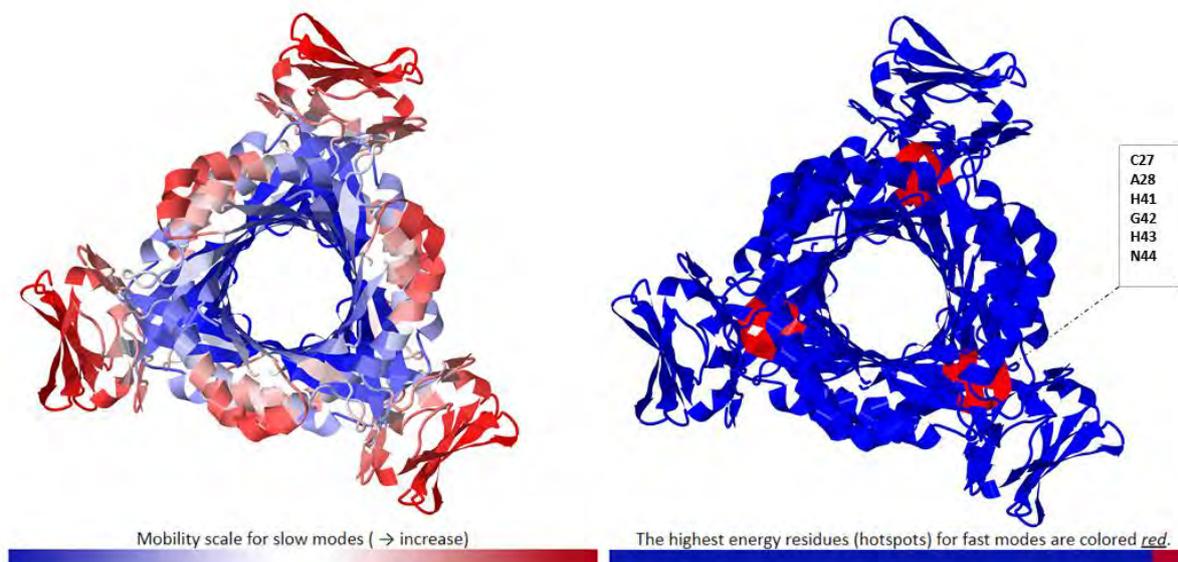


Figure 5.8. Colour-coded representation of the *P. falciparum* PTPS structure based on the mobility of the residues over the (A) 20 lowest-frequency and (B) 20 highest-frequency GNM modes. The colour scale varies from (blue) as the most rigid to (red) as most mobile sites. The boxed zoom shows the GNM kinetically hot residues/hotspots. The analyses and figures were obtained from the DynOmics portal (Li *et al.*, 2017). (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2020)).

5.3.9 Mechanical stiffness analysis

The mechanical stiffness analysis showed that residue pairs located in the active site surrounding region exhibited strong pairs of interactions, shown by the heatmap in Figure 5.9-A. This further suggests that the active site surrounding residues, including T127 and E161, are stiffer and exhibited strong resistance to deformation. A profile of the mechanical stiffness of individual residues to deformation is presented in Figure 5.9-C. From this profile, the active site residues H29, S17, L78, I83, F93, T127, G131, and E161, were identified as stiff sites. These residues were also among the highly conserved, high-frequency vibrating residues. Lastly, residues exhibiting the strongest effective force constant and their location within the protein structure were mapped onto the PTPS one chain (Figure 5.9-B). In this figure, the identified residues with the strongest effective force constant T127, E161, G131 were located around the active site area and were also among the ANM high vibrating and hinge residues.

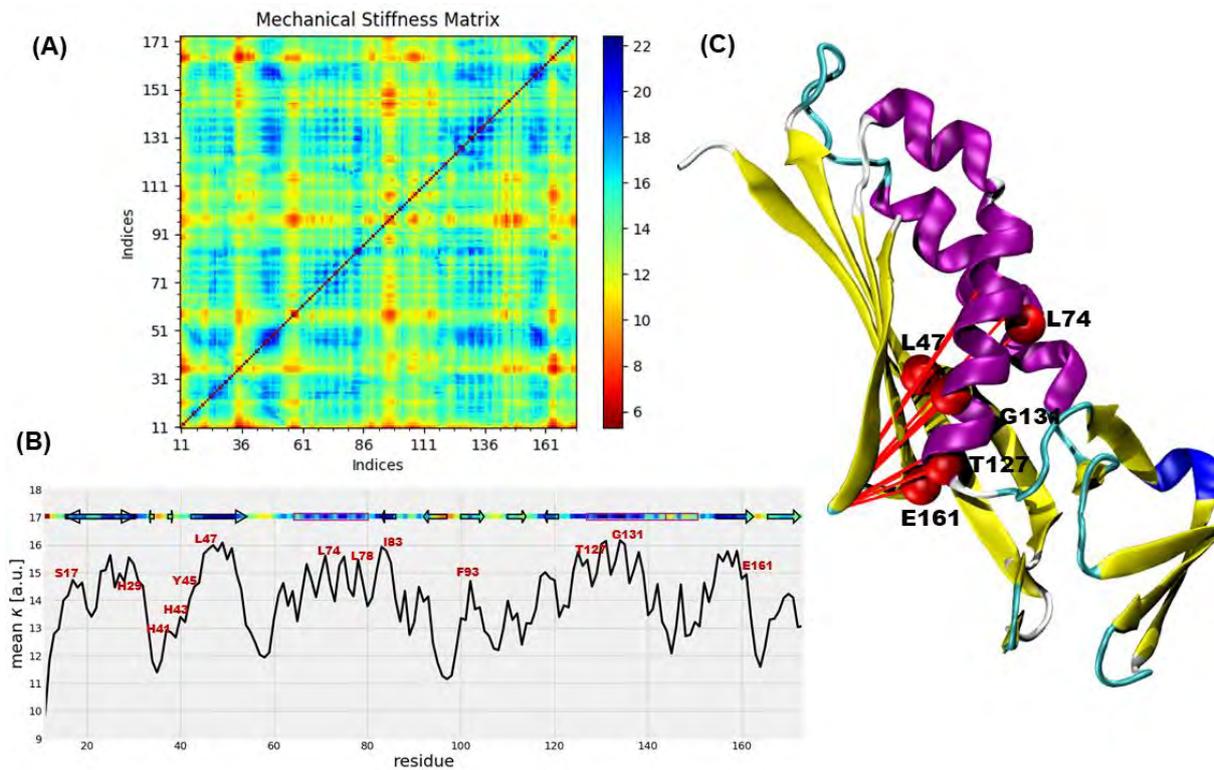


Figure 5.9. (A) ProDy mechanical stiffness heatmap of the *P. falciparum* PTPS. The colour bar illustrates the effective force constant values (blue as a strong /stiff pair of interactions and red as a weak/flexible pair of interactions). (B) The mean value of the effective spring constant over all pairs for each residue, the protein secondary structures, are shown along the upper abscissa. (D) A cartoon representation of a single chain of *P. falciparum* PTPS with the residues of the strongest interaction mapped onto the structure and shown as red spheres. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2020)).

5.4 Conclusions

Studying the conformational variability of proteins is fundamental in our knowledge of key biological processes. In this chapter, the structural dynamic of the malaria parasite *P. falciparum* PTPS enzyme was studied via the ENM based NMA. The analysis provided a detailed structural insight of key regions that modulate the enzyme conformational transitions and are imperative to its function. Within the selected low-frequency, non-degenerate essential global motions of functional significance were identified, including the central opening and closing of the tunnel that is accompanied by the N-terminal regions wringing and twisting motion. Sequence motifs uniquely conserved in the Plasmodium PTPS enzymes overlapped with sites of structural significance that modulate the PTPS global dynamics, including the N-terminal regions, the central helices and the protein tunnel. As a result, the characterized regions present potential allosteric sites for future antimalarial drug discovery efforts. The obtained results can pave the way for future selective inhibition as these regions are not conserved in the human PTPS. Finally, hinge residues were identified and found to overlap with high-frequency vibrating residues located in the protein core. Such residues can be regarded as hotspots central for maintaining the stability and fold of the proteins; thus, they can also be proposed as allosteric sites. Overall, the NMA provided key structural information about the *P. falciparum* PTPS and highlighted potential sites for its allosteric modulation.

Concluding Remarks and Future Perspectives

The discovery of new antimalarial drug targets and drugs of a novel mechanism of action is still necessary to overcome the current antimalarial drug resistance problem. Thus, the main focus of this thesis was to contribute to the existing and growing knowledge of antimalarial drug discovery and aid the efforts to overcome the drug resistance problem. The study demonstrated the sequence and structural aspects of the malaria parasite *de novo* folate biosynthesis pathway enzymes guanosine-5'-triphosphate (GTP) cyclohydrolase I (GCH1) and 6-pyruvoyl tetrahydropterin synthase (PTPS), further characterized the allosteric modulation of their conformational dynamics and produced force field parameters to enable studying the active site metal ions within the two enzymes. Based on various computational tools and approaches, the research study described in this thesis provided novel insights and described the two enzymes as attractive alternative drug targets, further identified potential allosteric drug targeting sites within their 3D structures. The impact of this research can pave the way for the discovery of new effective antimalarial antifolate drugs with novel mechanisms of action.

Chapter 1 provided a comprehensive literature review of malaria, current control regimes and treatments. The chapter further detailed the malaria parasite life cycle, antimalarial drug resistance problem, and the malaria parasite *de novo* folate biosynthesis pathway. The reviewed literature in this chapter established the context of the research and supported the importance of discovering new antifolate antimalarial drugs and metabolic targets to overcome the current antimalarial drug resistance. The chapter is then closed with the study aim, rationale and research primary objectives.

Chapter 2 described comparative sequence analysis, motif discovery, and phylogenetic analysis to disclose unique sequence and structure-based features of the malaria parasite GCH1 and PTPS. The study's main objective was to identify sequence similarities as well as differences between the malaria parasite enzymes and their human homologs. The sequence analysis of the two enzymes disclosed differences between the parasite and human protein sequences, including the active and surrounding regions. Two unique motifs were identified within the *Plasmodium falciparum* (*P. falciparum*) GCH1

enzyme and four within the *P. falciparum* PTPS enzyme. The phylogenetic tree calculations showed the distinct clustering of the Plasmodium enzymes and the different evolutionary histories of their human homologs. Overall, the analysis presented in this chapter has disclosed key differences between the parasite and human enzymes despite the functional and structural conservation. Such differences are crucial and can be exploited to develop antifolate antimalarial drugs of good selectivity and less toxicity to the host.

Chapter 3 described the development of force field parameters of the GCH1 and PTPS active site metal ions via quantum mechanics (QM) scans of a high level of theory. The resultant energy profiles from the QM scans were then fitted to terms in AMBER potential energy function via the least square method, and parameters were derived. The newly generated force field parameters were then evaluated via all-atom molecular dynamics simulations (MD) and found to adequately describe and maintain the coordination environment of the metal ions during the MD simulations. The reported force field parameters will enable the MD simulations of the two enzymes and the accurate description of their active sites metal ions, which can be of practical value to any future *in-silico* drug discovery attempts.

Chapter 4 presented normal mode analysis (NMA) of the *P. falciparum* GCH1 to understand its structural dynamics as well as its allosteric modulation mechanism relevant for drug discovery. Notable global motions of the *P. falciparum* GCH1 were characterized within four low-frequency non-degenerate modes. Such motions were accompanied by structural changes in the protein central five-helix bundle and driven by the high mobility of the protein N-terminal helices. Hot spot residues were also identified within the 20 highest-frequency normal modes of the *P. falciparum* GCH1; such residues are crucial for maintaining the protein fold stability and function and can be perceived as potential allosteric drug targeting sites. The study also invited attention to the mammalian GCH1 and its regulatory partner, the GCH1 feedback regulatory protein (GFRP). Six non-degenerate modes described the coupling of the two proteins and further explained the two proteins' structural dynamics. The MSF profile of the 20 highest-frequency modes revealed the local difference between the GCH1-GFRP inhibitory and stimulatory complexes, thus identifying residues responsible for modulating the

conformational transitions between the two complexes. Lastly, hotspot residues of the GCH1-GFRP complex were identified and found within the effector molecules tetrahydrobiopterin, and phenylalanine binding sites and the GCH1 central five-helix bundle. Overall, this study demonstrated the NMA approach's ability to disclose key structural sites responsible for modulating the global transition, such as hinges or hotspots that maintain the protein stability and function.

The final chapter of the thesis, **Chapter 5**, described the NMA of the subsequent enzyme PTPS to investigate its structural dynamics and the modulation of its conformational changes. Large-scale collective motions of the PTPS were characterized within eight low-frequency non-degenerate. The identified global motions demonstrated key aspects of the protein dynamics, such as a prominent regulation of the protein tunnel/barrel by the N-terminal and central helices regions. This further suggests their essential role in modulating the global protein dynamics. Sequence motifs that were uniquely conserved in the *P. falciparum* PTPS enzymes were found to overlap with sites of structural significance that have been shown to modulate the PTPS global dynamics, including the N-terminal regions, the central helices as well as the protein tunnel. Hinge residues were also identified and found to overlap with the high-frequency vibrating residues located within the protein core. The hinge regions were highly restricted during the global protein motions, as shown by the fluctuation profile within the 20 lowest-frequency normal modes, thus acted as pivots of the collective motions. Lastly, the identified key sites were regarded as potential allosteric drug targeting sites for the *P. falciparum* PTPS allosteric modulation.

Overall, the study presented in this thesis has provided novel insights on the structural dynamics of the malaria parasite *de novo* folate synthesis pathway enzymes GCH1 and PTPS further disclosed potential allosteric drug targeting sites within their 3D structures. The two enzymes have never been targeted before and can provide an excellent chance to overcome the drug resistance that is developed in the previously targeted enzymes, the dihydrofolate reductase (DHFR) and dihydropteroate synthase (DHPS). The research study exploited and further studied the unique structural features of the two enzymes to identify the new allosteric druggable sites. The NMA approach, in conjunction with the

sequence and structural analysis, was shown to be suitable for an efficient and speedy identification of allosteric drug targeting sites within the two proteins; such sites can then be targeted/screened against libraries of drug compounds.

Drugs targeting the allosteric sites will be more desirable, selective toward the parasite's enzymes, and less toxic to the human host. The study also reported key sequence and structure-based features that distinguished the parasite enzymes from their human homologs and thus can provide an opportunity for selective targeting of the two enzymes, consequently developing more effective and less toxic drugs. Accurate force field parameters of the GCH1 and PTPS active site metal ions were generated and further validated. The generated force field parameter will enable and deliver accurate molecular dynamic simulations studies involving these two enzymes. Finally, the reported results of this thesis provide a solid platform for the allosteric modulation of the two enzymes, which can contribute to and aid in the ongoing antimalarial drug discovery efforts to overcome the resistance problem.

If put forward, the impact of the research can decrease the strain imposed by the disease on the primary healthcare facilities, especially in malaria-endemic regions, thus improving the lives of suffering individuals and further promoting a healthy population.

Further research and improvement of this study may include a more systematic and quantitative analysis of the identified potential drug binding pockets using available pocket identifier tools such as Allosite and FTMAP, to mention a few. A comparison between the MD simulations and the NMA can also be of great value to the study; other available 3D structures of both GCH1 and PTPS can be studied and compared to the results of this research study.

Lastly, future work could also involve high throughput virtual screening of drug compounds targeting the identified allosteric sites. The screening can be performed using various online databases of natural compounds such as the South African Natural Compounds Database (SANCDB, <https://sancdb.rubi.ru.ac.za/>) or existing therapeutics compounds (for drug repurposing) such as the DrugBank database (<https://go.drugbank.com/>). Top promising compounds of high binding affinities and binding modes can be then selected for all-atom molecular dynamics (MD) studies and post-MD

trajectory analysis to assess the effects of the selected hit compound on the dynamics of the proteins. The study can be integrated with wet laboratory assays to further confirm the legitimacy of the *in silico* identified hits. The latter can be achieved through collaborative initiatives with research facilities with *in vitro* antimalarial drug efficacy assays.

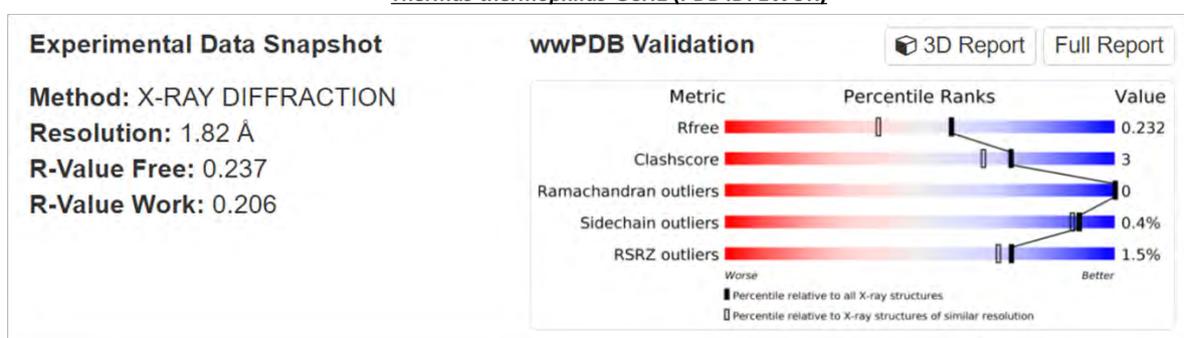
Appendices

Appendix 1

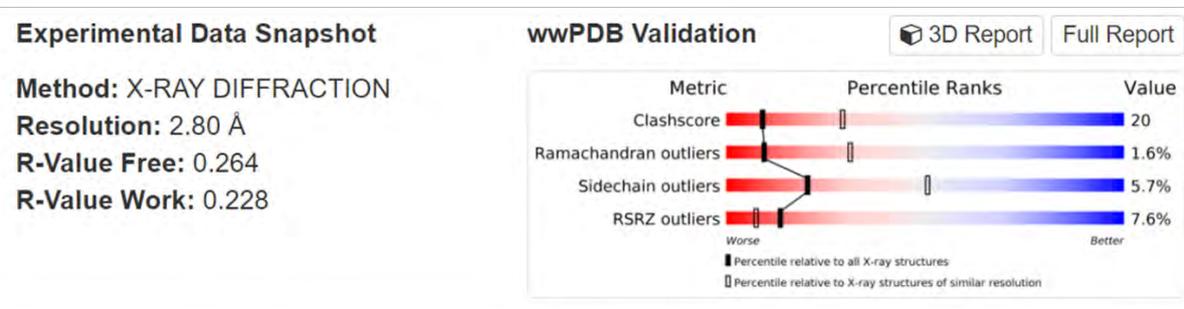
Organism	<i>Thermus thermophilus</i>
PDB ID	1WUR
E-value	3.50E-54
R-Value Free	0.238
R-Value Work	0.206
Structure resolution	1.82 Å
z-DOPE score	-1.15

Appendix 1.1. Template structure information (Tanaka *et al.*, 2005)

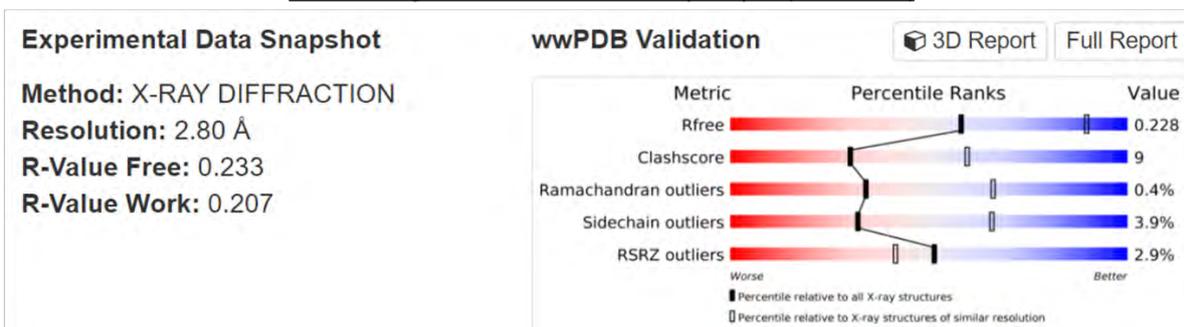
Thermus thermophilus GCH1 (PDB ID: 1WUR)



Rattus norvegicus GCH1-GFRP inhibitory complex (PDB ID: 1WPL)



Rattus norvegicus GCH1-GFRP stimulatory complex (PDB ID: 1IS7)



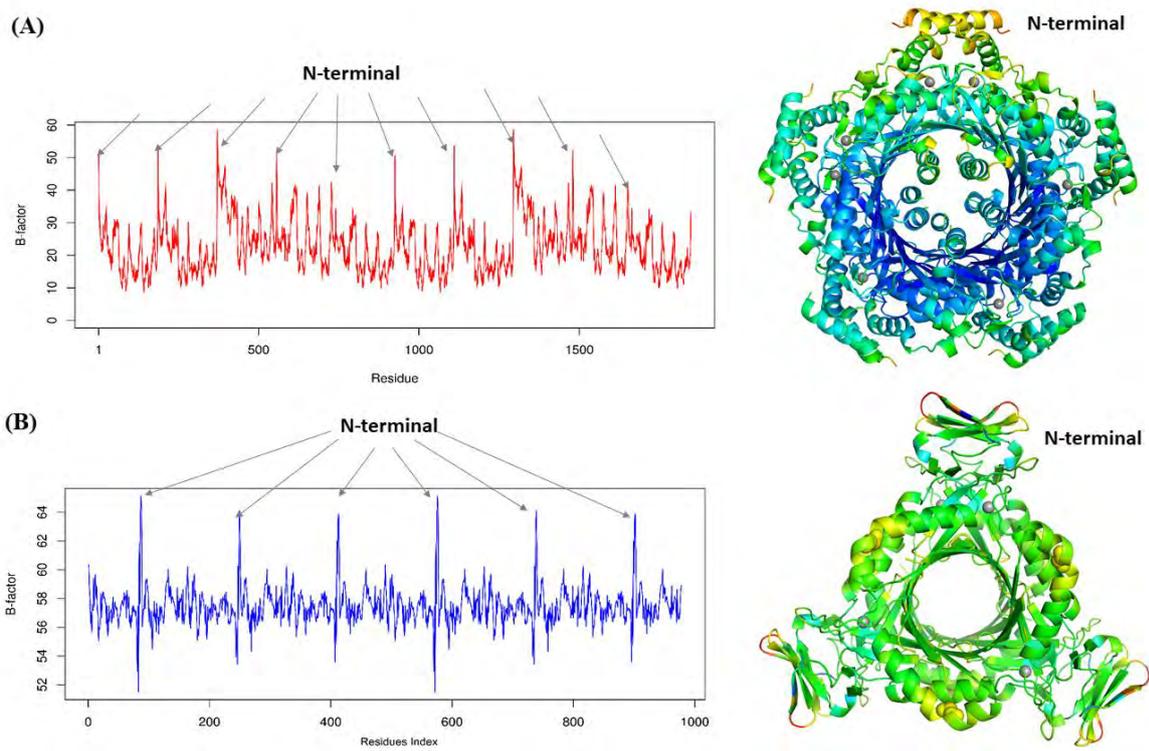
Appendix 1.2. wwPDB validation panels of the retrieved 3D structures, including the template structure information for the *Plasmodium falciparum* GCH1 homology model.

