

https://dspace.org/

Structural basis for the inhibition of Plasmodium falciparum glutathione S-transferase by bromosulphophthalein: insights from structure-function studies

Olfsen, Jessica

2020
https://hdl.handle.net/10539/31038

Downloaded from a DSpace repository. https://dspace.org/



UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

Structural basis for the inhibition of *Plasmodium falciparum* glutathione *S*-transferase by bromosulphophthalein: insights from structure-function studies

by

Jessica Olfsen

(831391)

Masters Thesis

Submitted in the fulfillment of the requirements for the degree

Masters of Science

in

Molecular and Cell Biology

In the Faculty of Science, University of Witwatersrand, Johannesburg, South Africa

Supervisor:

Dr. Ikechuwku A Achilonu

August 2020

Declaration- Plagiarism

I, Jessica Olfsen, declare that this research report is my own work. It is being submitted for the degree of BSc Masters to the University of the Witwatersrand, Johannesburg and has not been submitted for any other degree or examination to any other University.

Signature: 

Date: 20/04/2020

I would like to dedicate this work to my Dad and Mom. Your sacrifice, support and love through my life has been the root of my success. I owe it all to you and will never be able to repay you for what you have done. I love you both so very much.

Thank you.

"You're off to great places

Today is your day

Your mountain is waiting

So get on your way"

Dr Seuss

Acknowledgements

First and foremost, all praise and thanks to God for blessing and helping me throughout my research project and being a constant pillar of strength. I would like to sincerely thank my supervisor Doctor Ikechuwku A. Achilonu from the bottom of my heart for all his support, guidance, and encouragement throughout my project. I have grown so much as a scientist and I strive to be just as great as you one day. I will always be so grateful that I got a supervisor who challenged me constantly and that taught me nothing was impossible. You never gave up on me and I do not know how I will ever be able to repay you for all you have done and taught me. Thank you so very much Dr Ike. I would also like to thank: UNISA who generously allowed me to run my ITC experiments, the University of the Witwatersrand for the Postgraduate Merit Award and the National Research Foundation for funding me, and my colleagues at the Protein Structure-Function Research Unit. Specifically, I would like to thank Dr Roland Worth who helped and guided me through my research project. You always had ideas for me to try and were always there to help me out. I can never thank you enough for what you have done for me. You are not just a great scientist but a good friend. I want to thank my family, specifically my Dad, Mom, and Brother. Even though we live far from one another you never gave up on me and supported me in all avenues of my life. I will never ever be able to repay you Mom and Dad for everything you did and sacrificed for me. You are always my number one supporters and are always there no matter what. I was really blessed when I got you guys as my parents and I am just so thankful. To the Sims family, I want to thank you for your support, encouragement, and love during my studies through all the years. To the Dickson family, thank you so much for taking me in and for all your love and support, not only through my project, but also the writing of my thesis. I will never be able to thank you enough for all you have done for me. To my best friends and their families, specifically Joni and the Symon family, Rena and the Joao family, and Nabeelah and the Coovadia family. I always joke around and say that I have adopted your families but let us be real, I have. Thank you to you all for your love, delicious homecooked meals and support during these years. I am so lucky to call you girls not just my best friends but my sisters. Overall, I want to thank my family and friends as you have all played such an instrumental role in the person I have become and supported me through it all. I love you all so very much. Thank you.

Table of content

Declaration- Plagiarism i

Acknowledgements iii

Table of content iv

List of Abbreviations vi

List of Figures viii

List of Tables ix

Abstract x

Chapter 1 1

Introduction 1

1.1 Problem statement 1

1.2 Rationale 1

1.3 Aim 3

1.4 Objectives 3

1.5 Novelty 4

Chapter 2: 5

Literature review 5

2.1 Background on Malaria 5

2.2 Life cycle 6

2.3 Statistics and Epidemiology 7

2.4 Malaria in South Africa 9

2.5 Treatment 12

2.6 Redox implications in *Plasmodium falciparum* 12

2.7 Glutathione *S*-Transferase 13

2.8 Bromosulphophthalein 22

2.9 Computational Chemistry 24

Chapter 3 26

Expression, purification, and quantification of *Plasmodium falciparum* glutathione *S*-transferase 26