Appendix 1.4. PROCHECK local quality assessment of the *P. falciparum* GCH1 homology model.

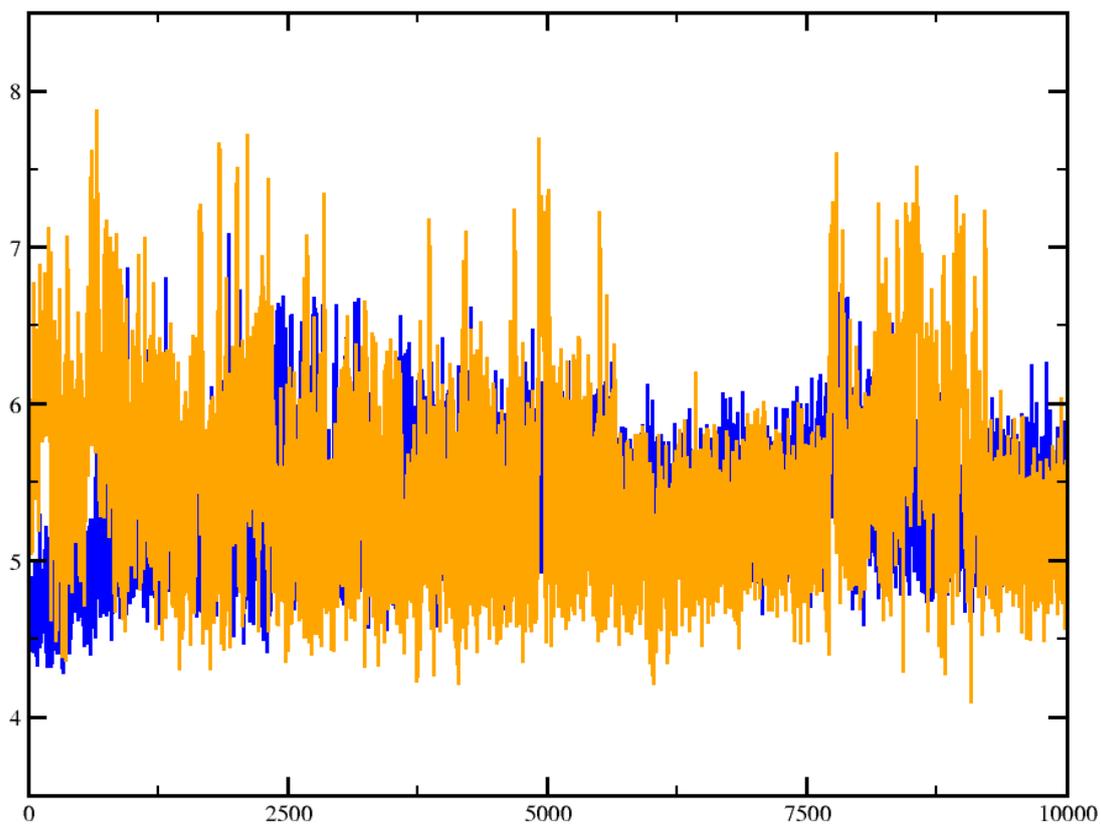
Organism	Model	PROCHECK validation report			
		Number of residues in favoured region/Core (Percentage value)	Number of residues in additional allowed region (Percentage value)	Number of residues in generously allowed region (Percentage value)	Number of residues in disallowed region (Percentage value)
<i>P. falciparum</i>	Model 1	90.6%core	7.0%allow	2.2%gener	0.2%disall
	Model 2	90.6% core	7.3% allow	2.0% gener	0.2% disall
	Model 3	91.0% core	6.7% allow	2.1% gener	0.2% disall
	Crystal structure(1WUR)	95.6% core	4.4% allow	0.0% gener	0.0% disall
Template					

Appendix 1.5. Crystal structures, QM optimized subsets, and MD simulations dihedral values.

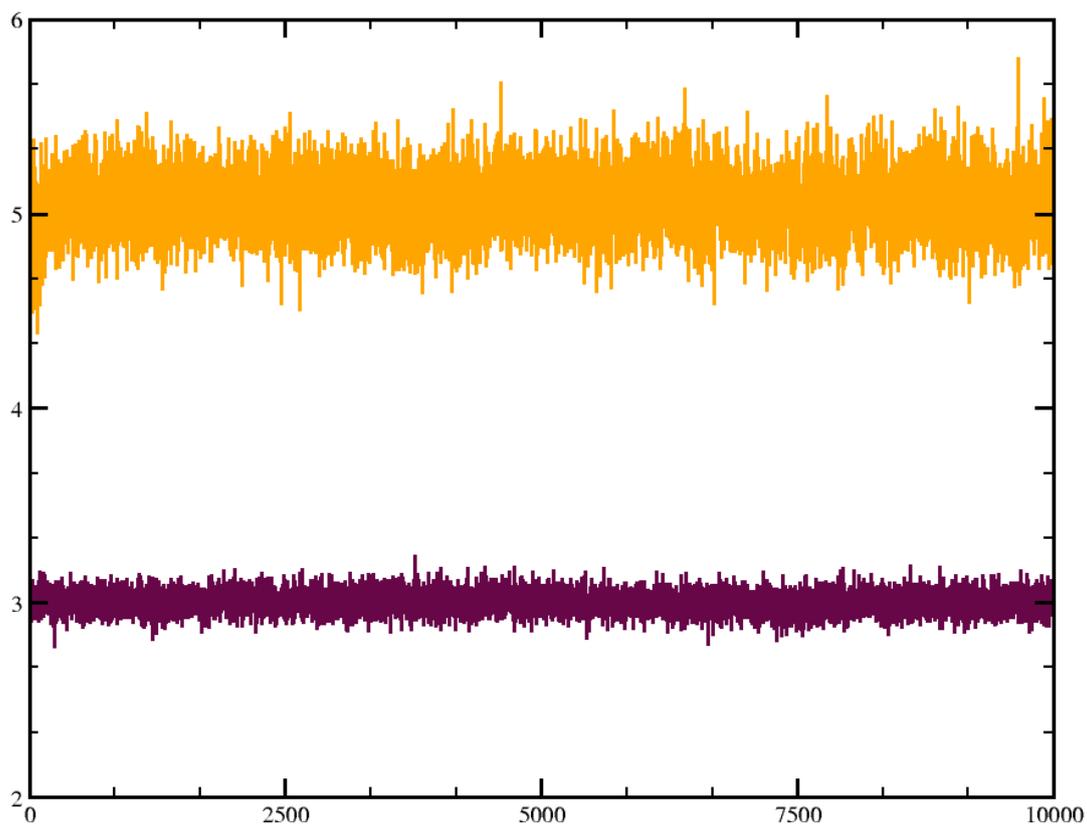
GCH1			
Dihedral angle	Crystal Structure	QM	MD
Cys-SG(108)-Zn ²⁺ -His-ND1(111)-His-CE1(111)	165.96	154.17	143.00 ± 7.27
Cys-SG(179)-Zn ²⁺ -His-ND1(111)-His-CG(111)	-42.49	-31.26	-57.85±10.57
PTPS			
His _{SNE2} (43)-Zn ²⁺ -His _{SNE2} (29)-His-CD2(29)	-53.57	-82.28	-50.86±8.27
His _{SNE2} (43)-Zn ²⁺ -His _{SNE2} (41)-His-CD2(41)	-3.53	7.95	-17.46±8.79
His _{SNE2} (29)-Zn ²⁺ -His _{SNE2} (41)-His-CD2(41)	-101.20	-116.27	-113.49±8.40



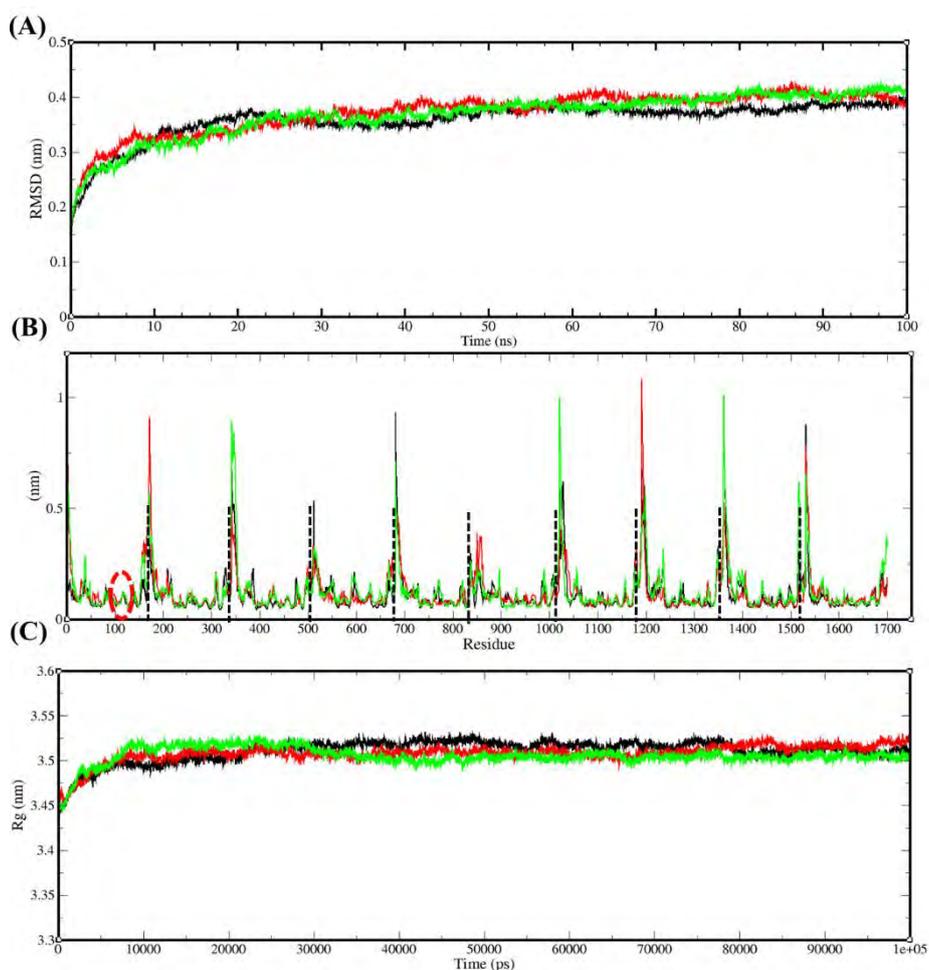
Appendix 1.6. Experimental B -factor plotted and mapped onto the 3D structures of (A) GCH1 (PDB ID: 1WUR) and (B) PTPS (PDB ID: 1Y13). Regions exhibiting the lowest B -factor are shown in blue. Higher values in green, yellow, and red as the most flexible. (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020)).



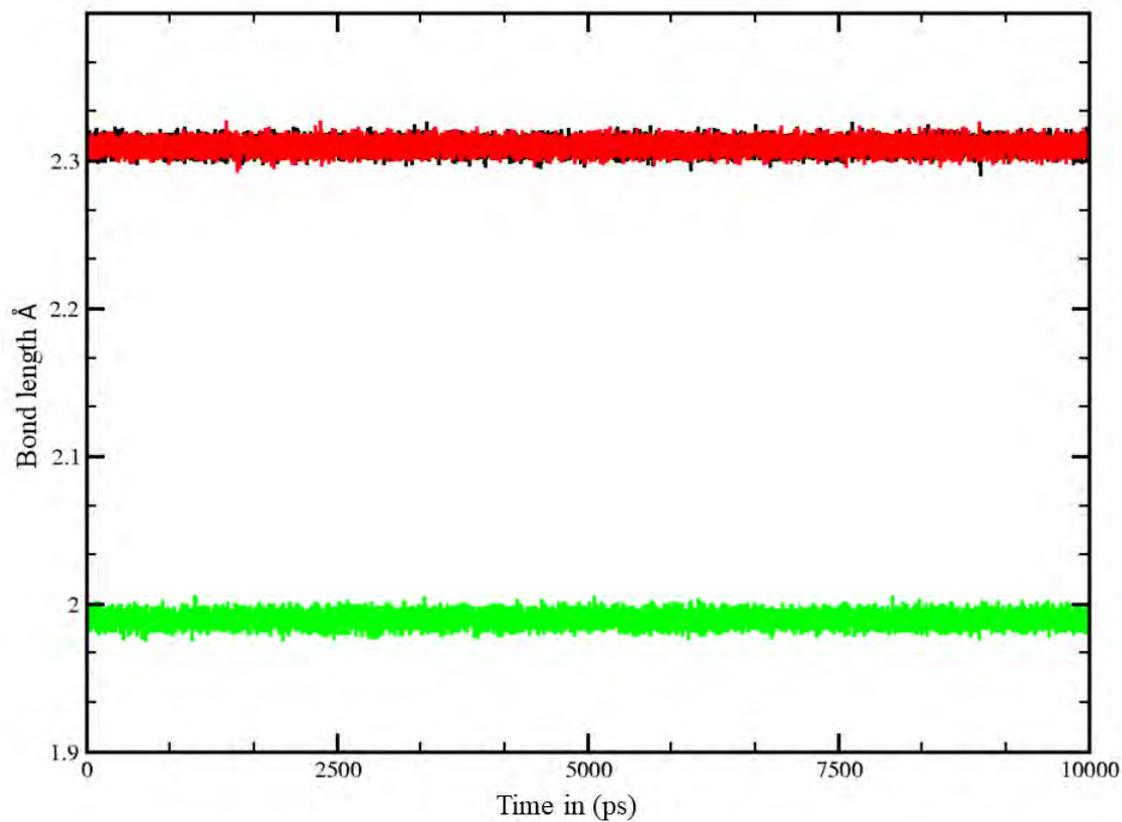
Appendix 1.7. Coordination distance to the Zn^{2+} ion of the GCH1 catalytically important residues His-ND1(110) (Yellow) and His-ND1(177) (Blue) calculated over 100 ns trajectories. (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020)).



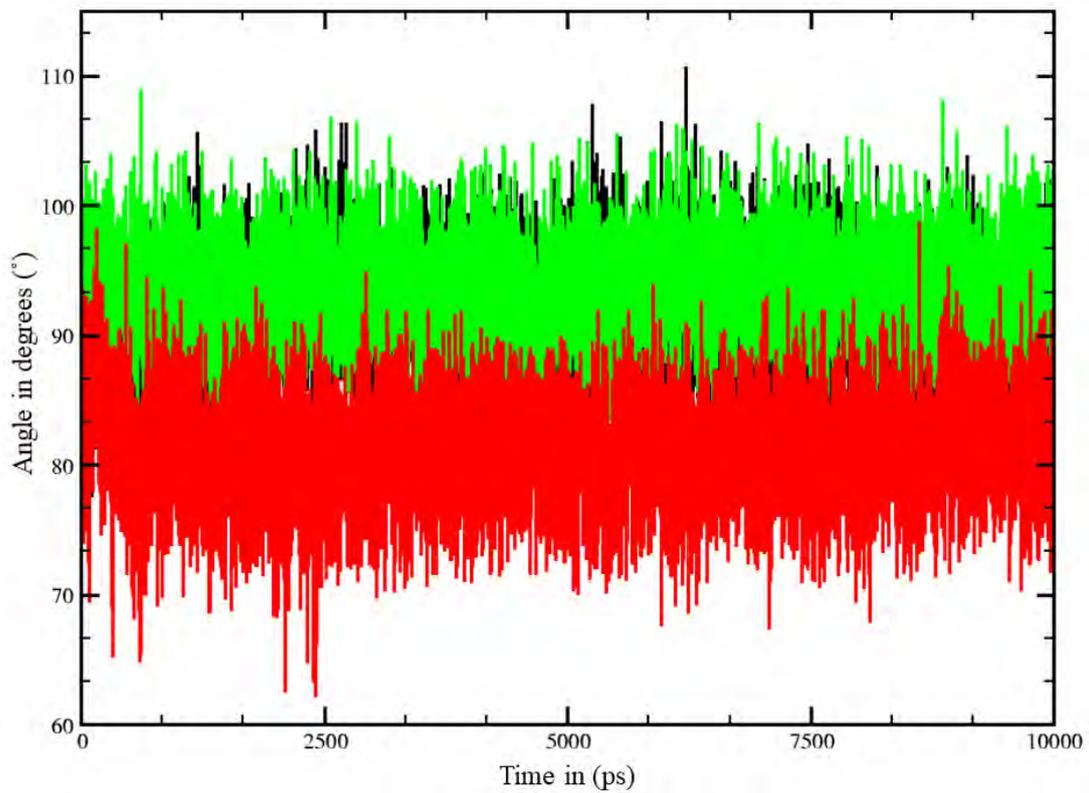
Appendix 1.8. Coordination distance to the Zn^{2+} ion of the PTPS catalytically important residues Glu-_{OE1}(38) (Yellow) and Glu-_{OE1}(161) (Maroon) calculated over 100 ns trajectories. (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020)).



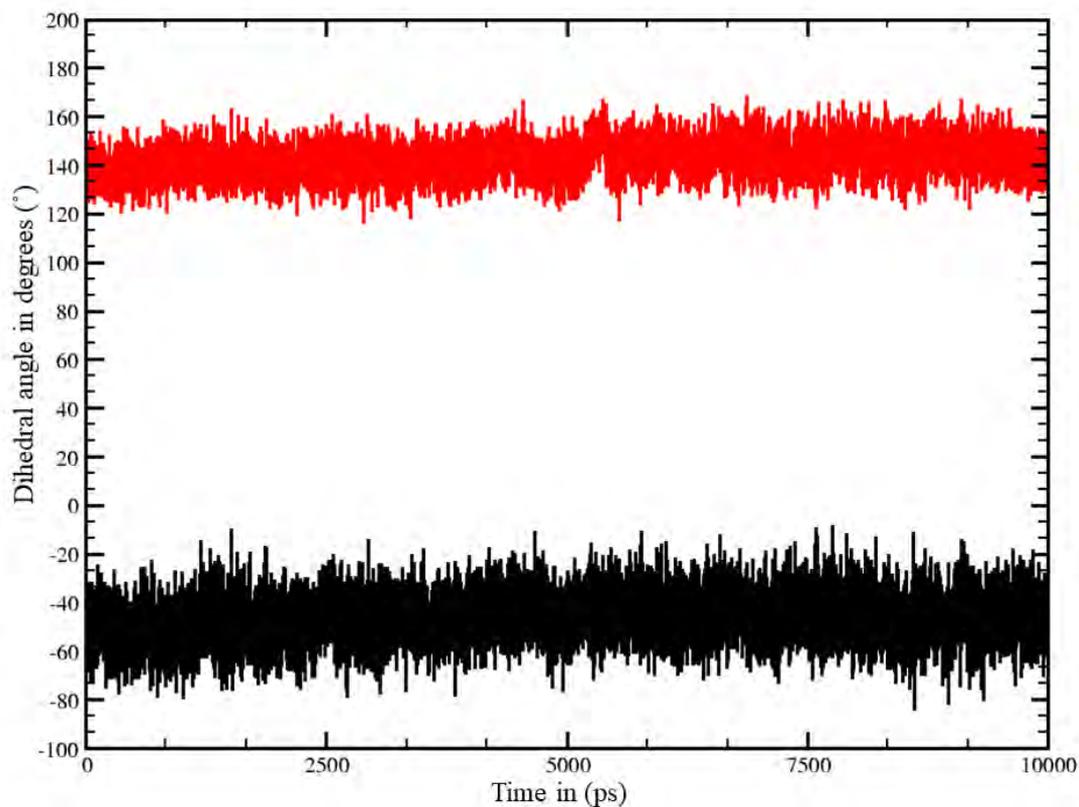
Appendix 1.9. Assessment of the *P. falciparum* GCH1 homology model structure stability by (A) root mean square deviation (RMSD), (B) root mean square fluctuation (RMSF), and (C) radius of gyration (R_g). The black dashed lines separate each chain of the protein. The active site residues responsible for the Zn^{2+} ion coordination are shown in dashed black circles. The MD replicates are shown in black (run 1), red (run 2), and green (run 3). (Reproduced with permission from (Khairallah, Tasthan Bishop and Moses, 2020)).



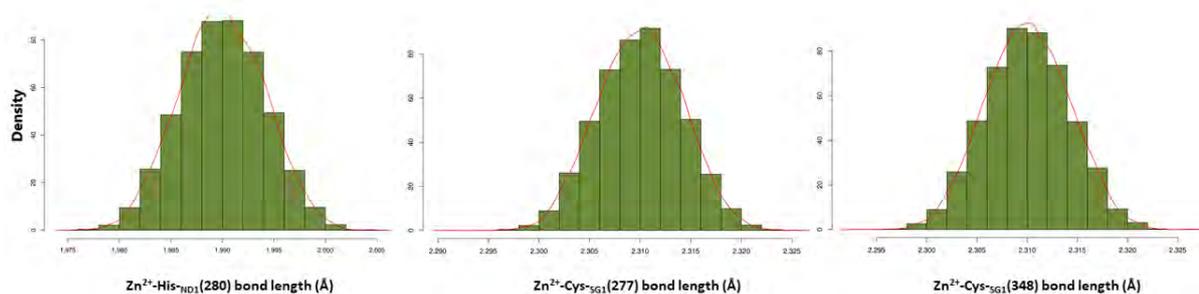
Appendix 1.10. *P. falciparum* GCH1 Zn²⁺ ion coordination distance during the 100 ns MD simulations. The black line represents the coordination distance of Zn²⁺-Cys-SG(277), the red line represents the coordination distance fluctuation of Zn²⁺-Cys-SG(348), and the green line represents the bond distance fluctuation of Zn²⁺-His-ND1(280). (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020)).



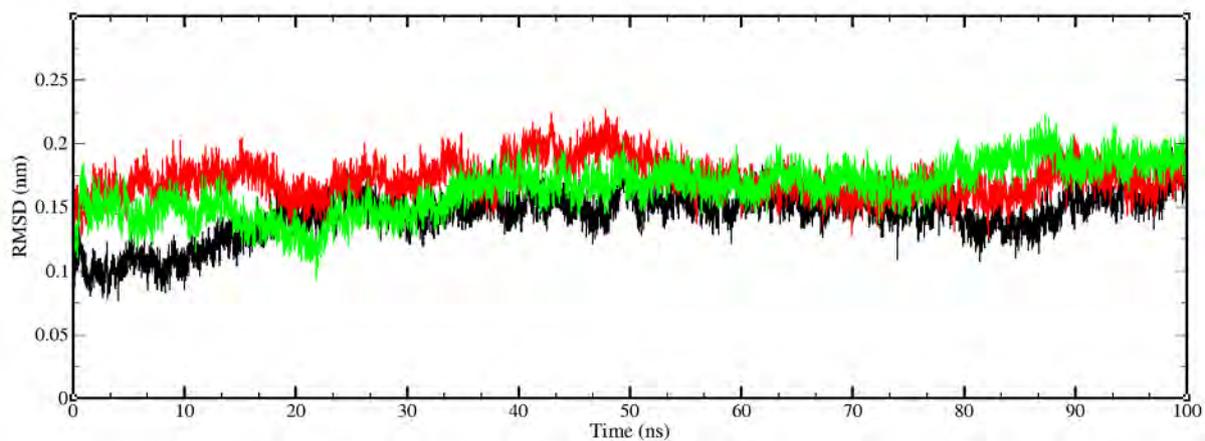
Appendix 1.11. *P. falciparum* GCH1 Zn²⁺ ion coordination angles during the 100 ns MD simulations. The black line represents the coordination angle of Cys-SG(277)-Zn²⁺-His-ND1(280); the green line represents the coordination angle of Cys-SG(348)-Zn²⁺-His-ND1(280), and the red line represents the coordination angle of Cys-SG(277)-Zn²⁺-Cys-SG(348). (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020)).



Appendix 1.12. *P. falciparum* GCH1 Zn^{2+} ion coordination dihedral angles during the 100 ns MD simulations. The red line represents the dihedral angle of Cys-SG(277)- Zn^{2+} - His-ND1(280)-His-CE1(280); the black line represents the dihedral angle of Cys-SG(348)- Zn^{2+} -His-ND1(280)-His-CG (280). (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020)).

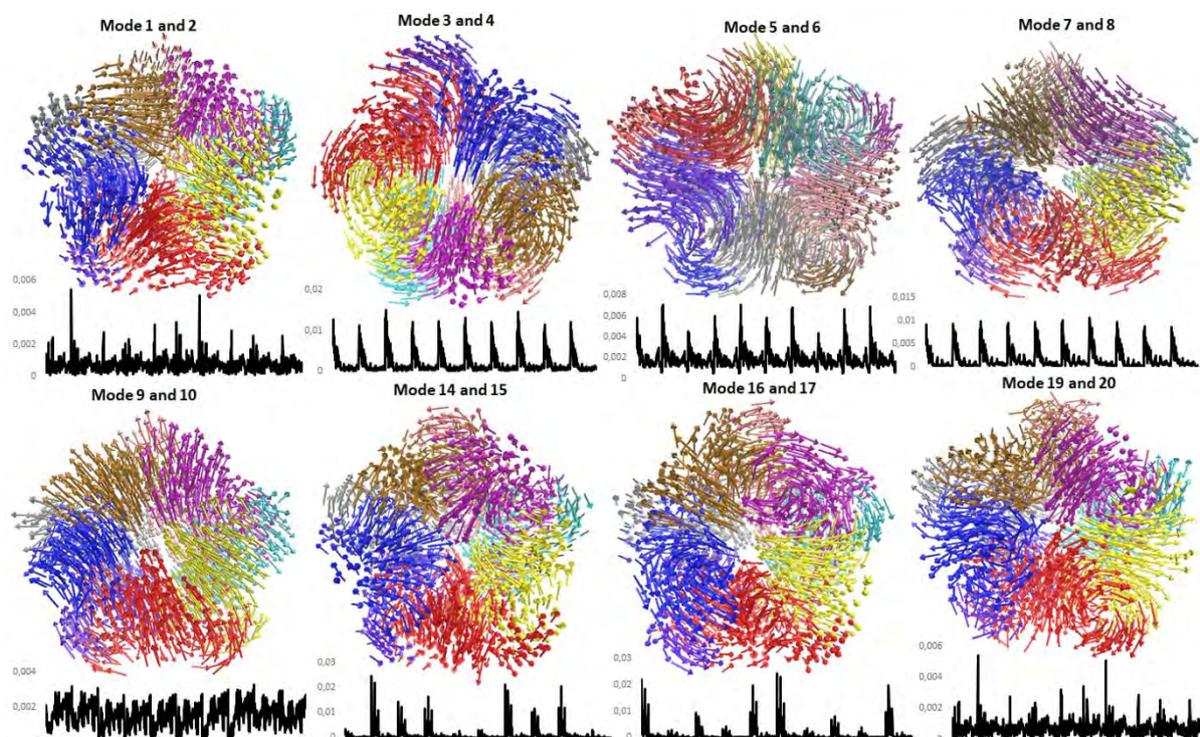


Appendix 1.13. Density histogram showing the distribution and occurrence of the Zn^{2+} ion coordination bond length during the 100 ns MD simulations of the *P. falciparum* GCH1 homology model. The red line is an overlaying density line that is drawn to show the kernel density of the histograms. (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020)).

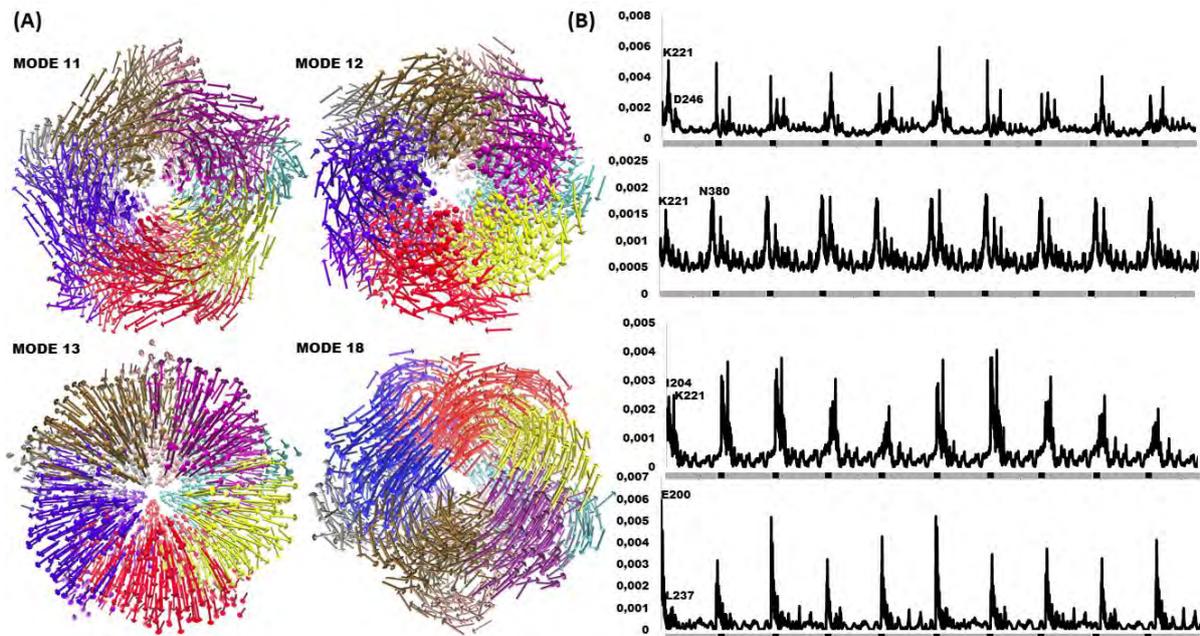


Appendix 1.14. RMSD values of the *P. falciparum* GCH1 Zn^{2+} ion over the 100 ns MD simulations. The MD replicates are shown in black (run 1), red (run 2), and green (run 3). (Reproduced with permission from (Khairallah, Tastan Bishop and Moses, 2020)).

Appendix 2



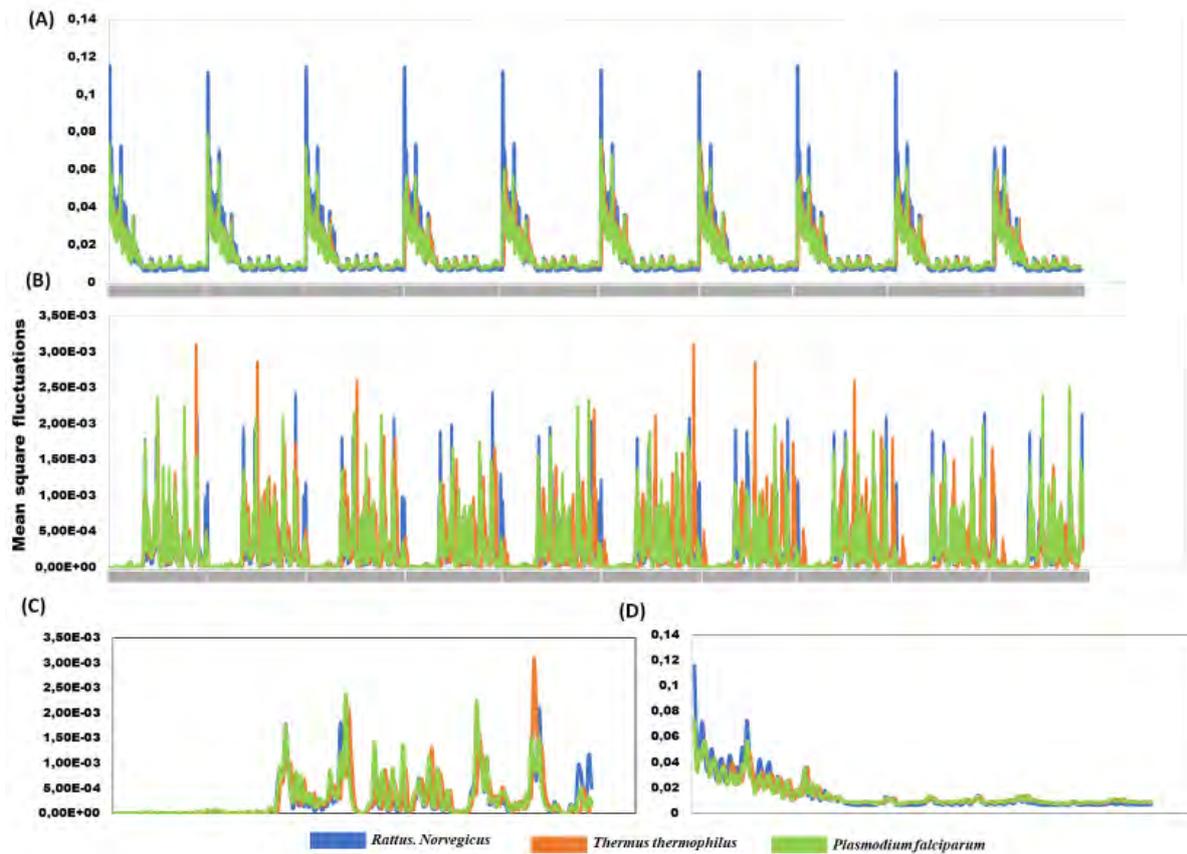
Appendix 2.1. The 3D motions (Top) and atomic MSF (Bottom) obtained from the summation of the degenerate modes that have the same eigenvalues within the first 20 non-trivial low-frequency modes of the *Plasmodium falciparum* GCH1 homology model. Each arrow denotes the protein C β atoms vector of movement. The MSF profiles are shown below each motion. (Reproduced with permission from (Khairallah, Ross and Tasthan Bishop, 2021)).



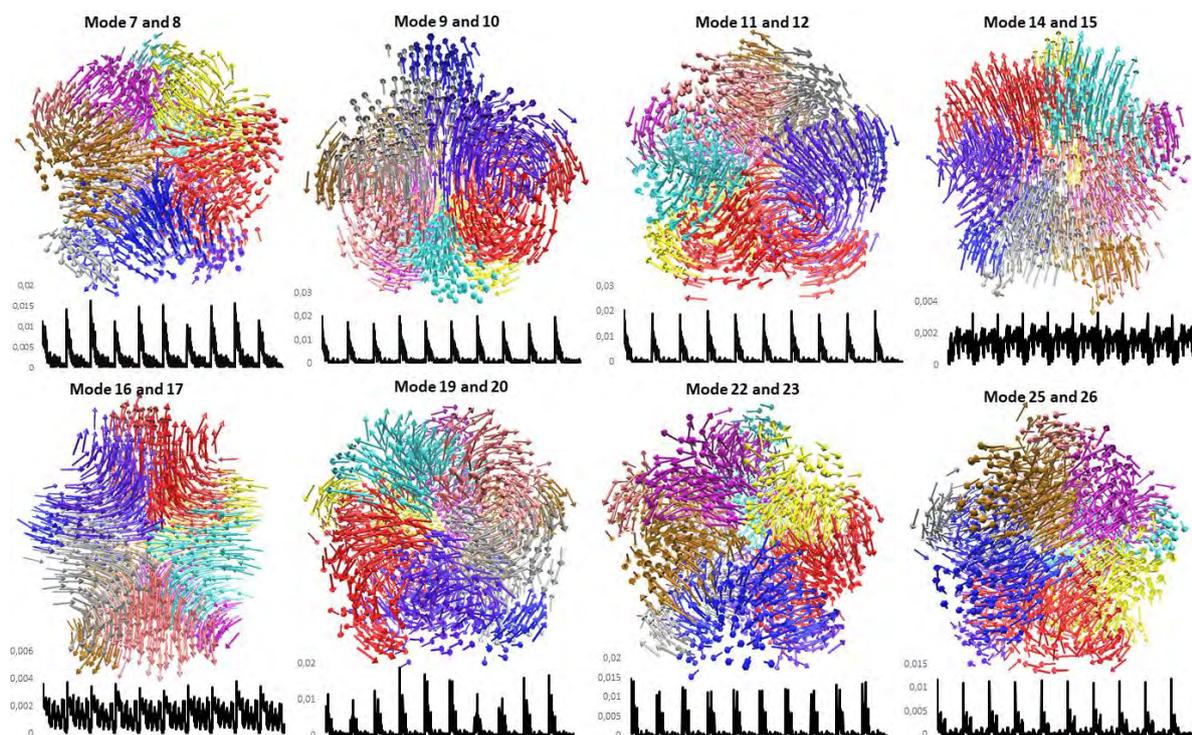
Appendix 2.2. (A) 3D motions identified from the four non-degenerate low-frequency modes of the *P. falciparum* GCH1 homolog. Each arrow denotes the protein C β atoms vector of movement. (B) Individual MSF profiles of the four non-degenerate low-frequency modes, from top to bottom mode 11, 12, 13, and 18. The grey blocks represent each chain of the protein homo- decameric structure. Residues of a notable fluctuation are labelled on the MSF plot. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2021)).

Species	Accession Number
<i>Plasmodium falciparum</i>	PF3D7_1224000
<i>Drosophila melanogaster</i>	E4NKN2
<i>Homo sapiens (Human)</i>	P30793
<i>Rattus norvegicus (Rat)</i>	P22288
<i>Escherichia coli</i>	P0A6T5
<i>Thermus thermophilus</i>	Q5SH52

Appendix 2.3. Organism names and accession numbers of the GTP cyclohydrolase sequences.



Appendix 2.4. MSF of the 20 non-trivial (A) Low-frequency and (B) High-frequency normal modes. (C) One chain representation of the high-frequency normal modes. (D) One chain representation of the low-frequency normal modes. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2021)).