3.1 Introduction 26

3.2 Materials 27

3.3 Methods 29

3.3.1 Transformation and Expression of PfGST 29

3.3.2 Purification of PfGST 29

3.4 Results 33

3.4.1 Expression and Purification of PfGST 33

3.4.2 Quantification of PfGST 33

3.5 Discussion 36
Chapter 4 37
Catalytic and ligandin activity of *Plasmodium falciparum* glutathione *S*-transferase in conjunction with X-ray crystallography 37
4.1 Introduction 37
4.2 Materials 37
4.3 Methods 38
4.3.1 Functional characterisation 38
4.3.2 Structural characterisation 42
4.4 Results 43
4.4.1 Enzyme kinetics of PfGST in the presence and absence of BSP 43
4.4.2 IC_{50} for the interaction between PfGST and BSP 43
4.4.3 Extrinsic fluorescence spectroscopy of PfGST in the presence and absence of BSP 43
4.4.4 Thermodynamics of PfGST-BSP interaction 44
4.4.5 Crystallisation and X-ray diffraction data collection of the PfGST-BSP complex 44
4.5 Discussion 54
Chapter 5 59
Molecular modelling of interactions between *Pf*GST and BSP 59
5.1 Introduction 59
5.2 Materials 59
5.3 Methods 60
5.3.1 Protein structure preparation and model validation 60
5.3.2 Protein and ligand structure preparation for molecular docking 61
5.3.3 MD simulation of PfGST in the absence and presence of BSP 62
5.4 Results 65
5.4.1 Construction and validating the homology modelling of PfGST 65
5.4.2 Induce Fit™ ligand docking, using Glide™ and Prime™, of PfGST and BSP 65
5.4.3 MD simulation of the apo- and BSP-bound conformation of PfGST 71
5.6 Discussion 76
Chapter 6 81
Overall discussion, conclusion, and future studies 81
References 84
Supplementary 90

List of Abbreviations

°C	Degrees Celsius
%	Percentage
Δ*H*°	Change in Enthalpy
Δ*S*°	Change in Entropy
ΔG°	Change in Gibbs free energy
2D	Two-dimensional
3D	Three-dimensional
A_{280}	Absorbance at 280 nm
AMBER	Assisted Model Building with Energy Refinement
ACT	Artemisinin-based combination therapy
ANS	1-anilinonaphthalene-8-sulfonate
BLAST	Basic Local Alignment Search Tool
BSP	Bromosulphophthalein
C_α	Carbon alpha
cDNA	Complementary DNA
CDNB	1-chloro-2,4-dinitrobenzene
CHARM	Chemistry at HARvard Macromolecular Mechanics
DDT	Dichloro-diphenyl-trichloroethane
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
E. coli	*Escherichia coli*
FDA	Food and Drug Administration
FR IX	Ferri/ferroprotoporphyrine IX
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GR	Glutathione reductase
Grx	Glutaredoxin
GSH	Glutathione
G-site	Glutathione-binding site
GSSG	Glutathione disulfide
GST	Glutathione *S*-transferase
HDX-MS	Hydrogen deuterium exchange mass spectrometry
hGSTA1-1	Human class Alpha GST with two type 1 subunits
H-site	Hydrophobic substrates
IC_{50}	Half maximal inhibitory concentration
IFMD	Induce-fit molecular docking

IDA	Iminodiacetic acid
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IRS	Indoor residual spray
ITC	Isothermal titration calorimetry
K_d	Dissociation constant
K_M	Michaelis constant
L-site	Dimer interface
MD	Molecular dynamics
MM	Molecular mechanics
mRNA	Messenger RNA
MUP	Mouse major urinary protein
n	Binding stoichiometry
NaEDTA	Sodium ethylenediaminetetraacetic acid
NMR	Nuclear magnetic resonance
NPT	Isothermal isobaric (temperature and pressure)
NTA	Nitrilotriacetic acid
OD_{600}	Optical density at 600 nm
ORF	Open Reading Frame
OPLS	Optimised Potentials for Liquid Simulations
PDB	Protein Data Bank
*Pf*GST	*Plasmodium falciparum* Glutathione *S*-Transferase
QM	Quantum mechanics
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
ROS	Reactive oxygen species
RPM	Revolutions per minute
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SID	Simulated interaction diagram
SOC media	Super optimal broth with catabolite repression
SPR	Surface plasmon resonance
TCEP	Tris(2-carboxyethyl) phosphine
OPLS	Optimised Potentials for Liquid Simulations
UV-Vis	Ultraviolet-visible
V_0	Initial velocity
V_{max}	Maximum velocity
γ-GCS	γ-glutamylcysteine synthetase

The IUPAC-IUBMB three and one letter codes for amino acids are used

List of Figures

Figure 2.1: Life Cycle of the Malaria Parasite 8
Figure 2.2: Global malaria risk 10
Figure 2.3: South African malaria risk map 11
Figure 2.4: Production of ROS 14
Figure 2.5: Maintenance of the redox balance, specifically looking at *Plasmodium falciparum* glutathione *S*-transferase 15
Figure 2.6: A homodimer of *Plasmodium falciparum* glutathione *S*-transferase (3FRC) 17
Figure 2.7: *Plasmodium falciparum* glutathione *S*-transferase bound to the universal inhibitor S-hexyl-GSH (1Q4J) 18
Figure 2.8: Transcriptome analysis of GST and GADPH in Plasmodium falciparum 21
Figure 2.9: The Chemical structure of BSP 23
Figure 3.1: A schematic displaying the vector construct, the *Pf*GST FASTA sequence and its translatable region 28
Figure 3.2: Expression and purification analysis of *Pf*GST 34
Figure 3.3: Absorbance Spectrum of *Pf*GST 35
Figure 4.1: The specific activity of *Pf*GST in the presence and absence of BSP 46
Figure 4.2: Steady State kinetics plots for *Pf*GST 47
Figure 4.3: Determining the IC_{50} in the presence of BSP 48
Figure 4.4: ANS fluorescence emission spectra of *Pf*GST in the presence and absence of BSP 49
Figure 4.5: The titration profile of *Pf*GST with BSP in the presence and absence of GSH 50
Figure 4.6: X-ray diffraction pattern of *Pf*GST-GSH-BSP 53
Figure 5.1: Ribbon representation of the alignment between *the* original *Pf*GST structure (green) and modelled *Pf*GST (blue) 66
Figure 5.2: *PF*GST-BSP complex 69
Figure 5.3: The ligand-protein interactions that occurred during IFMD of *Pf*GST and BSP 70
Figure 5.4: RMSD of the apo *Pf*GST vs the *Pf*GST in complex with BSP 72
Figure 5.5: RMSF of the residues side chain when the protein is complexed with and without BSP, over a 25 ns simulation 74
Figure 5.6: Interaction fractions of various residues during molecular dynamic simulations 75

List of Tables

Table: 3.1 Non-standard reagents and material, used in the different experimental methods, with their appropriate manufacturers 34

Table 4.1: Non-standard reagents and materials, used in the different experimental methods 44

Table 4.2: The thermodynamic of the *Pf*GST-BSP interaction, in the presence of GSH, based on a sequential two site model 58

Table 4.3: Crystallisation hits for the *apo* and complex forms of *Pf*GST in combination with GSH and BSP 59

Table 5.1: Analysis of the Ramachandran plot stereochemical properties of the *Pf*GST model before and after homology modelling 74

Table 5.2: A summary of the docking scores produced from IFMD of *Pf*GST with BSP 75