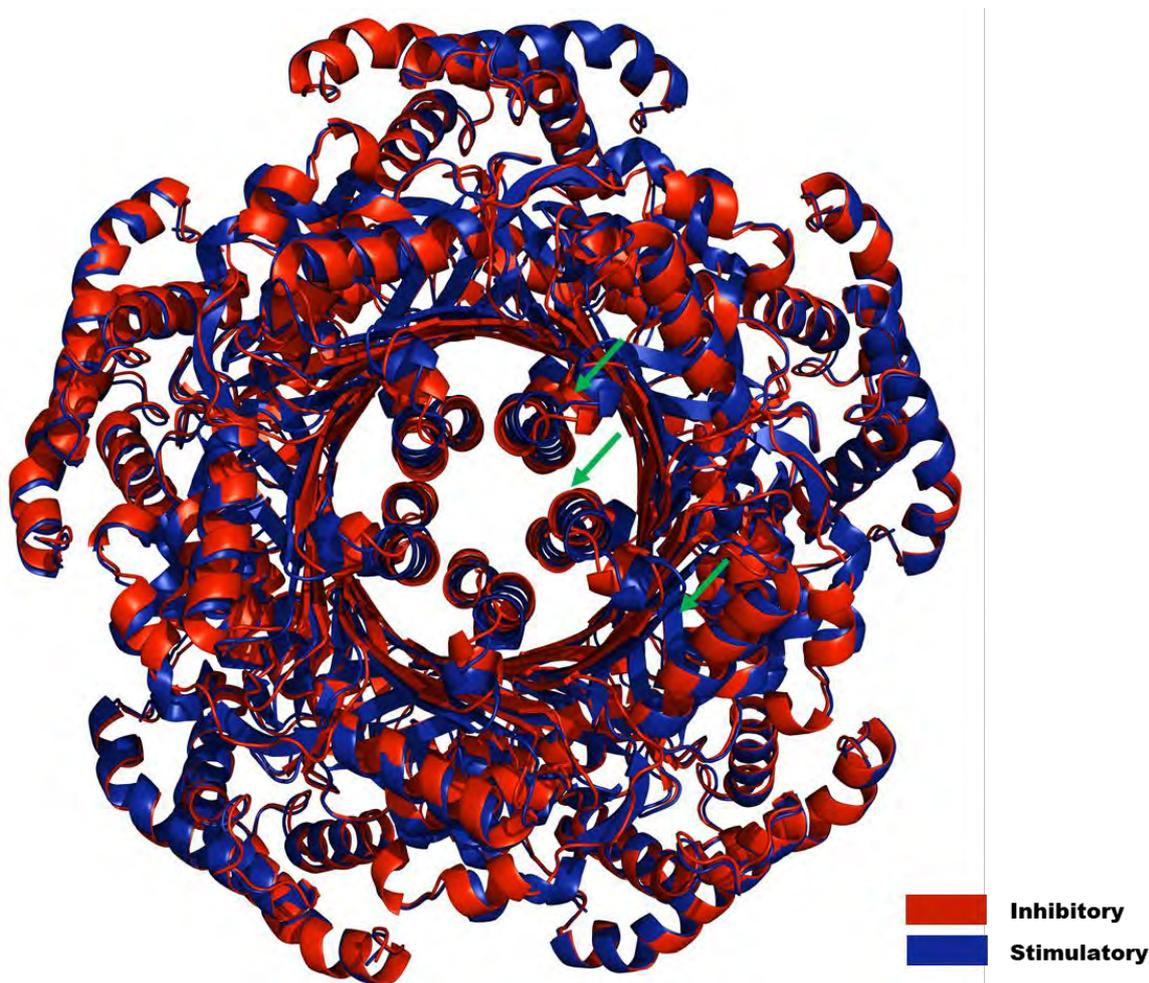
Appendix 2.5. The 3D motions (Top) and atomic MSF (Bottom) obtained from the summation of the degenerate modes that have the same eigenvalues within the 20 non-trivial low-frequency modes of the *Rattus norvegicus* stand-alone GCH1 structure. Each arrow denotes the protein C_{β} atoms vector of movement. The MSF profiles are shown below each motion. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2021)).

Mode	Overlap
Mode 1 and 2	0.00
Mode 3 and 4	0.00
Mode 5 and 6	0.00
Mode 8 and 9	0.01
Mod 10 and 11	0.02
Mod 13 and 14	0.00
Mod 16 and 17	0.00
Mod 19 and 20	0.00

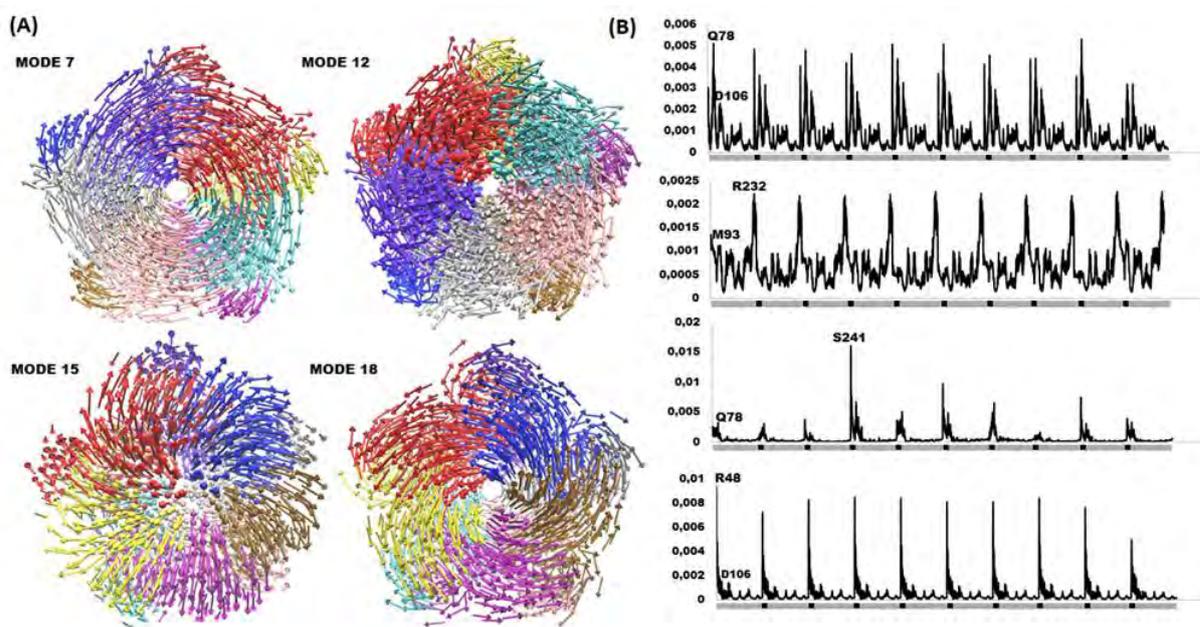
Appendix 2.6. Mode overlap analysis between the predicted motions in the combined degenerate modes within the 20 lowest-frequency modes of the stand-alone GCH1 from the inhibitory complex (PDB ID: 1WPL) and the crystal structure of the stand-alone GCH1 crystal structure from the stimulatory complex (PDB ID:1IS7).

Mode	Overlap
Mode 7	0.10
Mode 12	0.45
Mode 15	0.00
Mode 18	0.00

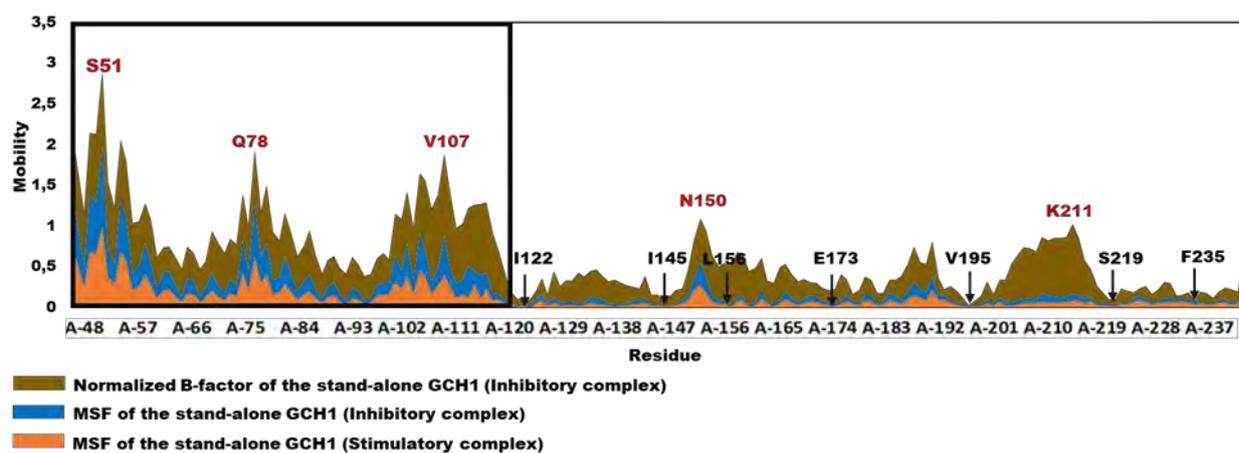
Appendix 2.7. Mode overlap analysis between the predicted motions in the non-degenerate modes within the 20 lowest-frequency modes of the stand-alone GCH1 from the inhibitory complex (PDB ID: 1WPL) and the crystal structure of stand-alone GCH1 from the stimulatory complex (PDB ID:1IS7).



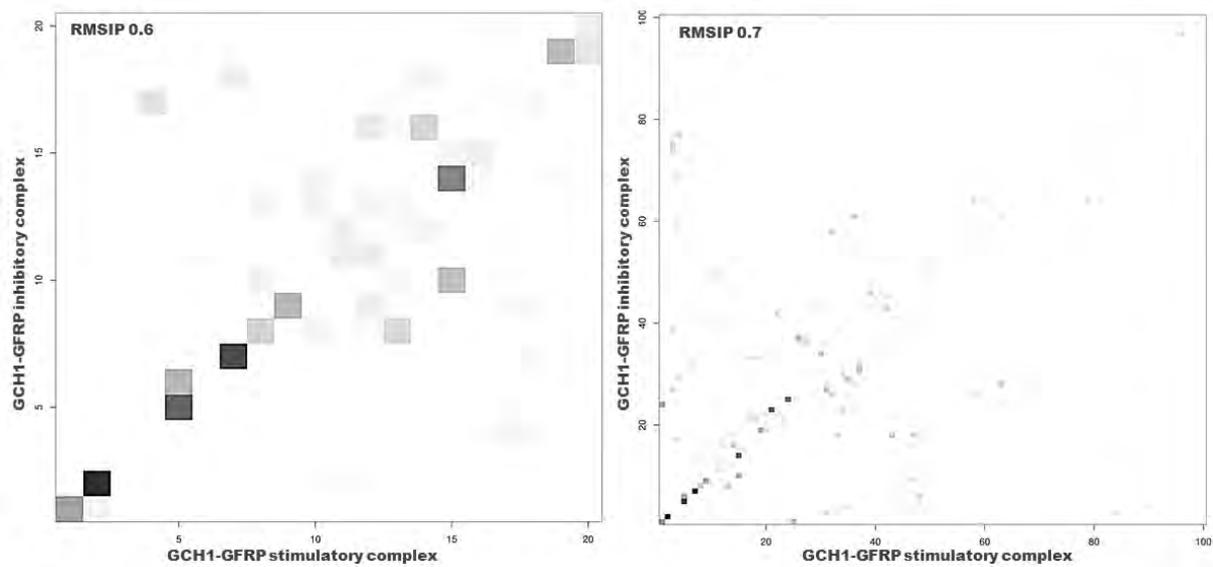
Appendix 2.8. Superimposition of the stand-alone GCH1 extracted from *Rattus norvegicus* GCH1-GFRP inhibitory complex (PDB ID:1WPL) and the stand-alone GCH1 extracted from *Rattus norvegicus* GCH1-GFRP stimulatory complex (PDB ID: 1IS7). The green arrows highlight structural differences (at the tip of the central helices and the last β strand) of the GCH1, which are noted by an overall inward contraction of the GCH1 as it transitions from stimulatory to inhibitory form. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2021)).



Appendix 2.9. (A) The 3D motions identified from the four non-degenerate low-frequency modes of the *R. norvegicus* stand-alone GCH1 structure. Each arrow denotes the protein C_{β} atoms vector of movement. (B) Individual MSF profiles of the four non-degenerate low-frequency modes, from top to bottom mode 7, 12, 15, and 18. The grey blocks represent each chain of the protein homo- decameric structure. Residues of a notable fluctuation are labelled on the MSF plot. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2021)).



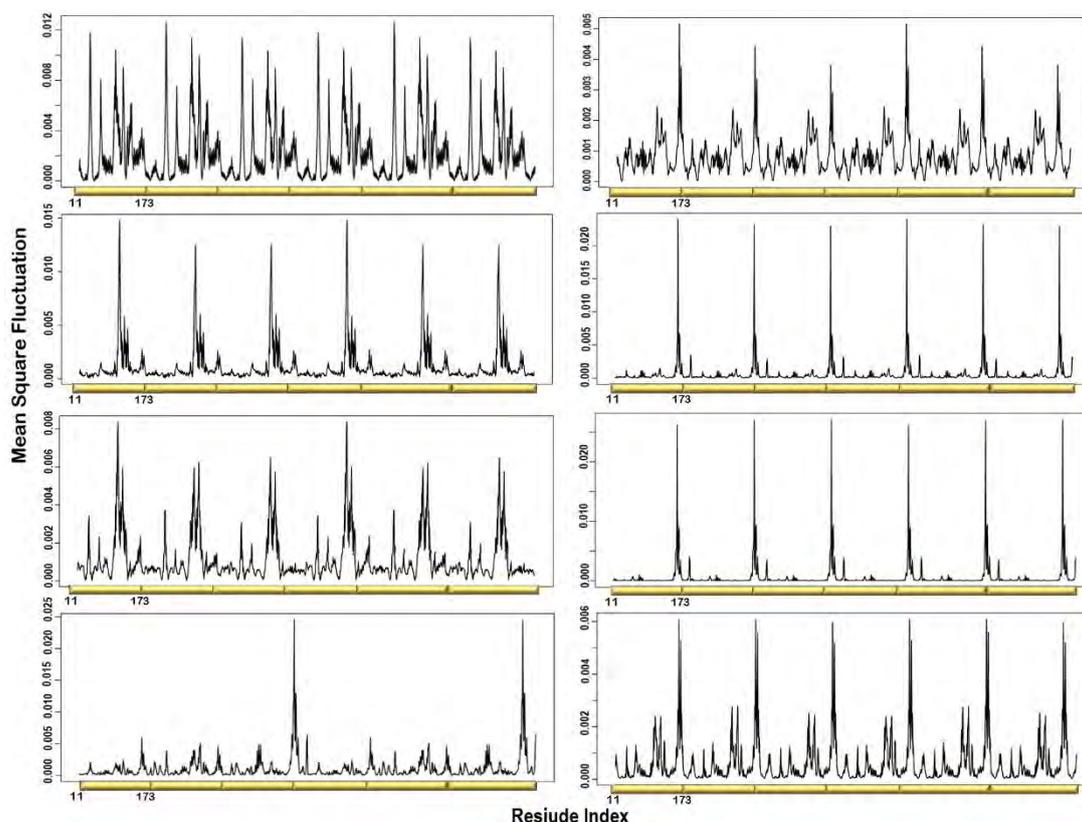
Appendix 2.10. A plot of the normalized experimental B -factor versus the MSF of all modes obtained from the stand-alone GCH1 structure from the inhibitory complex (PDB ID: 1WPL) and (PDB ID: IIS7), respectively. Residues of notable fluctuation are labelled in red; the black arrows indicate hinge residues of restricted mobility. The black box illustrates the GCH1 N-terminus. (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2021)).



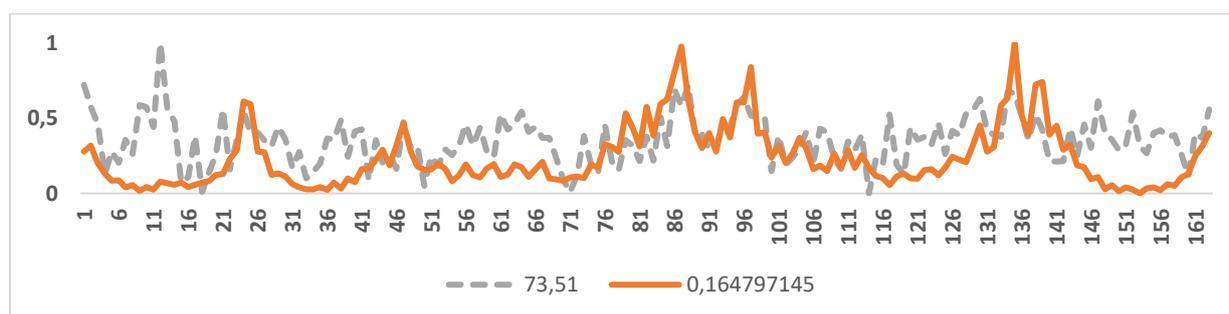
Appendix 2.11. The RMSIP of the first 20 non-trivial modes subspaces (left) and the first 100 non-trivial modes subspaces (right) of the GCH1-GFRP inhibitory complex from *R. norvegicus* (PDB ID: 1WPL), and the GCH1-GFRP stimulatory complex from *R. norvegicus* (PDB ID: 1IS7). (Reproduced with permission from (Khairallah, Ross and Tastan Bishop, 2021)).

Appendix 2.12. Movie illustrations are available at the website: https://rubi.ru.ac.za/GCH1_Modes

Appendix 3



Appendix 3.1. Individual MSF profiles of the eight non-degenerate modes. The yellow blocks represent each chain of the protein structure. The residues index is labelled for one chain along the lower abscissa and according to the *P. falciparum* (PDB ID: 1Y13) file. (Reproduced with permission from (Khairallah, Ross and Tasthan Bishop, 2020)).



Appendix 3.2. *P. falciparum* PTPS (PDB ID: 1Y13) experimental B-factor versus the MSF of all normal modes.

Appendix 3.3. Movie illustrations are available at the website: https://rubi.ru.ac.za/PTPS_Modes

References

- Abraham, M. J. *et al.* (2015) 'GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers', *SoftwareX*. DOI: 10.1016/j.softx.2015.06.001.
- Achan, J. *et al.* (2018) 'Malaria medicines to address drug resistance and support malaria elimination efforts', *Expert Review of Clinical Pharmacology*. DOI: 10.1080/17512433.2018.1387773.
- Aderibigbe, B. A. (2017) 'Design of drug delivery systems containing artemisinin and its derivatives', *Molecules*. DOI: 10.3390/molecules22020323.
- Agarwal, P. K. (2006) 'Enzymes: An integrated view of structure, dynamics and function', *Microbial Cell Factories*. DOI: 10.1186/1475-2859-5-2.
- Alp, N. J. *et al.* (2003) 'Tetrahydrobiopterin-dependent preservation of nitric oxide-mediated endothelial function in diabetes by targeted transgenic GTP-cyclohydrolase I overexpression', *Journal of Clinical Investigation*. DOI: 10.1172/JCI17786.
- Altschul, S. F. *et al.* (1990) 'Basic local alignment search tool', *Journal of Molecular Biology*. DOI: 10.1016/S0022-2836(05)80360-2.
- Amadei, A., Ceruso, M. A. and Di Nola, A. (1999) 'On the convergence of the conformational coordinates basis set obtained by the Essential Dynamics analysis of proteins' molecular dynamics simulations', *Proteins: Structure, Function and Genetics*. DOI: 10.1002/(SICI)1097-0134(19990901)36:4<419::AID-PROT5>3.0.CO;2-U.
- Amamuddy, O. S. *et al.* (2020) 'Integrated computational approaches and tools for allosteric drug discovery', *International Journal of Molecular Sciences*. DOI: 10.3390/ijms21030847.
- Amusengeri, A. and Bishop, Ö. T. (2019) 'Discorhabdin N, a South African natural compound, for Hsp72 and Hsc70 allosteric modulation: Combined study of molecular modeling and dynamic residue network analysis', *Molecules*. DOI: 10.3390/molecules24010188.
- Amusengeri, A., Tata, R. B. and Bishop, Ö. T. (2020) 'Understanding the pyrimethamine drug resistance mechanism via combined molecular dynamics and dynamic residue network analysis', *Molecules*. DOI: 10.3390/molecules25040904.
- Anderson, A. C. and Wright, D. L. (2014) 'Antifolate agents: A patent review (2010-2013)', *Expert Opinion on Therapeutic Patents*. DOI: 10.1517/13543776.2014.898062.
- Antony, H. A. and Parija, S. C. (2016) 'Antimalarial drug resistance: An overview', *Tropical Parasitology*. DOI: 10.4103/2229-5070.175081.
- Apweiler, R. (2008) 'The Universal Protein resource (UniProt)', *Nucleic Acids Research*. DOI: 10.1093/nar/gkm895.
- Armougom, F. *et al.* (2006) 'Expresso: Automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee', *Nucleic Acids Research*. DOI: 10.1093/nar/gkl092.
- Arodola, O. A. and Soliman, M. E. S. (2017) 'Quantum mechanics implementation in drug-design workflows: Does it really help?', *Drug Design, Development and Therapy*. DOI: 10.2147/DDDT.S126344.
- Atilgan, A. R. *et al.* (2001) 'Anisotropy of fluctuation dynamics of proteins with an elastic network model', *Biophysical Journal*. DOI: 10.1016/S0006-3495(01)76033-X.
- Atilgan, A. R., Turgut, D. and Atilgan, C. (2007) 'Screened nonbonded interactions in native proteins manipulate optimal paths for robust residue communication.', *Biophysical journal*. DOI: 10.1529/biophysj.106.099440.