Abstract

Malaria, caused by the protozoan parasite Plasmodium, is the second most common cause of infectious disease-related death in the world. The species responsible for most of the deaths is *Plasmodium falciparum*. Presently, there are limitations of current drugs for malaria including inhibitors that are specific for their target. *P. falciparum* glutathione *S*-transferase (*Pf*GST) has emerged as a promising therapeutic target as it is one of the several components that are responsible for maintaining the redox balance in the malaria parasite. Therefore, this study aimed to understand the inhibition of *Pf*GST by bromosulphophthalein (BSP), a non-substrate anionic ligand. Recombinant *Pf*GST was over-expressed in *Escherichia coli* and purified using metal affinity chromatography. The enzyme has been purified and the specific activity was assessed using the glutathione-1-chloro-2,4-dinitrobenzene (GSH-CDNB) conjugation reaction, in the presence and absence of BSP. Once inhibition was evident, an IC_{50} was performed. Extrinsic fluorescence spectroscopy was then performed in order to determine the binding of BSP to *Pf*GST. The thermodynamics of the *Pf*GST-BSP interaction was then determined, using isothermal titration calorimetry (ITC). Crystal studies, along with molecular modelling was done in order to determine the binding site of BSP. The specific activity of *Pf*GST in the absence and presence of BSP was 0.84 µmol/min/mg and 0.34 µmol/min/mg respectively, at concentration of 537 µM, BSP inhibits the *Pf*GST by 50 %. The K_M of *Pf*GST was 5.41 mM for GSH. Fluorescence spectroscopy suggests that BSP binds near the H-site of the enzyme, therefore, inhibiting the conjugation of glutathione to CDNB at the G-site. ITC showed that *Pf*GST has two types of co-operative binding sites for BSP which are exergonic. The low affinity site was proposed to be BSP binding to the H-site, which was enthalpically favourable and entropically unfavourable. The high affinity site was proposed to be BSP binding in the L-site, which was which enthalpically unfavourable and entropically favourable. The binding of BSP at the L-site was supported by molecular docking and dynamics studies. The top ranked pose had a negative Glidescore and a Glide emodel score of -9.08 and -122.35 $kcal.mol^{-1}$ respectively, which suggests tight binding of BSP to the L-site. Both *in-silico* studies showed that TYR 108 of *Pf*GST was a key residue in π-π interaction with BSP. The interaction of TYR108 and the ligand has been showed in other enzyme studies as a key interaction in the inhibition of the enzyme. Therefore, the detoxification and redox balance activities of *Pf*GST is paramount for the survival of the parasite; hence, BSP could be a valuable inhibitor of this important biological processes within the parasite.

Chapter 1

Introduction

1.1 Problem statement

Malaria can be considered as an emerging disease due to the increase in the resistance to antimalarials and growing geographical distribution of *Plasmodium falciparum*, therefore, threatening individual's health and welfare (Conrad and Rosenthal, 2019). Malaria is the second most common cause of infectious disease-related death in the world with 219 million new cases globally in 2017, of which the African region accounts for 92% of the global cases (World Health Organisation, 2018). There were an estimated 435 000 deaths (93% are accounted for by Africa region) of which 266 000 deaths occurred in children under the age of 5 globally (World Health Organisation, 2018). Even though there was a global decrease of 20 million cases of malaria between 2010 and 2017, there was no substantial progress in the overall reduction of malarial cases globally (World Health Organisation, 2018). Due limitations faced by current drugs for malaria which include; drug resistance in both first- and second-line of drug therapy, low efficacy, safety issues, poor compliance and the high cost of artemisinin-based combination therapy (ACT) (Kar and Kar, 2010), there is a need to develop new drugs that are directed to novel targets, which can be used in combination with existing antimalarial therapies (Yadav *et al*., 2019).

1.2 Rationale

The parasite is susceptible to oxidative stress and could be a potential therapeutic target for drug action (Savvides and Karplus, 1996). Non-enzymatic degradation of free heam (ferri/ferroprotoporphyrine IX (FR IX)) or the reduction of glutaredoxin (Grx) causes the formation of glutathione disulfide (GSSG) (Müller, 2004). The GSSG is either reduced back to GSH via the enzyme glutathione reductase (GR) or removed via GSSG-efflux system (Müller, 2004). Due to a loss of some of the GSSG via the GSSG-efflux system, γ-glutamylcysteine synthetase (γ-GCS) and glutathione synthetase replenish the GSH-pool by producing GSH (Müller, 2004). An enzyme known as glutathione S-transferase (GST) detoxifies the parasite through its binding to electrophilic harmful compounds, which are then removed from the parasite by a GS-X pump (Müller, 2004). It also functions as an alkyl hydroperoxide (ROOH)

peroxidase. Therefore, GSSG produced from the reductive reactions of the GST goes through the glutathione redox cycle via GR enzyme and be recycled back to GSH (Müller, 2004). Consequently, inhibition of these enzymes could cause an increase in oxidative stress in the asexual erythrocytic stage of the *Plasmodium* parasite leading to the inhibition of transmission of the parasite (Sarma *et al.*, 2003).

Genome sequencing of *Plasmodium falciparum* revealed that *Pf*GST is the only isoform (Deponte and Becker, 2005). In addition, it cannot be assigned to a GST subclass as its H-site is more solvent exposed, due to the truncation of its amino acid sequence (Perbandt *et al.*, 2015). It has been suggested that the H-site, being more solvent exposed, allows for winder range of inhibitors to bind (Fritz-Wolf *et al.*, 2003). Therefore, amphiphilic compounds can access the H-site of the *Pf*GST, therefore, evading the human isoform's H-site (Fritz-Wolf *et al.*, 2003). Therefore, the binding of the compound will result in the disturbance of the GSH conjugation to xenobiotic compounds, causing an increase level of cytotoxic peroxides, which may lead to an increase in toxic haemin (Deponte and Becker, 2005). Haemin, in the presence of GSH, can inhibit the *Pf*GST in the lower micromolar range, suggesting that the free hemin, *in vivo*, might be buffered by the enzyme (Harwaldt *et al.*, 2002; Liebau *et al.*, 2002). In addition, chloroquine has been shown to inhibit the hemin catabolism which then causes increased intracellular haemin (Famin *et al.*, 1999). Therefore, the inhibitors of *Pf*GST and chloroquine could be interdependent (Deponte and Becker, 2005). Another difference of this enzyme is that, where there is no GSH present, the enzymes form non-covalently linked higher aggregates, which are inactive. However, in the presence of GSH, the enzyme becomes a functionally active dimer (Tripathi *et al.*, 2007). Therefore, if the GSH binding is disrupted, the enzyme will be rendered inactive, therefore, causing intracellular oxidative stress (Deponte and Becker, 2005). Salinas and Wong (1999) showed that drug resistance development for both cancer therapy and the therapy of parasitic diseases was a result of GST, which aids in the clearance of the drugs. The activity of GST is shown to be related to the drug pressure in chloroquine-resistant parasites (Dubois *et al.*, 1995; Srivastava *et al.*, 1999). Using plasmoDB (https://plasmodb.org/plasmo/) *Pf*GST has also been found at comparable levels to GAPDH, a known housekeeping enzyme in the parasite. This suggests that *Pf*GST is an important enzyme during the blood stage

of the parasite. Therefore, the loss of the *Pf*GST, could cause a build-up of cytotoxic species in the parasite, leading to the elimination of the parasite.

The inhibitor explored in this study was bromosulphophthalein (BSP). The dye was is used in the testing of the liver function, however, it was shown by Kolobe *et al.* (2004), as a non-substrate anionic ligand that interacts and inhibits the low affinity binding site of the human class Alpha GST with two type 1 subunits (hGSTA1-1). Therefore, due to the binding of subclass Alpha of the GST family, this study investigated BSP's interaction and possible inhibition of *Pf*GST enzyme. Therefore, inhibition of *Pf*GST could allow for the scaffolding of the BSP molecular, resulting in novel inhibitors, targeted only to *Pf*GST, which effectively contribute to the cytotoxic effects and inhibits the formation of hemozoin, leading to deleterious effects and elimination of the parasites (Müller, 2004).