- Auerbach, G *et al.* (2000) 'Zinc plays a key role in human and bacterial GTP cyclohydrolase I.', *Proceedings of the National Academy of Sciences of the United States of America*. DOI: 10.1073/pnas.240463497.
- Auerbach, Günter *et al.* (2000) 'Zinc plays a key role in human and bacterial GTP cyclohydrolase I', *Proceedings of the National Academy of Sciences*. DOI: 10.1073/pnas.240463497.
- Aurrecochea, C. *et al.* (2009) 'PlasmoDB: A functional genomic database for malaria parasites', *Nucleic Acids Research*. DOI: 10.1093/nar/gkn814.
- Baggish, A. L. and Hill, D. R. (2002) 'Antiparasitic agent atovaquone', *Antimicrobial Agents and Chemotherapy*. DOI: 10.1128/AAC.46.5.1163-1173.2002.
- Bahar, I. *et al.* (1998) 'Vibrational Dynamics of Folded Proteins: Significance of Slow and Fast Motions in Relation to Function and Stability', *Physical Review Letters*. DOI: 10.1103/PhysRevLett.80.2733.
- Bahar, I., Lezon, T. R., Yang, L.-W., *et al.* (2010) 'Global Dynamics of Proteins: Bridging Between Structure and Function', *Annual Review of Biophysics*. DOI: 10.1146/annurev.biophys.093008.131258.
- Bahar, I., Lezon, T. R., Bakan, A., *et al.* (2010) 'Normal mode analysis of biomolecular structures: Functional mechanisms of membrane proteins', *Chemical Reviews*. DOI: 10.1021/cr900095e.
- Bahar, I., Atilgan, A. R. and Erman, B. (1997) 'Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential', *Folding and Design*. DOI: 10.1016/S1359-0278(97)00024-2.
- Bahar, I., Chennubhotla, C. and Tobi, D. (2007) 'Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation', *Current Opinion in Structural Biology*. DOI: 10.1016/j.sbi.2007.09.011.
- Bahar, I. and Jernigan, R. L. (1998) 'Vibrational Dynamics of Transfer RNAs: Comparison of the Free and Synthetase-bound Forms', pp. 1–14.
- Bailey, T. L. *et al.* (2015) 'The MEME Suite', *Nucleic Acids Research*. DOI: 10.1093/nar/gkv416.
- Bailey, T. L. and Elkan, C. (1994) 'Fitting a mixture model by expectation maximization to discover motifs in biopolymers.', *Proceedings / ... International Conference on Intelligent Systems for Molecular Biology ; ISMB. International Conference on Intelligent Systems for Molecular Biology*.
- Bailey, T. L. and Gribskov, M. (1998) 'Combining evidence using p-values: Application to sequence homology searches', *Bioinformatics*. DOI: 10.1093/bioinformatics/14.1.48.
- Bakan, A. *et al.* (2014) 'Evol and ProDy for bridging protein sequence evolution and structural dynamics', *Bioinformatics*. DOI: 10.1093/bioinformatics/btu336.
- Bakan, A., Meireles, L. M. and Bahar, I. (2011) 'ProDy: Protein dynamics inferred from theory and experiments', *Bioinformatics*. DOI: 10.1093/bioinformatics/btr168.
- Bao, G. (2002) 'Mechanics of biomolecules', *Journal of the Mechanics and Physics of Solids*. DOI: 10.1016/S0022-5096(02)00035-2.
- Basu, S. and Sahi, P. K. (2017) 'Malaria: An Update', *Indian Journal of Pediatrics*. DOI: 10.1007/s12098-017-2332-2.
- Bauer, J. A., Pavlović, J. and Bauerová-Hlinková, V. (2019) 'Normal Mode Analysis as a Routine Part of a Structural Investigation', *Molecules*. MDPI AG. DOI: 10.3390/molecules24183293.
- Bauserman, M. *et al.* (2019) 'An overview of malaria in pregnancy', *Seminars in Perinatology*. DOI: 10.1053/j.semperi.2019.03.018.

- Bayly, C. I. *et al.* (1993) ‘A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model’, *Journal of Physical Chemistry*. DOI: 10.1021/j100142a004.
- Becke, A. D. (1988) ‘Density-functional exchange-energy approximation with correct asymptotic behavior’, *Physical Review A*. DOI: 10.1103/PhysRevA.38.3098.
- Becker, W. *et al.* (2018) ‘Investigating Protein–Ligand Interactions by Solution Nuclear Magnetic Resonance Spectroscopy’, *ChemPhysChem*. DOI: 10.1002/cphc.201701253.
- Bendall, J. K. *et al.* (2014) ‘Tetrahydrobiopterin in cardiovascular health and disease’, *Antioxidants and Redox Signaling*. DOI: 10.1089/ars.2013.5566.
- Benkert, P., Biasini, M. and Schwede, T. (2011) ‘Toward the estimation of the absolute quality of individual protein structure models’, *Bioinformatics*. DOI: 10.1093/bioinformatics/btq662.
- Berry, A. *et al.* (2006) ‘Prevalence of Plasmodium falciparum cytochrome b gene mutations in isolates imported from Africa, and implications for atovaquone resistance’, *Transactions of the Royal Society of Tropical Medicine and Hygiene*. DOI: 10.1016/j.trstmh.2006.02.004.
- Bhagwat, M. and Aravind, L. (2007) ‘PSI-BLAST tutorial’, *Methods in Molecular Biology*. DOI: 10.1385/1-59745-514-8:177.
- Bhardwaj, V. K. and Purohit, R. (2020) ‘A new insight into protein-protein interactions and the effect of conformational alterations in PCNA’, *International Journal of Biological Macromolecules*. DOI: 10.1016/j.ijbiomac.2020.01.212.
- Bhattacharjee, A. K. and Karle, J. M. (1996) ‘Molecular electronic properties of a series of 4-quinolinecarbinolamines define antimalarial activity profile’, *Journal of Medicinal Chemistry*. DOI: 10.1021/jm960358z.
- Blasco, B., Leroy, Di. and Fidock, D. A. (2017) ‘Antimalarial drug resistance: Linking Plasmodium falciparum parasite biology to the clinic’, *Nature Medicine*. DOI: 10.1038/nm.4381.
- Bochevarov, A. D. *et al.* (2013) ‘Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences’, *International Journal of Quantum Chemistry*. DOI: 10.1002/qua.24481.
- Boratyn, G. M. *et al.* (2013) ‘BLAST: a more efficient report with usability improvements.’, *Nucleic acids research*. DOI: 10.1093/nar/gkt282.
- Born, M. and Oppenheimer, R. (1927) ‘Zur Quantentheorie der Molekeln’, *Annalen der Physik*. DOI: 10.1002/andp.19273892002.
- Bourne, C. R. (2014) ‘Utility of the biosynthetic folate pathway for targets in antimicrobial discovery’, *Antibiotics*. DOI: 10.3390/antibiotics3010001.
- Bowling, K. M. *et al.* (2008) ‘Direct Binding of GTP Cyclohydrolase and Tyrosine Hydroxylase’, *Journal of Biological Chemistry*. DOI: 10.1074/jbc.M802552200.
- Brandt, B. W. and Heringa, J. (2009) ‘webPRC: The Profile Comparer for alignment-based searching of public domain databases’, *Nucleic Acids Research*. DOI: 10.1093/nar/gkp279.
- Bridgford, J. L. *et al.* (2018) ‘Artemisinin kills malaria parasites by damaging proteins and inhibiting the proteasome’, *Nature Communications*. DOI: 10.1038/s41467-018-06221-1.
- Brooks, B. and Karplus, M. (1985) ‘Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme.’, *Proceedings of the National Academy of Sciences of the United States of America*. DOI: 10.1073/pnas.82.15.4995.
- Burger, S. K. *et al.* (2012) ‘Automated parametrization of AMBER force field terms from vibrational

analysis with a focus on functionalizing dinuclear zinc(II) scaffolds', *Journal of Chemical Theory and Computation*. DOI: 10.1021/ct2007742.

Bürgisser, D. M. *et al.* (1994) 'Expression and characterization of recombinant human and rat liver 6-pyruvoyl tetrahydropterin synthase. Modified cysteine residues inhibit the enzyme activity', *European Journal of Biochemistry*. DOI: 10.1111/j.1432-1033.1994.tb19964.x.

Bürgisser, D. M. *et al.* (1995) '6-pyruvoyl tetrahydropterin synthase, an enzyme with a novel type of active site involving both zinc binding and an intersubunit catalytic triad motif; Site-directed mutagenesis of the proposed active center, characterization of the metal binding site and', *Journal of Molecular Biology*. DOI: 10.1006/jmbi.1995.0558.

Buxbaum, E. (2007) *Fundamentals of protein structure and function, Fundamentals of Protein Structure and Function*. DOI: 10.1007/978-0-387-68480-2.

Bylaska, E. J. *et al.* (2009) 'Hard scaling challenges for ab initio molecular dynamics capabilities in NWChem: Using 100,000 CPUs per second', in *Journal of Physics: Conference Series*. DOI: 10.1088/1742-6596/180/1/012028.

Bylaska, E. J. *et al.* (2017) 'Performance evaluation of nwchem ab-initio molecular dynamics (AIMD) simulations on the intel® xeon phi™ processor', in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. DOI: 10.1007/978-3-319-67630-2_30.

Di Cera, E. (2020) 'Mechanisms of ligand binding', *Biophysics Reviews*. AIP Publishing. DOI: 10.1063/5.0020997.

Chan-Yao-Chong, M., Durand, D. and Ha-Duong, T. (2019) 'Molecular Dynamics Simulations Combined with Nuclear Magnetic Resonance and/or Small-Angle X-ray Scattering Data for Characterizing Intrinsically Disordered Protein Conformational Ensembles', *Journal of Chemical Information and Modeling*. DOI: 10.1021/acs.jcim.8b00928.

Chan, B. (2017) 'Use of Low-Cost Quantum Chemistry Procedures for Geometry Optimization and Vibrational Frequency Calculations: Determination of Frequency Scale Factors and Application to Reactions of Large Systems', *Journal of Chemical Theory and Computation*. DOI: 10.1021/acs.jctc.7b00721.

Chatzou, M. *et al.* (2016) 'Multiple sequence alignment modeling: Methods and applications', *Briefings in Bioinformatics*. DOI: 10.1093/BIB/BBV099.

Chennubhotla, C. *et al.* (2005) 'Elastic network models for understanding biomolecular machinery: From enzymes to supramolecular assemblies', in *Physical Biology*. DOI: 10.1088/1478-3975/2/4/S12.

Choi, M. ki *et al.* (2017) 'Dynamic characteristics of a flagellar motor protein analyzed using an elastic network model', *Journal of Molecular Graphics and Modelling*. DOI: 10.1016/j.jmgm.2017.10.001.

Chowdhury, B. and Garai, G. (2017) 'A review on multiple sequence alignment from the perspective of genetic algorithm', *Genomics*. DOI: 10.1016/j.ygeno.2017.06.007.

Chuaiphichai, S. *et al.* (2014) 'Cell-autonomous role of endothelial GTP cyclohydrolase 1 and tetrahydrobiopterin in blood pressure regulation', *Hypertension*. DOI: 10.1161/HYPERTENSIONAHA.114.03089.

Chughlay, M. F. *et al.* (2020) 'First-in-human clinical trial to assess the safety, tolerability and pharmacokinetics of P218, a novel candidate for malaria chemoprotection', *British Journal of Clinical Pharmacology*. DOI: 10.1111/bcp.14219.

Cieplak, P. *et al.* (1995) 'Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins', *Journal of*

Computational Chemistry. DOI: 10.1002/jcc.540161106.

Clare, G. M. and Gronenborn, A. M. (1989) 'Determination of three-dimensional structures of proteins and nucleic acids in solution by nuclear magnetic resonance spectroscopy', *Critical Reviews in Biochemistry and Molecular Biology*. DOI: 10.3109/10409238909086962.

Colloc'h, N., Poupon, A. and Mornon, J.-P. (2000) 'Sequence and structural features of the T-fold, an original tunnelling building unit', *Proteins: Structure, Function, and Genetics*. DOI: 10.1002/(SICI)1097-0134(20000501)39:2<142::AID-PROT4>3.0.CO;2-X.

Conrad, M. D. and Rosenthal, P. J. (2019) 'Antimalarial drug resistance in Africa: the calm before the storm?', *The Lancet Infectious Diseases*. DOI: 10.1016/S1473-3099(19)30261-0.

Cornell, W. D. *et al.* (1993) 'Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation', *Journal of the American Chemical Society*. DOI: 10.1021/ja00074a030.

Cowell, A. N. and Winzeler, E. A. (2019) 'The genomic architecture of antimalarial drug resistance', *Briefings in Functional Genomics*. DOI: 10.1093/bfgp/elz008.

de Crécy-Lagard, V. *et al.* (2007) 'Comparative genomics of bacterial and plant folate synthesis and salvage: Predictions and validations', *BMC Genomics*. DOI: 10.1186/1471-2164-8-245.

Daniels, K. G., Suo, Y. and Oas, T. G. (2015) 'Conformational kinetics reveals affinities of protein conformational states', *Proceedings of the National Academy of Sciences of the United States of America*. DOI: 10.1073/pnas.1502084112.

Darden, T., York, D. and Pedersen, L. (1993) 'Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems', *The Journal of Chemical Physics*. DOI: 10.1063/1.464397.

Daugelaite, J., O' Driscoll, A. and Sleator, R. D. (2013) 'An Overview of Multiple Sequence Alignments and Cloud Computing in Bioinformatics', *ISRN Biomathematics*. DOI: 10.1155/2013/615630.

Dayasiri, K. C. *et al.* (2019) 'GTP-Cyclohydrolase i deficiency presenting as malignant hyperphenylalaninemia, recurrent hyperthermia and progressive neurological dysfunction in a South Asian child - A case report', *BMC Pediatrics*. DOI: 10.1186/s12887-019-1580-x.

Dayhoff, M., Eck, R. and Park, C. (1972) 'A Model of Evolutionary Change in Proteins Accepted Point Mutations', in *Atlas of protein sequence and structure*.

Dayhoff, M. O., Schwartz, R. M. and Orcutt, B. C. (1978) 'A model of evolutionary change in proteins', in *Atlas of protein sequence and structure*.

DeLano L, W. (2002) 'PyMOL: An Open-Source Molecular Graphics Tool', *Ccp4 Newsletter on Protein Crystallography*.

DeLano, W. L. (2014) 'The PyMOL Molecular Graphics System, Version 1.8', *Schrödinger LLC*, p. <http://www.pymol.org>. DOI: 10.1038/hr.2014.17.

Delarue, M. and Dumas, P. (2004) 'On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models', *Proceedings of the National Academy of Sciences of the United States of America*. DOI: 10.1073/pnas.0400301101.

Delves, M. *et al.* (2019) 'Fueling Open Innovation for Malaria Transmission-Blocking Drugs: Hundreds of Molecules Targeting Early Parasite Mosquito Stages', *Frontiers in Microbiology*. DOI: 10.3389/fmicb.2019.02134.

Demerdash, O. N. A., Daily, M. D. and Mitchell, J. C. (2009) 'Structure-based predictive models for allosteric hot spots', *PLoS Computational Biology*. DOI: 10.1371/journal.pcbi.1000531.

- Demirel, M. C. *et al.* (1998) 'Identification of kinetically hot residues in proteins', *Protein Science*. DOI: 10.1002/pro.5560071205.
- Depiereux, E. and Feytmans, E. (1992) 'Match-box: A fundamentally new algorithm for the simultaneous alignment of several protein sequences', *Bioinformatics*. DOI: 10.1093/bioinformatics/8.5.501.
- Ditchfield, R., Hehre, W. J. and Pople, J. A. (1971) 'Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules', *The Journal of Chemical Physics*. DOI: 10.1063/1.1674902.
- Dittrich, S. *et al.* (2008) 'An atypical orthologue of 6-pyruvoyltetrahydropterin synthase can provide the missing link in the folate biosynthesis pathway of malaria parasites', *Molecular Microbiology*. DOI: 10.1111/j.1365-2958.2007.06073.x.
- Dokomajilar, C. *et al.* (2006) 'Selection of Plasmodium falciparum pfmdr1 alleles following therapy with artemether-lumefantrine in an area of Uganda where malaria is highly endemic', *Antimicrobial Agents and Chemotherapy*. DOI: 10.1128/AAC.50.5.1893-1895.2006.
- Donini, O. A. T. and Kollman, P. A. (2000) 'Calculation and Prediction of Binding Free Energies for the Matrix Metalloproteinases', *Journal of Medicinal Chemistry*. DOI: 10.1021/jm000040d.
- Doruker, P., Atilgan, A. R. and Bahar, I. (2000) 'Dynamics of proteins predicted by molecular simulations and analytical approaches: Application to α -amylase inhibitor', *Proteins: Structure, Function and Genetics*. DOI: 10.1002/1097-0134(20000815)40:3<512::AID-PROT180>3.0.CO;2-M.
- Draper, S. J. *et al.* (2018) 'Malaria Vaccines: Recent Advances and New Horizons', *Cell Host and Microbe*. DOI: 10.1016/j.chom.2018.06.008.
- Duan, Y. *et al.* (2003) 'A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations', *Journal of Computational Chemistry*. DOI: 10.1002/jcc.10349.
- Duarte, F. *et al.* (2014) 'Force field independent metal parameters using a nonbonded dummy model', *Journal of Physical Chemistry B*. DOI: 10.1021/jp501737x.
- Ducker, G. S. and Rabinowitz, J. D. (2017) 'One-Carbon Metabolism in Health and Disease', *Cell Metabolism*. DOI: 10.1016/j.cmet.2016.08.009.
- Duffy, P. E. and Patrick Gorres, J. (2020) 'Malaria vaccines since 2000: progress, priorities, products', *npj Vaccines*. DOI: 10.1038/s41541-020-0196-3.
- Dutta, A. and Bahar, I. (2010) 'Metal-binding sites are designed to achieve optimal mechanical and signaling properties', *Structure*. DOI: 10.1016/j.str.2010.06.013.
- Ebenhoch, R. *et al.* (2020) 'A hybrid approach reveals the allosteric regulation of GTP cyclohydrolase I', *Proceedings of the National Academy of Sciences*. DOI: 10.1073/pnas.2013473117.
- Ebert, J. and Brutlag, D. (2006) 'Development and validation of a consistency based multiple structure alignment algorithm', *Bioinformatics*. DOI: 10.1093/bioinformatics/btl046.
- Edgar, R. C. (2004) 'MUSCLE: A multiple sequence alignment method with reduced time and space complexity', *BMC Bioinformatics*. DOI: 10.1186/1471-2105-5-113.
- Eisenmesser, E. Z. *et al.* (2005) 'Intrinsic dynamics of an enzyme underlies catalysis', *Nature*. DOI: 10.1038/nature04105.
- Elstner, M. *et al.* (2003) 'Modeling zinc in biomolecules with the self consistent charge-density functional tight binding (SCC-DFTB) method: Applications to structural and energetic analysis', *Journal of Computational Chemistry*. DOI: 10.1002/jcc.10201.