1.3 Aim

The aim of the study was to biophysically characterise the *Pf*GST enzyme and evaluate the binding and effect of BSP on the *Pf*GST's enzyme activity.

1.4 Objectives

In order to accomplish the aim of the study, the objectives of the study were to:

- Overexpress in *Escherichia coli* (*E. coli*) and purify recombinant *Pf*GST using immobilised metal affinity chromatography (IMAC).
- Quantify and assess the quality of *Pf*GST using ultraviolet (UV-Vis) spectrophotometry.
- Determine the specific activity of *Pf*GST in the presence and absence of BSP.
- Determine the IC_{50} of BSP inhibition of *Pf*GST.
- Determine the tertiary structure of *Pf*GST using extrinsic florescence in the presence and absence of BSP.
- Crystallise *Pf*GST in complex with BSP.
- Determine the thermodynamic parameters of the interaction between *Pf*GSTs and BSP, in the presence of GSH.

- Use Molecular modelling studies to understand and describe the interactions between *Pf*GST and BSP.

1.5 Novelty

Malaria, due to *Plasmodium falciparum*, is still a relevant disease based on the high infection and mortality rates, along with resistance and increase cost of certain treatments. Therefore, there is a need for the improvement of current treatment of the disease. BSP is known to inhibit hGSTA1-1 (Kolobe *et al.*, 2004). However, there has been no study done on the interaction and possible inhibition of *Pf*GST using BSP. Both enzyme kinetic studies, along with inhibition studies, extrinsic fluorescence, and isothermal titration calorimetry (ITC) were performed to further understand the binding and inhibition of *Pf*GST by BSP. In addition, the *Pf*GST vector was synthesised with the addition of a his-tag, which was then successfully expressed and purified using IMAC.

Chapter 2:

Literature review

2.1 Background on Malaria

Malaria, caused by the protozoan parasite Plasmodium, is known to be of ancient origins and has been documented in 2700 BC in China, 2000 BC in Mesopotamia, 1570 BC in Egypt and from the 6th century in Hindu texts (Cox, 2010). Malaria, meaning spoilt air, was first believed to be caused by miasmas rising from swamps which caused poor health, malarial fevers and larger spleens (Cox, 2010). However, Charles Louis Alphonse Laveran discovered parasites in 1880 and, specifically, that the vectors of the parasite were female *Anopheles* mosquito (Cox, 2010). There are two clinical types of malaria; uncomplicated and severe malaria (Autino *et al.*, 2012). In uncomplicated malaria, symptoms develop within 10-15 days of the infection and include a cold sensation, shivering, fever, headaches, sweating and seizures (Autino *et al.*, 2012). Severe malaria, characterised microscopically by a parasitemia level that is greater than 5%, is due to infection caused by *P. falciparum (*Autino *et al.*, 2012*).* Symptoms include organ damage, blood abnormalities, cerebral malaria, haemolysis which may lead to severe anaemia, pulmonary edema, acute respiratory distress syndrome, thrombocytopenia, renal failure, and cardiovascular collapse (Autino *et al.*, 2012). This ultimately leads to death (Kar and Kar, 2010).

There are five major species which cause malaria in humans and are namely: *P. vivax, P. malariae, P. ovale, P. knowlesi* and *P. falciparum* (Autino *et al.*, 2012). *P. falciparum,* which is responsible for most fatalities (Schantz-Dunn and Nour, 2009), is predominantly found in tropical Africa (Autino, *et al*., 2012). *P. vivax* is found dominantly in South America (Autino, *et al*., 2012). Both, *P. falciparum* and *P. vivax* are found in South-Eastern Asia, along with. *P. knowlesi,* and Western Pacific (Autino, *et al*., 2012). *P. malariae* is usually found at lower levels in all malarial areas; however, co-infection with *P. falciparum* occurs in tropical Africa (Autino, *et al.*, 2012). *P. ovale is found* in tropical Africa (Autino, *et al*., 2012).