- Emekli, U. *et al.* (2008) ‘HingeProt: Automated prediction of hinges in protein structures’, *Proteins: Structure, Function and Genetics*. DOI: 10.1002/prot.21613.
- Eyal, E. and Bahar, I. (2008) ‘Toward a molecular understanding of the anisotropic response of proteins to external forces: Insights from elastic network models’, *Biophysical Journal*. DOI: 10.1529/biophysj.107.120733.
- Fenwick, R. B., Esteban-Martín, S. and Salvatella, X. (2011) ‘Understanding biomolecular motion, recognition, and allostery by use of conformational ensembles’, *European Biophysics Journal*. DOI: 10.1007/s00249-011-0754-8.
- Finn, R. D., Clements, J. and Eddy, S. R. (2011) ‘HMMER web server: Interactive sequence similarity searching’, *Nucleic Acids Research*. DOI: 10.1093/nar/gkr367.
- Flores, S. C. *et al.* (2007) ‘Hinge Atlas: Relating protein sequence to sites of structural flexibility’, *BMC Bioinformatics*. DOI: 10.1186/1471-2105-8-167.
- Frisch, M. J. *et al.* (2009) ‘Gaussian 09, Revision B.01’, *Gaussian 09, Revision B.01*, Gaussian, Inc., Wallingford CT.
- Frisch, M. J. *et al.* (2010) ‘Gaussian09 Revision D.01, Gaussian Inc. Wallingford CT’, *Gaussian 09 Revision C.01*.
- Frischknecht, F. and Matuschewski, K. (2017) ‘Plasmodium sporozoite biology’, *Cold Spring Harbor Perspectives in Medicine*. DOI: 10.1101/cshperspect.a025478.
- Fröhlking, T. *et al.* (2020) ‘Toward empirical force fields that match experimental observables’, *Journal of Chemical Physics*. DOI: 10.1063/5.0011346.
- Froimowitz, M. (1993) ‘HyperChem(TM): A software package for computational chemistry and molecular modeling’, *BioTechniques*.
- Fuglebakk, E., Tiwari, S. P. and Reuter, N. (2015) ‘Comparing the intrinsic dynamics of multiple protein structures using elastic network models’, *Biochimica et Biophysica Acta - General Subjects*. DOI: 10.1016/j.bbagen.2014.09.021.
- Funderburk, C. D. *et al.* (2006) ‘A typical N-terminal Extensions Confer Novel Regulatory Properties on GTP Cyclohydrolase Isoforms in Drosophila melanogaster’, *Journal of Biological Chemistry*. DOI: 10.1074/jbc.M602196200.
- Gaalswyk, K., Muniyat, M. I. and MacCallum, J. L. (2018) ‘The emerging role of physical modeling in the future of structure determination’, *Current Opinion in Structural Biology*. DOI: 10.1016/j.sbi.2018.03.005.
- Galactionova, K. *et al.* (2015) ‘Effective coverage and systems effectiveness for malaria case management in Sub-Saharan African countries’, *PLoS ONE*. DOI: 10.1371/journal.pone.0127818.
- Gall, A. *et al.* (2015) ‘Conformational Switching in a Light-Harvesting Protein as Followed by Single-Molecule Spectroscopy’, *Biophysical Journal*. DOI: 10.1016/j.bpj.2015.04.017.
- Go, N., Noguti, T. and Nishikawa, T. (1983) ‘Dynamics of a small globular protein in terms of low-frequency vibrational modes.’, *Proceedings of the National Academy of Sciences of the United States of America*. DOI: 10.1073/pnas.80.12.3696.
- Göbl, C. *et al.* (2014) ‘NMR approaches for structural analysis of multidomain proteins and complexes in solution’, *Progress in Nuclear Magnetic Resonance Spectroscopy*. DOI: 10.1016/j.pnmrs.2014.05.003.
- Grant, B. J. *et al.* (2006) ‘Bio3d: An R package for the comparative analysis of protein structures’, *Bioinformatics*. DOI: 10.1093/bioinformatics/btl461.

- Grant, B. J., Gorfe, A. A. and McCammon, J. A. (2010) 'Large conformational changes in proteins: Signaling and other functions', *Current Opinion in Structural Biology*. DOI: 10.1016/j.sbi.2009.12.004.
- Gräwert, T., Fischer, M. and Bacher, A. (2013) 'Structures and reaction mechanisms of GTP cyclohydrolases', *IUBMB Life*. DOI: 10.1002/iub.1153.
- Gresh, N., Piquemal, J. P. and Krauss, M. (2005) 'Representation of Zn(II) complexes in polarizable molecular mechanics. Further refinements of the electrostatic and short-range contributions. Comparisons with parallel Ab initio computations', *Journal of Computational Chemistry*. DOI: 10.1002/jcc.20244.
- Griffith, E. C. *et al.* (2018) 'The structural and functional basis for recurring sulfa drug resistance mutations in *Staphylococcus aureus* dihydropteroate synthase', *Frontiers in Microbiology*. DOI: 10.3389/fmicb.2018.01369.
- Grimberg, B. T. and Mehlotra, R. K. (2011) 'Expanding the antimalarial drug arsenal-now, but how?', *Pharmaceuticals*. DOI: 10.3390/ph4050681.
- Groenhof, G. (2013) 'Introduction to QM/MM simulations', *Methods in Molecular Biology*. DOI: 10.1007/978-1-62703-17-5_3.
- Gross, E. K. U. and Kohn, W. (1990) 'Time-dependent density-functional theory', *Advances in Quantum Chemistry*. DOI: 10.1016/S0065-3276(08)60600-0.
- Guarnera, E. and Berezovsky, I. N. (2020) 'Allosteric drugs and mutations: chances, challenges, and necessity', *Current Opinion in Structural Biology*. DOI: 10.1016/j.sbi.2020.01.010.
- Gunasekaran, K., Ma, B. and Nussinov, R. (2004) 'Is allostery an intrinsic property of all dynamic proteins?', *Proteins: Structure, Function and Genetics*. DOI: 10.1002/prot.20232.
- Guo, J. and Zhou, H. X. (2016) 'Protein Allostery and Conformational Dynamics', *Chemical Reviews*. DOI: 10.1021/acs.chemrev.5b00590.
- H, F. *et al.* (2020) 'Tetrahydrobiopterin (BH4) pathway: from metabolism to neuropsychiatry', *Current Neuropharmacology*. DOI: 10.2174/1570159x18666200729103529.
- Haas, K. L. and Franz, K. J. (2009) 'Application of metal coordination chemistry to explore and manipulate cell biology', *Chemical Reviews*. DOI: 10.1021/cr900134a.
- Hafner, J. (2008) 'Ab-initio simulations of materials using VASP: Density-functional theory and beyond', *Journal of Computational Chemistry*. DOI: 10.1002/jcc.21057.
- Haliloglu, T. *et al.* (2005) 'How Similar Are Protein Folding and Protein Binding Nuclei? Examination of Vibrational Motions of Energy Hot Spots and Conserved Residues', *Biophysical Journal*. Elsevier. DOI: 10.1529/biophysj.104.051342.
- Haliloglu, T., Bahar, I. and Erman, B. (1997) 'Gaussian dynamics of folded proteins', *Physical Review Letters*. DOI: 10.1103/PhysRevLett.79.3090.
- Han, X. *et al.* (2018) 'Study of the variability of the native protein structure', in *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*. DOI: 10.1016/B978-0-12-809633-8.20148-9.
- Hanboonkunupakarn, B. and White, N. J. (2015) 'The threat of antimalarial drug resistance', *Tropical Diseases, Travel Medicine and Vaccines*. DOI: 10.1186/s40794-016-0027-8.
- Harada, T., Kagamiyama, H. and Hatakeyama, K. (1993) 'Feedback regulation mechanisms for the control of GTP cyclohydrolase I activity', *Science*. DOI: 10.1126/science.8502995.
- Harpole, T. J. and Delemotte, L. (2018) 'Conformational landscapes of membrane proteins delineated

- by enhanced sampling molecular dynamics simulations', *Biochimica et Biophysica Acta - Biomembranes*. DOI: 10.1016/j.bbamem.2017.10.033.
- Haspel, N. *et al.* (2010) 'Tracing conformational changes in proteins', *BMC Structural Biology*. DOI: 10.1186/1472-6807-10-S1-S1.
- Hatherley, R. *et al.* (2015) 'Plasmodium falciparum Hop: Detailed analysis on complex formation with Hsp70 and Hsp90', *Biochemical and Biophysical Research Communications*. DOI: 10.1016/j.bbrc.2014.11.103.
- Hatherley, R. *et al.* (2016) 'PRIMO: An interactive homology modeling pipeline', *PLoS ONE*. DOI: 10.1371/journal.pone.0166698.
- Hehre, W. J., Ditchfield, K. and Pople, J. A. (1972) 'Self-consistent molecular orbital methods. XII. Further extensions of gaussian-type basis sets for use in molecular orbital studies of organic molecules', *The Journal of Chemical Physics*. DOI: 10.1063/1.1677527.
- Heinberg, A. *et al.* (2013) 'Direct evidence for the adaptive role of copy number variation on antifolate susceptibility in Plasmodium falciparum', *Molecular Microbiology*. DOI: 10.1111/mmi.12162.
- Heinberg, A. and Kirkman, L. (2015) 'The molecular basis of antifolate resistance in Plasmodium falciparum: Looking beyond point mutations', *Annals of the New York Academy of Sciences*. DOI: 10.1111/nyas.12662.
- Henzler-Wildman, K. and Kern, D. (2007) 'Dynamic personalities of proteins', *Nature*. DOI: 10.1038/nature06522.
- Higgins, C. E. and Gross, S. S. (2011) 'The N-terminal peptide of mammalian GTP cyclohydrolase I is an autoinhibitory control element and contributes to binding the allosteric regulatory protein GFRP', *Journal of Biological Chemistry*. DOI: 10.1074/jbc.M110.196204.
- Hildebrand, A. *et al.* (2009) 'Fast and accurate automatic structure prediction with HHpred', *Proteins: Structure, Function and Bioinformatics*. DOI: 10.1002/prot.22499.
- Hinsen, K. (1998) 'Analysis of domain motions by approximate normal mode calculations', *Proteins: Structure, Function and Genetics*. DOI: 10.1002/(SICI)1097-0134(19981115)33:3<417::AID-PROT10>3.0.CO;2-8.
- Hinsen, K., Thomas, A. and Field, M. J. (1999) 'Analysis of domain motions in large proteins', *Proteins: Structure, Function and Genetics*. DOI: 10.1002/(SICI)1097-0134(19990215)34:3<369::AID-PROT9>3.0.CO;2-F.
- Hohenberg, P. and Kohn, W. (1964) 'Inhomogeneous electron gas', *Physical Review*. DOI: 10.1103/PhysRev.136.B864.
- Hong, M., Zhang, Y. and Hu, F. (2012) 'Membrane protein structure and dynamics from NMR spectroscopy', *Annual Review of Physical Chemistry*. DOI: 10.1146/annurev-physchem-032511-143731.
- Hoops, S. C., Anderson, K. W. and Merz, K. M. (1991) 'Force Field Design for Metalloproteins', *Journal of the American Chemical Society*. DOI: 10.1021/ja00022a010.
- Horta, M. F. *et al.* (2020) 'Cell invasion by intracellular parasites - the many roads to infection', *Journal of cell science*. DOI: 10.1242/jcs.232488.
- Hossain, M. D. *et al.* (2020) 'Reaction mechanism and kinetics for CO₂ reduction on nickel single atom catalysts from quantum mechanics', *Nature Communications*. DOI: 10.1038/s41467-020-16119-6.
- Hsieh, Y. C. *et al.* (2016) 'Comparative normal mode analysis of the dynamics of DENV and ZIKV

- capsids', *Frontiers in Molecular Biosciences*. DOI: 10.3389/fmolb.2016.00085.
- Hu, L. and Ryde, U. (2011) 'Comparison of methods to obtain force-field parameters for metal sites', *Journal of Chemical Theory and Computation*. DOI: 10.1021/ct100725a.
- Humphrey, W., Dalke, A. and Schulten, K. (1996) 'VMD: Visual molecular dynamics', *Journal of Molecular Graphics*. DOI: 10.1016/0263-7855(96)00018-5.
- Hussein, D. *et al.* (2015) 'Validating the GTP-cyclohydrolase 1-feedback regulatory complex as a therapeutic target using biophysical and in vivo approaches', *British Journal of Pharmacology*. DOI: 10.1111/bph.13202.
- Hyde, J. E. (2005) 'Exploring the folate pathway in *Plasmodium falciparum*', *Acta Tropica*, 94(3 SPEC. ISS.). DOI: 10.1016/j.actatropica.2005.04.002.
- Hyde, J. E. (2007) 'Drug-resistant malaria - An insight', *FEBS Journal*. DOI: 10.1111/j.1742-4658.2007.05999.x.
- Ichinose, H. *et al.* (2013) 'GTP cyclohydrolase regulation. Implications for brain development and function.', in *Advances in Pharmacology*. DOI: 10.1016/B978-0-12-411512-5.00003-8.
- Illergård, K., Ardell, D. H. and Elofsson, A. (2009) 'Structure is three to ten times more conserved than sequence - A study of structural response in protein cores', *Proteins: Structure, Function and Bioinformatics*. DOI: 10.1002/prot.22458.
- Isin, B. *et al.* (2012) 'Identification of motions in membrane proteins by elastic network models and their experimental validation', *Methods in Molecular Biology*. DOI: 10.1007/978-1-62703-23-6_17.
- Isin, B., Doruker, P. and Bahar, I. (2002) 'Functional Motions of Influenza Virus Hemagglutinin: A Structure-Based Analytical Approach', *Biophysical Journal*. DOI: 10.1016/S0006-3495(02)75422-2.
- Jaroszewski, L. *et al.* (2011) 'FFAS server: Novel features and applications', *Nucleic Acids Research*. DOI: 10.1093/nar/gkr441.
- Jiang, Z. *et al.* (2019) 'Effects of an electric field on the conformational transition of the protein: A molecular dynamics simulation study', *Polymers*. DOI: 10.3390/polym11020282.
- Jones, D. T., Taylor, W. R. and Thornton, J. M. (1992) 'The rapid generation of mutation data matrices from protein sequences', *Bioinformatics*. DOI: 10.1093/bioinformatics/8.3.275.
- Josling, G. a and Llinás, M. (2015) 'Sexual development in *Plasmodium* parasites: knowing when it's time to commit.', *Nature reviews. Microbiology*. DOI: 10.1038/nrmicro3519.
- Kahsai, A. W. *et al.* (2014) 'Monitoring protein conformational changes and dynamics using stable-isotope labeling and mass spectrometry', *Nature Protocols*. DOI: 10.1038/nprot.2014.075.
- Kanehisa, M. *et al.* (2002) 'The KEGG databases at GenomeNet', *Nucleic Acids Research*. DOI: 10.1093/nar/30.1.42.
- Katoh, K. (2002) 'MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform', *Nucleic Acids Research*. DOI: 10.1093/nar/gkf436.
- Kawasaki, H., Soma, N. and Kretsinger, R. H. (2019) 'Molecular Dynamics Study of the Changes in Conformation of Calmodulin with Calcium Binding and/or Target Recognition', *Scientific Reports*. DOI: 10.1038/s41598-019-47063-1.
- Keskin, O., Jernigan, R. L. and Bahar, I. (2000) 'Proteins with Similar Architecture Exhibit Similar Large-Scale Dynamic Behavior', *Biophysical Journal*. DOI: [https://doi.org/10.1016/S0006-3495\(00\)76756-7](https://doi.org/10.1016/S0006-3495(00)76756-7).
- Khairallah, A., Ross, C. J. and Tastan Bishop, Ö. (2020) 'Probing the Structural Dynamics of the *Plasmodium falciparum* Tunneling-Fold Enzyme 6-Pyruvoyl Tetrahydropterin Synthase to Reveal

Allosteric Drug Targeting Sites', *Frontiers in Molecular Biosciences*. Frontiers. DOI: 10.3389/fmolb.2020.575196.

Khairallah, A., Ross, C. J. and Tastan Bishop, Ö. (2021) 'GTP Cyclohydrolase I as a Potential Drug Target: New Insights into Its Allosteric Modulation via Normal Mode Analysis', *Journal of Chemical Information and Modeling*. American Chemical Society. DOI: 10.1021/acs.jcim.1c00898.

Khairallah, A., Tastan Bishop, Ö. and Moses, V. (2021) 'AMBER force field parameters for the Zn (II) ions of the tunneling-fold enzymes GTP cyclohydrolase I and 6-pyruvoyl tetrahydropterin synthase', *Journal of Biomolecular Structure and Dynamics*. Taylor & Francis. DOI: 10.1080/07391102.2020.1796800.

Kibret, S. *et al.* (2019) 'Environmental and meteorological factors linked to malaria transmission around large dams at three ecological settings in Ethiopia', *Malaria Journal*. DOI: 10.1186/s12936-019-2689-y.

Kim, J. K. *et al.* (2020) 'Elucidating the role of metal ions in carbonic anhydrase catalysis', *Nature Communications*. DOI: 10.1038/s41467-020-18425-5.

Kleckner, I. R. and Foster, M. P. (2011) 'An introduction to NMR-based approaches for measuring protein dynamics', *Biochimica et Biophysica Acta - Proteins and Proteomics*. DOI: 10.1016/j.bbapap.2010.10.012.

Kompis, I. M., Islam, K. and Then, R. L. (2005) 'DNA and RNA synthesis: Antifolates', *Chemical Reviews*. DOI: 10.1021/cr0301144.

Koonin, E. V. (2005) 'Orthologs, Paralogs, and Evolutionary Genomics', *Annual Review of Genetics*. DOI: 10.1146/annurev.genet.39.073003.114725.

Kovacs, J. A., Chacón, P. and Abagyan, R. (2004) 'Predictions of protein flexibility: First-order measures', *Proteins: Structure, Function and Genetics*. DOI: 10.1002/prot.20151.

Krishnakumar, S. *et al.* (2000) 'Functional interactions between GTP cyclohydrolase I and tyrosine hydroxylase in *Drosophila*', *Journal of Neurogenetics*. DOI: 10.3109/01677060009083474.

Krungkrai, J. and Krungrai, S. R. (2016) 'Antimalarial qinghaosu/artemisinin: The therapy worthy of a Nobel Prize', *Asian Pacific Journal of Tropical Biomedicine*. DOI: 10.1016/j.apjtb.2016.03.010.

Krungkrai, J., Yuthavong, Y. and Webster, H. K. (1985) 'Guanosine triphosphate cyclohydrolase in *Plasmodium falciparum* and other *Plasmodium* species', *Molecular and Biochemical Parasitology*. DOI: 10.1016/0166-6851(85)90001-5.

Krylov, A. I. and Gill, P. M. W. (2013) 'Q-Chem: An engine for innovation', *Wiley Interdisciplinary Reviews: Computational Molecular Science*. DOI: 10.1002/wcms.1122.

Kuhn, S., Gill, M. J. and Kain, K. C. (2005) 'Emergence of atovaquone-proguanil resistance during treatment of *Plasmodium falciparum* malaria acquired by a non-immune North American traveller to West Africa', *American Journal of Tropical Medicine and Hygiene*. DOI: 10.4269/ajtmh.2005.72.407.

Kumar, S. *et al.* (2018) 'Drug targets for resistant malaria: Historic to future perspectives', *Biomedicine and Pharmacotherapy*. DOI: 10.1016/j.biopha.2018.05.009.

Kumar, S. C. M. (2017) 'Drug Resistance in Malaria', in *Drug Resistance in Bacteria, Fungi, Malaria, and Cancer*. Cham: Springer International Publishing. DOI: 10.1007/978-3-319-48683-3_19.

Kumar, S., Stecher, G. and Tamura, K. (2016) 'MEGA7: Molecular evolutionary genetics analysis version 7.0. molecular biology and evolution', *Molecular Biology and Evolution*.

Kümpornsinn, K., Kotanan, N., *et al.* (2014a) 'Biochemical and functional characterization of

- Plasmodium falciparum GTP cyclohydrolase I', *Malaria Journal*. DOI: 10.1186/1475-2875-13-150.
- Kümpornsin, K., Kotanan, N., *et al.* (2014b) 'Biochemical and functional characterization of Plasmodium falciparum GTP cyclohydrolase I', *Malaria Journal*. DOI: 10.1186/1475-2875-13-150.
- Kümpornsin, K., Modchang, C., *et al.* (2014) 'Origin of robustness in generating drug-resistant malaria parasites', *Molecular Biology and Evolution*. DOI: 10.1093/molbev/msu140.
- Kyte, J. and Doolittle, R. F. (1982) 'A simple method for displaying the hydropathic character of a protein', *Journal of Molecular Biology*, 157(1). DOI: 10.1016/0022-2836(82)90515-0.
- Laskowski, R. A. *et al.* (1993) 'PROCHECK: a program to check the stereochemical quality of protein structures', *Journal of Applied Crystallography*, 26(2). DOI: 10.1107/S0021889892009944.
- Lassmann, T. and Sonnhammer, E. L. L. (2005) 'Kalign - An accurate and fast multiple sequence alignment algorithm', *BMC Bioinformatics*. DOI: 10.1186/1471-2105-6-298.
- Latremoliere, A. and Costigan, M. (2011) 'GCH1, BH4 and Pain', *Current Pharmaceutical Biotechnology*. DOI: 10.2174/138920111798357393.
- Lee, C., Yang, W. and Parr, R. G. (1988) 'Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density', *Physical Review B*. DOI: 10.1103/PhysRevB.37.785.
- Lee, J. Y. *et al.* (2017) 'Allosteric Modulation of Intact γ -Secretase Structural Dynamics', *Biophysical Journal*. DOI: 10.1016/j.bpj.2017.10.012.
- Leo-Macias, A. *et al.* (2005) 'An analysis of core deformations in protein superfamilies', *Biophysical Journal*. DOI: 10.1529/biophysj.104.052449.
- Levitt, M., Sander, C. and Stern, P. S. (1985) 'Protein normal-mode dynamics: Trypsin inhibitor, crambin, ribonuclease and lysozyme', *Journal of Molecular Biology*. DOI: 10.1016/0022-2836(85)90230-X.
- Levy, R. M. and Karplus, M. (1979) 'Vibrational approach to the dynamics of an α -helix', *Biopolymers*. DOI: 10.1002/bip.1979.360181008.
- Lezon, T. R., Sali, A. and Bahar, I. (2009) 'Global motions of the nuclear pore complex: Insights from elastic network models', *PLoS Computational Biology*. DOI: 10.1371/journal.pcbi.1000496.
- Li, H. *et al.* (2017) 'DynOmics: dynamics of structural proteome and beyond', *Nucleic Acids Research*, 45(W1). DOI: 10.1093/nar/gkx385.
- Li, P. and Merz, K. M. (2014) 'Taking into account the ion-induced dipole interaction in the nonbonded model of ions', *Journal of Chemical Theory and Computation*. DOI: 10.1021/ct400751u.
- Li, P. and Merz, K. M. (2016) 'MCPB.py: A Python Based Metal Center Parameter Builder', *Journal of Chemical Information and Modeling*. DOI: 10.1021/acs.jcim.5b00674.
- Li, P. and Merz, K. M. (2017) 'Metal Ion Modeling Using Classical Mechanics', *Chemical Reviews*. DOI: 10.1021/acs.chemrev.6b00440.
- Li, P., Song, L. F. and Merz, K. M. (2015) 'Systematic parameterization of monovalent ions employing the nonbonded model', *Journal of Chemical Theory and Computation*. DOI: 10.1021/ct500918t.
- Liao, Q. *et al.* (2017) 'Extending the Nonbonded Cationic Dummy Model to Account for Ion-Induced Dipole Interactions', *Journal of Physical Chemistry Letters*. DOI: 10.1021/acs.jpcclett.7b02358.
- Lindorff-Larsen, K. *et al.* (2005) 'Simultaneous determination of protein structure and dynamics', *Nature*. DOI: 10.1038/nature03199.

- Liu, W. *et al.* (2017) ‘Molecular dynamics simulations elucidate conformational selection and induced fit mechanisms in the binding of PD-1 and PD-L1’, *Molecular BioSystems*. DOI: 10.1039/c7mb00036g.
- Liu, X. *et al.* (2008) ‘Mechanics of Channel Gating of the Nicotinic Acetylcholine Receptor’, *PLoS Computational Biology*. DOI: 10.1371/journal.pcbi.0040019.
- Liu, Y. and Bahar, I. (2010) ‘Toward understanding allosteric signaling mechanisms in the atpase domain of molecular chaperones’, in *Pacific Symposium on Biocomputing 2010, PSB 2010*. DOI: 10.1142/9789814295291_0029.
- Lopes, P. E. M., Guvench, O. and Mackerell, A. D. (2015) ‘Current status of protein force fields for molecular dynamics simulations’, *Methods in Molecular Biology*. DOI: 10.1007/978-1-4939-1465-4_3.
- Lötsch, J. *et al.* (2010) ‘A GTP cyclohydrolase 1 genetic variant delays cancer pain’, *Pain*. DOI: 10.1016/j.pain.2009.10.021.
- Loutchko, D. and Flechsig, H. (2020) ‘Allosteric communication in molecular machines via information exchange: what can be learned from dynamical modeling’, *Biophysical Reviews*. DOI: 10.1007/s12551-020-00667-8.
- Lu, C. H. *et al.* (2012) ‘Prediction of metal ion-binding sites in proteins using the fragment transformation method’, *PLoS ONE*. DOI: 10.1371/journal.pone.0039252.
- Lu, S. I. and Gao, L. T. (2018) ‘Calculations of Electronic Excitation Energies and Excess Electric Dipole Moments of Solvated p-Nitroaniline with the EOM-CCSD-PCM Method’, *Journal of Physical Chemistry A*. DOI: 10.1021/acs.jpca.8b02321.
- Lu, S., Shen, Q. and Zhang, J. (2019) ‘Allosteric Methods and Their Applications: Facilitating the Discovery of Allosteric Drugs and the Investigation of Allosteric Mechanisms’, *Accounts of Chemical Research*. DOI: 10.1021/acs.accounts.8b00570.
- Ma, B. and Nussinov, R. (2010) ‘Enzyme dynamics point to stepwise conformational selection in catalysis’, *Current Opinion in Chemical Biology*. DOI: 10.1016/j.cbpa.2010.08.012.
- Mackenzie, C. O. and Grigoryan, G. (2017) ‘Protein structural motifs in prediction and design’, *Current Opinion in Structural Biology*. DOI: 10.1016/j.sbi.2017.03.012.
- Mahajan, S. and Sanejouand, Y. H. (2015) ‘On the relationship between low-frequency normal modes and the large-scale conformational changes of proteins’, *Archives of Biochemistry and Biophysics*. DOI: 10.1016/j.abb.2014.12.020.
- Maita, N. *et al.* (2002) ‘Crystal structure of the stimulatory complex of GTP cyclohydrolase I and its feedback regulatory protein GFRP’, *Proceedings of the National Academy of Sciences of the United States of America*. DOI: 10.1073/pnas.022646999.
- Maita, N. *et al.* (2004) ‘Structural Basis of Biopterin-induced Inhibition of GTP Cyclohydrolase I by GFRP, Its Feedback Regulatory Protein’, *Journal of Biological Chemistry*. DOI: 10.1074/jbc.M409440200.
- Manathunga, M. *et al.* (2020) ‘Parallel Implementation of Density Functional Theory Methods in the Quantum Interaction Computational Kernel Program’, *ChemRxiv*. DOI: 10.26434/chemrxiv.12018963.v1.
- Mäntsälä, P. and Niemi, J. (2009) ‘Enzymes: The Biological Catalysts of Life’, *Physiology and Maintenance*.
- Mardirossian, N. and Head-Gordon, M. (2017) ‘Thirty years of density functional theory in computational chemistry: An overview and extensive assessment of 200 density functionals’, *Molecular Physics*. DOI: 10.1080/00268976.2017.1333644.

- Mather, M. W. and Ke, H. (2019) 'para-Aminobenzoate Synthesis versus Salvage in Malaria Parasites', *Trends in Parasitology*. DOI: 10.1016/j.pt.2019.01.010.
- Maximova, T. *et al.* (2016) 'Principles and Overview of Sampling Methods for Modeling Macromolecular Structure and Dynamics', *PLoS Computational Biology*. DOI: 10.1371/journal.pcbi.1004619.
- Mazal, H. and Haran, G. (2019) 'Single-molecule FRET methods to study the dynamics of proteins at work', *Current Opinion in Biomedical Engineering*. DOI: 10.1016/j.cobme.2019.08.007.
- McNeill, E. and Channon, K. M. (2012) 'The role of tetrahydrobiopterin in inflammation and cardiovascular disease', *Thrombosis and Haemostasis*. DOI: 10.1160/TH12-06-0424.
- Menard, D. *et al.* (2005) 'Association of failures of seven-day courses of artesunate in a non-immune population in Bangui, Central African Republic with decreased sensitivity of *Plasmodium falciparum*', *American Journal of Tropical Medicine and Hygiene*. DOI: 10.4269/ajtmh.2005.73.616.
- Mencacci, N. E. *et al.* (2014) 'Parkinson's disease in GTP cyclohydrolase 1 mutation carriers', *Brain*. DOI: 10.1093/brain/awu179.
- Meshnick, S. R. (2002) 'Artemisinin: Mechanisms of action, resistance and toxicity', in *International Journal for Parasitology*. DOI: 10.1016/S0020-7519(02)00194-7.
- Mikulska-Ruminska, K. *et al.* (2017) 'Nanomechanics of multidomain neuronal cell adhesion protein contactin revealed by single molecule AFM and SMD', *Scientific Reports*. DOI: 10.1038/s41598-017-09482-w.
- Mitchell, M. R., Tlustý, T. and Leibler, S. (2016) 'Strain analysis of protein structures and low dimensionality of mechanical allosteric couplings', *Proceedings of the National Academy of Sciences of the United States of America*. DOI: 10.1073/pnas.1609462113.
- Mlugu, E. M. *et al.* (2020) 'Effectiveness of sulfadoxine–pyrimethamine for intermittent preventive treatment of malaria and adverse birth outcomes in pregnant women', *Pathogens*. DOI: 10.3390/pathogens9030207.
- Mohapatra, P. K. *et al.* (2014) 'Molecular evidence of increased resistance to anti-folate drugs in *Plasmodium falciparum* in North-east India: A signal for potential failure of artemisinin plus sulphadoxine-pyrimethamine combination therapy', *PLoS ONE*. DOI: 10.1371/journal.pone.0105562.
- Molina-Franky, J. *et al.* (2020) '*Plasmodium falciparum* pre-erythrocytic stage vaccine development', *Malaria Journal*. DOI: 10.1186/s12936-020-3141-z.
- Morgenstern, B. (1999) 'DIALIGN 2: Improvement of the segment-to-segment approach to multiple sequence alignment', in *Bioinformatics*. DOI: 10.1093/bioinformatics/15.3.211.
- Mount, D. W. (2008) 'Comparison of the PAM and BLOSUM amino acid substitution matrices', *Cold Spring Harbor Protocols*. DOI: 10.1101/pdb.ip59.
- Müller-Sienerth, N. *et al.* (2020) 'A panel of recombinant proteins from human-infective *Plasmodium* species for serological surveillance', *Malaria Journal*. DOI: 10.1186/s12936-020-3111-5.
- Müller, I. B. and Hyde, J. E. (2010) 'Antimalarial drugs: Modes of action and mechanisms of parasite resistance', *Future Microbiology*. DOI: 10.2217/fmb.10.136.
- Müller, I. B. and Hyde, J. E. (2013) 'Folate metabolism in human malaria parasites - 75 years on', *Molecular and Biochemical Parasitology*. DOI: 10.1016/j.molbiopara.2013.02.008.
- Na, H. and Song, G. (2016) 'The effective degeneracy of protein normal modes', *Physical Biology*. DOI: 10.1088/1478-3975/13/3/036002.
- Nair, S. *et al.* (2008) 'Adaptive copy number evolution in malaria parasites', *PLoS Genetics*. DOI:

10.1371/journal.pgen.1000243.

Nar, H. *et al.* (1994) 'Three-dimensional structure of 6-pyruvoyl tetrahydropterin synthase, an enzyme involved in tetrahydrobiopterin biosynthesis.', *The EMBO Journal*. DOI: 10.1002/j.1460-2075.1994.tb06377.x.

Nar, H., Huber, R., Auerbach, G., *et al.* (1995) 'Active site topology and reaction mechanism of GTP cyclohydrolase I', *Proceedings of the National Academy of Sciences of the United States of America*. DOI: 10.1073/pnas.92.26.12120.

Nar, H., Huber, R., Meining, W., *et al.* (1995) 'Atomic structure of GTP cyclohydrolase I', *Structure*. DOI: 10.1016/S0969-2126(01)00179-4.

Nar, H. (2011) '6-Pyruvoyl-Tetrahydropterin Synthase', in *Encyclopedia of Inorganic and Bioinorganic Chemistry*. Chichester, UK: John Wiley & Sons, Ltd. DOI: 10.1002/9781119951438.eibc0475.

Nduati, E. *et al.* (2008) 'Effect of folate derivatives on the activity of antifolate drugs used against malaria and cancer', *Parasitology Research*. DOI: 10.1007/s00436-008-0897-4.

Needleman, S. B. and Wunsch, C. D. (1970) 'A general method applicable to the search for similarities in the amino acid sequence of two proteins', *Journal of Molecular Biology*. DOI: 10.1016/0022-2836(70)90057-4.

Nghochuzie, N. N. *et al.* (2020) 'Pausing the Fight Against Malaria to Combat the COVID-19 Pandemic in Africa: Is the Future of Malaria Bleak?', *Frontiers in Microbiology*. DOI: 10.3389/fmicb.2020.01476.

Nilsson, K. *et al.* (2003) 'An automatic method to generate force-field parameters for hetero-compounds', *Acta Crystallographica - Section D Biological Crystallography*. DOI: 10.1107/S0907444902021431.

Noguti, T. and Go, N. (1982) 'Collective variable description of small-amplitude conformational fluctuations in a globular protein', *Nature*. DOI: 10.1038/296776a0.

Notredame, C., Higgins, D. G. and Heringa, J. (2000) 'T-coffee: A novel method for fast and accurate multiple sequence alignment', *Journal of Molecular Biology*. DOI: 10.1006/jmbi.2000.4042.

Nussinov, R. and Ma, B. (2012) 'Protein dynamics and conformational selection in bidirectional signal transduction', *BMC Biology*. DOI: 10.1186/1741-7007-10-2.

Nussinov, R. and Tsai, C. J. (2013) 'Allostery in disease and in drug discovery', *Cell*. DOI: 10.1016/j.cell.2013.03.034.

Nyunt, M. M. and Plowe, C. V. (2007) 'Pharmacologic advances in the global control and treatment of malaria: Combination therapy and resistance', *Clinical Pharmacology and Therapeutics*. DOI: 10.1038/sj.clpt.6100361.

Nzila, A., Ward, Steve A., *et al.* (2005a) 'Comparative folate metabolism in humans and malaria parasites (part I): pointers for malaria treatment from cancer chemotherapy', *Trends in Parasitology*. DOI: 10.1016/j.pt.2005.04.002.

Nzila, A., Ward, Steve A., *et al.* (2005b) 'Comparative folate metabolism in humans and malaria parasites (part II): Activities as yet untargeted or specific to Plasmodium', *Trends in Parasitology*. DOI: 10.1016/j.pt.2005.05.008.

Nzila, A., Ward, Steve A., *et al.* (2005) 'Comparative folate metabolism in humans and malaria parasites (part II): Activities as yet untargeted or specific to Plasmodium', *Trends in Parasitology*. DOI: 10.1016/j.pt.2005.05.008.

Nzila, A. (2006) 'The past, present and future of antifolates in the treatment of Plasmodium

- falciparum infection', *Journal of Antimicrobial Chemotherapy*. DOI: 10.1093/jac/dk1104.
- O'Sullivan, O. *et al.* (2004) '3DCoffee: Combining protein sequences and structures within multiple sequence alignments', *Journal of Molecular Biology*. DOI: 10.1016/j.jmb.2004.04.058.
- Okrasa, K. and Kazlauskas, R. J. (2006) 'Manganese-substituted carbonic anhydrase as a new peroxidase', *Chemistry - A European Journal*. DOI: 10.1002/chem.200501413.
- Onwujekwe, O. *et al.* (2013) 'The economic burden of malaria on households and the health system in Enugu state southeast Nigeria', *PLoS ONE*. DOI: 10.1371/journal.pone.0078362.
- Opperdoes, F. R. and Lemey, P. (2012) 'Phylogenetic analysis using protein sequences', in *The Phylogenetic Handbook*. DOI: 10.1017/cbo9780511819049.011.
- Orellana, L. (2019) 'Large-Scale Conformational Changes and Protein Function: Breaking the in silico Barrier', *Frontiers in Molecular Biosciences*. DOI: 10.3389/fmolb.2019.00117.
- Orozco, M. (2014) 'A theoretical view of protein dynamics', *Chemical Society Reviews*. DOI: 10.1039/c3cs60474h.
- Tastan Bishop, Ö, de Beer, T. A. P. and Joubert, F. (2008) 'Protein homology modelling and its use in South Africa', *South African Journal of Science*, 104 (February), pp. 2–6.
- Palamini, M., Canciani, A. and Forneris, F. (2016) 'Identifying and visualizing macromolecular flexibility in structural biology', *Frontiers in Molecular Biosciences*. DOI: 10.3389/fmolb.2016.00047.
- Páll, S. *et al.* (2020) 'Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS', *Journal of Chemical Physics*. DOI: 10.1063/5.0018516.
- Pan, L. *et al.* (2011) 'GTP-cyclohydrolase deficiency responsive to sapropterin and 5-HTP supplementation: Relief of treatment-refractory depression and suicidal behaviour', *BMJ Case Reports*. DOI: 10.1136/bcr.03.2011.3927.
- Pang, Y. P. *et al.* (2000) 'Successful molecular dynamics simulation of the zinc-bound farnesyltransferase using the cationic dummy atom approach.', *Protein science: a publication of the Protein Society*.
- Pearson, W. R. and Lipman, D. J. (1988) 'Improved tools for biological sequence comparison.', *Proceedings of the National Academy of Sciences of the United States of America*. DOI: 10.1073/pnas.85.8.2444.
- Pei, J. and Grishin, N. V. (2007) 'PROMALS: Towards accurate multiple sequence alignments of distantly related proteins', *Bioinformatics*. DOI: 10.1093/bioinformatics/btm017.
- Penkler, D. *et al.* (2017) 'Perturbation-Response Scanning Reveals Key Residues for Allosteric Control in Hsp70', *Journal of Chemical Information and Modeling*. DOI: 10.1021/acs.jcim.6b00775.
- Penkler, D. L., Atilgan, C. and Tastan Bishop, Ö. (2018) 'Allosteric Modulation of Human Hsp90 α Conformational Dynamics', *Journal of Chemical Information and Modeling*. DOI: 10.1021/acs.jcim.7b00630.
- Peters, M. B. *et al.* (2010) 'Structural survey of zinc-containing proteins and development of the zinc AMBER force field (ZAFF)', *Journal of Chemical Theory and Computation*. DOI: 10.1021/ct1002626.
- Petrovic, D. *et al.* (2018) 'Conformational dynamics and enzyme evolution', *Journal of the Royal Society Interface*. DOI: 10.1098/rsif.2018.0330.
- Ploom, T. *et al.* (1999) 'Crystallographic and kinetic investigations on the mechanism of 6-pyruvoyl tetrahydropterin synthase', *Journal of Molecular Biology*. DOI: 10.1006/jmbi.1998.2511.