2.2 Life cycle

The malaria life cycle is one of complexity, involving two living organisms mainly the *Anopheles* mosquito (the vector) and humans (vertebrate host) (Floren, *et al*., 2002). In Figure 2.1, the life cycle of malaria is provided. It consists of two hosts: the mosquito (sexual stage of life cycle) and the human (asexual stage of life cycle). This transmission is promoted by the mutually beneficial relationship between the mosquito and the *Plasmodium* parasite (Ferguson and Read, 2004). The transmission of the malarial parasite occurs when an infected female *Anopheles* mosquito draws blood from a human (Schantz-Dunn and Nour, 2009). The parasite is transferred from the infected mosquito's saliva, which enters the human capillary bed as *Plasmodium* sporozoites. This is classified as the asexual stage of the mosquito's life cycle and consists of both a hepatic and an erythrocytic infection (Schantz-Dunn & Nour, 2009). The sporozoites then migrate to the liver of the human host and replicate to form merozoites (Schantz-Dunn & Nour, 2009). Once fully formed, the merozoites rupture the liver cells (Florens *et al*., 2002) and migrate into the bloodstream invading erythrocytes by attaching to them. The invasion by merozoites is done through low affinity interactions between the parasite and the erythrocyte, followed by the reorientation of the parasite and formation of electron dense contacts, also known as a tight junction, to the erythrocyte (Salina *et al*., 2019). A moving junction with an active myosin motor allows for the invasion of the parasite, which sheds its surface protein coat and forms a parasitophorous vacuole within the erythrocyte (Salina *et al*., 2019). Once the invasion is complete, asexual replication within the parasitophorous vacuole occurs. This allows for the formation of the daughter merozoites which then rupture the erythrocyte and released into the blood stream where it can then invade other erythrocytes (ref). The invasion, replication and lysis of the erythrocytes along with phagocytosis of the infected blood cells by the spleen may cause anaemia and folic acid defiance in the body (Schantz-Dunn & Nour, 2009). Malaria symptoms then arise due to the infection of the red blood cells, along with the inflammatory response of the immune system (Schantz-Dunn & Nour, 2009).

A small percent of the parasite, during the blood stage, form the male and female gametocytes, which remain in the bloodstream (Barillas-Mury and Kumar, 2005) until a non-infected mosquito draws blood from the infected human. This is the beginning

of the sexual (*sporogony*) of the mosquito's life cycle (Florens *et al.*, 2002). The gametes, once in the gut of the mosquito, fuse to form the zygote which then forms into actively moving ookinetes. The ookinetes burrow into the midgut wall of the mosquito where they develop into oocysts which grow and divided into many active haploid sporozoites (Barillas-Mury and Kumar, 2005). The sexual stage of the life cycle occurs within 8-15 days and once completed, the sporozoites are released into the mosquito's body cavity and invade the mosquito's salivary glands. This is then the start of the asexual stage of the life cycle (Barillas-Mury and Kumar, 2005).

2.3 Statistics and Epidemiology

Malaria is the second most common cause of infectious disease-related death in the world with 219 million new cases globally in 2017, of which the African region accounts for 92% of the global cases (World Health Organisation, 2018). Between 2010 and 2017, there was a global decrease (20 million cases) in malarial cases, however, there was no substantial progress in the overall reduction of malarial cases globally. In 2017, there were an estimated 435 000 deaths, of which 93% are accounted for by Africa regions (World Health Organisation, 2018). There were 266 000 deaths which occurred in children under the age of 5 globally, of which 247 380 occurred in African region (World Health Organisation, 2018). For the African malaria cases in 2017, *P. falciparum* accounted for 99.7% of the cases. It also accounted for 62,8% of South-East Asia cases, 69% of the Eastern Mediterranean cases and 71.9% of the Western Pacific cases. Below, the transmission of malaria in different countries and regions can be observed for 2017 in Figure 2.2 A. The species of greatest importance i.e. *P. falciparum*, due to its increase mortality and the region in which it is found (tropical Africa), distribution is observed below in Figure 2.2 B.