- Plowe, C. V. *et al.* (1995) 'Pyrimethamine and proguanil resistance-conferring mutations in *Plasmodium falciparum* dihydrofolate reductase: Polymerase chain reaction methods for surveillance in Africa', *American Journal of Tropical Medicine and Hygiene*. DOI: 10.4269/ajtmh.1995.52.565.
- Poltev, V. (2017) 'Molecular mechanics: Principles, history, and current status', in *Handbook of Computational Chemistry*. DOI: 10.1007/978-3-319-27282-5_9.
- Postils, V. *et al.* (2018) 'Quantum mechanics/molecular mechanics studies on the relative reactivities of compound I and II in cytochrome p450 enzymes', *International Journal of Molecular Sciences*. DOI: 10.3390/ijms19071974.
- Prejanò, M. *et al.* (2020) 'The effects of the metal ion substitution into the active site of metalloenzymes: A theoretical insight on some selected cases', *Catalysts*. DOI: 10.3390/catal10091038.
- R Core Team (2019) 'R: A language and environment for statistical computing.', *R Foundation for Statistical Computing*.
- Reader, J. *et al.* (2021) 'Multistage and transmission-blocking targeted antimalarials discovered from the open-source MMV Pandemic Response Box', *Nature Communications*. DOI: 10.1038/s41467-020-20629-8.
- Rebelo, J. *et al.* (2003) 'Biosynthesis of pteridines. Reaction mechanism of GTP cyclohydrolase I', *Journal of Molecular Biology*. DOI: 10.1016/S0022-2836(02)01303-7.
- Rédei, G. P. (2008) 'CLUSTAL W (improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice)', in *Encyclopedia of Genetics, Genomics, Proteomics and Informatics*. DOI: 10.1007/978-1-4020-6754-9_3188.
- Remmert, M. *et al.* (2012) 'HHblits: Lightning-fast iterative protein sequence searching by HMM-HMM alignment', *Nature Methods*. DOI: 10.1038/nmeth.1818.
- Riordan, J. F. (1977) 'The role of metals in enzyme activity', *Annals of Clinical and Laboratory Science*.
- Rivera, J. C. *et al.* (2017) 'Tetrahydrobiopterin (BH4) deficiency is associated with augmented inflammation and microvascular degeneration in the retina', *Journal of Neuroinflammation*. DOI: 10.1186/s12974-017-0955-x.
- Robert, X. and Gouet, P. (2014) 'Deciphering key features in protein structures with the new ENDscript server', *Nucleic Acids Research*. DOI: 10.1093/nar/gku316.
- Rosca, F. *et al.* (2002) 'Investigations of anharmonic low-frequency oscillations in heme proteins', *Journal of Physical Chemistry A*. DOI: 10.1021/jp0129277.
- Rosenberg, I. H. (2012) 'A history of the isolation and identification of folic acid (folate)', *Annals of Nutrition and Metabolism*. DOI: 10.1159/000343112.
- Rosenhead, L. and Goldstein, H. (1951) 'Classical Mechanics', *The Mathematical Gazette*. DOI: 10.2307/3610571.
- Ross, C. *et al.* (2018) 'MODE-TASK: large-scale protein motion tools', *Bioinformatics*. Edited by A. Valencia. Narnia. DOI: 10.1093/bioinformatics/bty427.
- Ross, C. J. *et al.* (2018) 'Unraveling the Motions behind Enterovirus 71 Uncoating', *Biophysical Journal*, 114(4). DOI: 10.1016/j.bpj.2017.12.021.s
- Ross, C., Knox, C. and Tastan Bishop, Ö. (2017) 'Interacting motif networks located in hotspots associated with RNA release are conserved in Enterovirus capsids', *FEBS Letters*. DOI: 10.1002/1873-3468.12663.

- Sacquin-Mora, S. (2018) ‘Mechanical variations in proteins with large-scale motions highlight the formation of structural locks’, *Journal of Structural Biology*. DOI: 10.1016/j.jsb.2018.05.006.
- Sander, C. and Schneider, R. (1991) ‘Database of homology-derived protein structures and the structural meaning of sequence alignment’, *Proteins: Structure, Function, and Bioinformatics*. DOI: 10.1002/prot.340090107.
- Sarma, H. *et al.* (2018) ‘Phylogenetic analysis: Early evolution of life’, in *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*. DOI: 10.1016/B978-0-12-809633-8.20171-4.
- Sarma, N. *et al.* (2019) ‘The economic burden of Malaria: Revisiting the evidence’, *American Journal of Tropical Medicine and Hygiene*. DOI: 10.4269/ajtmh.19-0386.
- Sato, S. (2021) ‘Plasmodium—a brief introduction to the parasites causing human malaria and their basic biology’, *Journal of Physiological Anthropology*. DOI: 10.1186/s40101-020-00251-9.
- Scherrer, A. *et al.* (2017) ‘On the mass of atoms in molecules: Beyond the born-oppenheimer approximation’, *Physical Review X*. DOI: 10.1103/PhysRevX.7.031035.
- Schlagenhauf, P. *et al.* (2018) ‘Malaria chemoprophylaxis’, in *Travel Medicine*. DOI: 10.1016/B978-0-323-54696-6.00015-X.
- Schmidt, M. W. *et al.* (2016) ‘GAMESS’, *Journal of Computational Chemistry*.
- Schramek, N. *et al.* (2002) ‘Reaction mechanism of GTP cyclohydrolase I: Single turnover experiments using a kinetically competent reaction intermediate’, *Journal of Molecular Biology*. DOI: 10.1006/jmbi.2001.5339.
- Schrödinger, E. (1926) ‘An undulatory theory of the mechanics of atoms and molecules’, *Physical Review*. DOI: 10.1103/PhysRev.28.1049.
- Schweitzer, B. I., Dicker, A. P. and Bertino, J. R. (1990) ‘Dihydrofolate reductase as a therapeutic target.’, *The FASEB Journal*. DOI: 10.1096/fasebj.4.8.2185970.
- Seminario, J. M. (1996) ‘Calculation of intramolecular force fields from second-derivative tensors’, *International Journal of Quantum Chemistry*. DOI: 10.1002/(SICI)1097-461X(1996)60:7<1271::AID-QUA8>3.0.CO;2-W.
- Senn, H. M. and Thiel, W. (2009) ‘QM/MM methods for biomolecular systems’, *Angewandte Chemie - International Edition*. DOI: 10.1002/anie.200802019.
- Seno, Y. and Go, N. (1990) ‘Deoxymyoglobin studied by the conformational normal mode analysis. I. Dynamics of globin and the heme-globin interaction’, *Journal of Molecular Biology*. DOI: 10.1016/S0022-2836(05)80063-4.
- Seritan, S., Bannwarth, C., Fales, Bryan S., *et al.* (2020) ‘TeraChem: A graphical processing unit-accelerated electronic structure package for large-scale ab initio molecular dynamics’, *Wiley Interdisciplinary Reviews: Computational Molecular Science*. DOI: 10.1002/wcms.1494.
- Seritan, S., Bannwarth, C., Fales, B. Scott, *et al.* (2020) ‘TeraChem: Accelerating electronic structure and ab initio molecular dynamics with graphical processing units’, *The Journal of chemical physics*. DOI: 10.1063/5.0007615.
- Sevене, E., González, R. and Menéndez, C. (2010) ‘Current knowledge and challenges of antimalarial drugs for treatment and prevention in pregnancy’, *Expert Opinion on Pharmacotherapy*. DOI: 10.1517/14656561003733599.
- Shaw, D. E. *et al.* (2009) ‘Millisecond-scale molecular dynamics simulations on Anton’, in. DOI: 10.1145/1654059.1654099.

- Shen, M. and Sali, A. (2006) 'Statistical potential for assessment and prediction of protein structures', *Protein Science*, 15(11). DOI: 10.1110/ps.062416606.
- Shi, Y. and Wu, J. (2007) 'Structural basis of protein-protein interaction studied by NMR', *Journal of Structural and Functional Genomics*. DOI: 10.1007/s10969-007-9021-8.
- Shibeshi, M. A., Kifle, Z. D. and Atnafie, S. A. (2020) 'Antimalarial drug resistance and novel targets for antimalarial drug discovery', *Infection and Drug Resistance*. DOI: 10.2147/IDR.S279433.
- Shoemaker, S. C. and Ando, N. (2018) 'X-rays in the Cryo-Electron Microscopy Era: Structural Biology's Dynamic Future', *Biochemistry*. DOI: 10.1021/acs.biochem.7b01031.
- Shrivastava, I. H. *et al.* (2020) 'Allostery as Structure-Encoded Collective Dynamics', in *Structural Biology in Drug Discovery*. DOI: 10.1002/9781118681121.ch6.
- Shrivastava, I. H. and Bahar, I. (2006) 'Common Mechanism of Pore Opening Shared by Five Different Potassium Channels', *Biophysical Journal*. DOI: 10.1529/biophysj.105.080093.
- Siciliano, G. and Alano, P. (2015) 'Enlightening the malaria parasite life cycle: Bioluminescent Plasmodium in fundamental and applied research', *Frontiers in Microbiology*. DOI: 10.3389/fmicb.2015.00391.
- Sievers, F. *et al.* (2011) 'Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega', *Molecular Systems Biology*. DOI: 10.1038/msb.2011.75.
- Simonet, S. *et al.* (2021) 'GTP-cyclohydrolase deficiency induced peripheral and deep microcirculation dysfunction with age', *Microvascular Research*. DOI: 10.1016/j.mvr.2020.104078.
- Singh, A. *et al.* (2018) 'Application of molecular dynamic simulation to study food proteins: A review', *Critical Reviews in Food Science and Nutrition*. DOI: 10.1080/10408398.2017.1341864.
- Sköld, O. (2000) 'Sulfonamide resistance: Mechanisms and trends', *Drug Resistance Updates*. DOI: 10.1054/drup.2000.0146.
- Smith, T. F. and Waterman, M. S. (1981) 'Identification of common molecular subsequences', *Journal of Molecular Biology*. DOI: 10.1016/0022-2836(81)90087-5.
- Sparta, M. and Alexandrova, A. N. (2012) 'How Metal Substitution Affects the Enzymatic Activity of Catechol-O-Methyltransferase', *PLoS ONE*. DOI: 10.1371/journal.pone.0047172.
- Srivastava, Ashutosh *et al.* (2018) 'Role of computational methods in going beyond x-ray crystallography to explore protein structure and dynamics', *International Journal of Molecular Sciences*. DOI: 10.3390/ijms19113401.
- Staines, H. M. *et al.* (2018) 'Clinical implications of Plasmodium resistance to atovaquone/proguanil: A systematic review and meta-analysis', *Journal of Antimicrobial Chemotherapy*. DOI: 10.1093/jac/dkx431.
- Stewart, E. T. (1970) 'Quantum Chemistry', *Nature*. DOI: 10.1038/226383b0.
- Stote, R. H. and Karplus, M. (1995) 'Zinc binding in proteins and solution: A simple but accurate nonbonded representation', *Proteins: Structure, Function, and Genetics*. DOI: 10.1002/prot.340230104.
- Su, J. G. *et al.* (2011) 'Identification of key residues for protein conformational transition using elastic network model', *Journal of Chemical Physics*. DOI: 10.1063/1.3651480.
- Sugawara, E. and Nikaido, H. (2014) 'Properties of AdeABC and AdeIJK efflux systems of *Acinetobacter baumannii* compared with those of the AcrAB-TolC system of *Escherichia coli*.', *Antimicrobial agents and chemotherapy*. DOI: 10.1128/AAC.03728-14.
- Suplatov, D. and Švedas, V. (2015) 'Study of functional and allosteric sites in protein superfamilies',

- Swick, L. and Kapatos, G. (2006) 'A yeast 2-hybrid analysis of human GTP cyclohydrolase I protein interactions', *Journal of Neurochemistry*. DOI: 10.1111/j.1471-4159.2006.03836.x.
- Tama, F. *et al.* (2003) 'Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy', *Proceedings of the National Academy of Sciences of the United States of America*. DOI: 10.1073/pnas.1632476100.
- Tama, F. and Sanejouand, Y.-H. (2001) 'Conformational change of proteins arising from normal mode calculations', *Protein Engineering, Design and Selection*. DOI: 10.1093/protein/14.1.1.
- Tamura, K. *et al.* (2011) 'MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods', *Molecular Biology and Evolution*. DOI: 10.1093/molbev/msr121.
- Tanaka, Y. *et al.* (2005) 'Novel reaction mechanism of GTP cyclohydrolase I. High-resolution X-ray crystallography of *Thermus thermophilus* HB8 enzyme complexed with a transition state analogue, the 8-oxoguanine derivative', *Journal of Biochemistry*. DOI: 10.1093/jb/mvi120.
- Tardieux, I. and Baum, J. (2016) 'Reassessing the mechanics of parasite motility and host-cell invasion', *Journal of Cell Biology*. DOI: 10.1083/JCB.201605100.
- Tazawa, M. *et al.* (2000) 'GTP cyclohydrolase I from *Tetrahymena pyriformis*: Cloning of cDNA and expression', *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology*. DOI: 10.1016/S0305-0491(00)00239-X.
- Teilum, K., Olsen, J. G. and Kragelund, B. B. (2009) 'Functional aspects of protein flexibility', *Cellular and Molecular Life Sciences*. DOI: 10.1007/s00018-009-0014-6.
- Temiz, N. A. and Bahar, I. (2002) 'Inhibitor binding alters the directions of domain motions in HIV-1 reverse transcriptase', *Proteins: Structure, Function and Genetics*. DOI: 10.1002/prot.10183.
- Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) 'CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice', *Nucleic Acids Research*. DOI: 10.1093/nar/22.22.4673.
- THÖNY, B., AUERBACH, G. and BLAU, N. (2000) 'Tetrahydrobiopterin biosynthesis, regeneration and functions', *Biochemical Journal*. DOI: 10.1042/0264-6021:3470001.
- Tirion, M. M. (1996) 'Large amplitude elastic motions in proteins from a single-parameter, atomic analysis', *Physical Review Letters*, 77(9). DOI: 10.1007/s11199-005-4202-z.
- Tjong, E. and Mohiuddin, S. S. (2019) *Biochemistry, Tetrahydrofolate, StatPearls*.
- Togashi, Y. and Flechsig, H. (2018) 'Coarse-grained protein dynamics studies using elastic network models', *International Journal of Molecular Sciences*. DOI: 10.3390/ijms19123899.
- Di Tommaso, P. *et al.* (2011) 'T-Coffee: A web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension', *Nucleic Acids Research*. DOI: 10.1093/nar/gkr245.
- Towler, P. *et al.* (2004) 'ACE2 X-Ray Structures Reveal a Large Hinge-bending Motion Important for Inhibitor Binding and Catalysis', *Journal of Biological Chemistry*. DOI: 10.1074/jbc.M311191200.
- Tse, E. G., Korsik, M. and Todd, M. H. (2019) 'The past, present and future of anti-malarial medicines', *Malaria Journal*. DOI: 10.1186/s12936-019-2724-z.
- Turner, P. (2005) 'XMGRACE, Version 5.1. 19.', *Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology, Beaverton*.
- Ung, P. M. U. *et al.* (2013) 'Identification of Key Hinge Residues Important for Nucleotide-

- Dependent Allosterity in *E. coli* Hsp70/DnaK', *PLoS Computational Biology*. DOI: 10.1371/journal.pcbi.1003279.
- Uwimana, A. *et al.* (2020) 'Emergence and clonal expansion of in vitro artemisinin-resistant *Plasmodium falciparum* kelch13 R561H mutant parasites in Rwanda', *Nature Medicine*. DOI: 10.1038/s41591-020-1005-2.
- Vaughan, A. M., Aly, A. S. I. and Kappe, S. H. I. (2008) 'Malaria Parasite Pre-Erythrocytic Stage Infection: Gliding and Hiding', *Cell Host and Microbe*. DOI: 10.1016/j.chom.2008.08.010.
- Verzier, L. H. *et al.* (2019) 'Plasmodium knowlesi as a model system for characterising Plasmodium vivax drug resistance candidate genes', *PLoS Neglected Tropical Diseases*. DOI: 10.1371/journal.pntd.0007470.
- Visser, M. J. E. and Pretorius, E. (2019) 'Atomic Force Microscopy: The Characterisation of Amyloid Protein Structure in Pathology', *Current Topics in Medicinal Chemistry*. DOI: 10.2174/1568026619666191121143240.
- Vosko, S. H., Wilk, L. and Nusair, M. (1980) 'Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis', *Canadian Journal of Physics*, 58(8). DOI: 10.1139/p80-159.
- Wako, H. and Endo, S. (2011) 'Ligand-induced conformational change of a protein reproduced by a linear combination of displacement vectors obtained from normal mode analysis', *Biophysical Chemistry*. DOI: 10.1016/j.bpc.2011.07.004.
- Wako, H. and Endo, S. (2017) 'Normal mode analysis as a method to derive protein dynamics information from the Protein Data Bank', *Biophysical Reviews*. DOI: 10.1007/s12551-017-0330-2.
- Walker, R. C. and Götz, A. W. (2016) 'Electronic Structure Calculations on Graphics Processing Units: From Quantum Chemistry to Condensed Matter Physics, Electronic Structure Calculations on Graphics Processing Units: From Quantum Chemistry to Condensed Matter Physics'. DOI: 10.1002/9781118670712.
- Wang, D. and Geva, E. (2005) 'Protein structure and dynamics from single-molecule fluorescence resonance energy transfer', *Journal of Physical Chemistry B*. DOI: 10.1021/jp0478864.
- Wartak, M. S. *et al.* (2019) 'Born–Oppenheimer Approximation', in *Field Guide to Solid State Physics*. DOI: 10.1117/3.2510243.ch29.
- Webb, B. and Sali, A. (2017) 'Protein structure modeling with MODELLER', in *Methods in Molecular Biology*. DOI: 10.1007/978-1-4939-7231-9_4.
- Whelan, S. and Goldman, N. (2001) 'A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach', *Molecular Biology and Evolution*. DOI: 10.1093/oxfordjournals.molbev.a003851.
- White, N. (1999) 'Antimalarial drug resistance and combination chemotherapy', *Philosophical Transactions of the Royal Society B: Biological Sciences*. DOI: 10.1098/rstb.1999.0426.
- White, N. J. (2004) 'Antimalarial drug resistance', *Journal of Clinical Investigation*. DOI: 10.1172/JCI21682.
- Wiederstein, M. and Sippl, M. J. (2007) 'ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins', *Nucleic Acids Research*, 35(Web Server). DOI: 10.1093/nar/gkm290.
- Woodrow, C. J. and White, N. J. (2017) 'The clinical impact of artemisinin resistance in Southeast Asia and the potential for future spread', *FEMS Microbiology Reviews*. DOI: 10.1093/femsre/fuw037.
- World Health Organization (2020) *World Malaria Report: 20 years of global progress and*

challenges, *World Health*.

Wuthrich, K. (1990) 'Protein structure determination in solution by NMR spectroscopy', *Journal of Biological Chemistry*. DOI: 10.1142/9789812795830_0001.

Xia, X. *et al.* (2009) 'MICAlign: A sequence-to-structure alignment tool integrating multiple sources of information in conditional random fields', *Bioinformatics*. DOI: 10.1093/bioinformatics/btp251.

Xiong, J. (2006) *Essential Bioinformatics, Essential Bioinformatics*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511806087.

Yang, L. W. and Bahar, I. (2005) 'Coupling between catalytic site and collective dynamics: A requirement for mechanochemical activity of enzymes', *Structure*. DOI: 10.1016/j.str.2005.03.015.

Yazhini, A. and Srinivasan, N. (2020) 'How good are comparative models in the understanding of protein dynamics?', *Proteins: Structure, Function and Bioinformatics*. DOI: 10.1002/prot.25879.

Yu, Z., Li, P. and Merz, K. M. (2018) 'Extended Zinc AMBER Force Field (EZAFF)', *Journal of Chemical Theory and Computation*. DOI: 10.1021/acs.jctc.7b00773.

Yuthavong, Y. *et al.* (2005) 'Malarial (*Plasmodium falciparum*) dihydrofolate reductase-thymidylate synthase: Structural basis for antifolate resistance and development of effective inhibitors', *Parasitology*. DOI: 10.1017/S003118200400664X.

Zhang, P., Shen, L. and Yang, W. (2019) 'Solvation Free Energy Calculations with Quantum Mechanics/Molecular Mechanics and Machine Learning Models', *Journal of Physical Chemistry B*. DOI: 10.1021/acs.jpcc.8b11905.

Zhang, Y. *et al.* (2020) 'Intrinsic dynamics is evolutionarily optimized to enable allosteric behavior', *Current Opinion in Structural Biology*. DOI: 10.1016/j.sbi.2019.11.002.

Zhang, Y. and Zheng, J. (2020) 'Bioinformatics of metalloproteins and metalloproteomes', *Molecules*. DOI: 10.3390/molecules25153366.

Zheng, F. and Grigoryan, G. (2017) 'Sequence statistics of tertiary structural motifs reflect protein stability', *PLoS ONE*. DOI: 10.1371/journal.pone.0178272.

Zheng, S. *et al.* (2016) 'VFFDT: A New Software for Preparing AMBER Force Field Parameters for Metal-Containing Molecular Systems', *Journal of Chemical Information and Modeling*. DOI: 10.1021/acs.jcim.5b00687.

Zheng, W., Brooks, B. R. and Thirumalai, D. (2007) 'Allosteric transitions in the chaperonin GroEL are captured by a dominant normal mode that is most robust to sequence variations', *Biophysical Journal*. DOI: 10.1529/biophysj.107.105270.