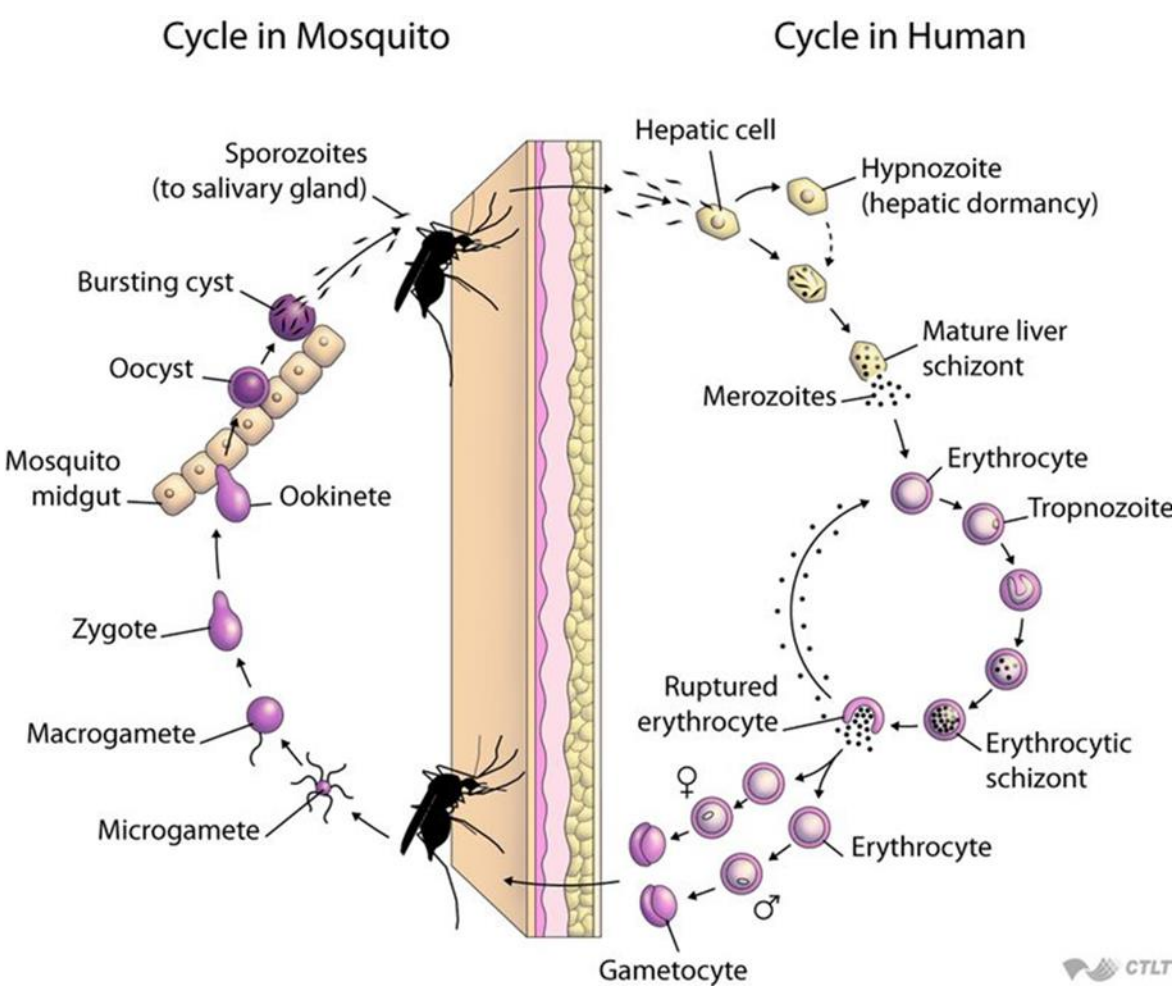
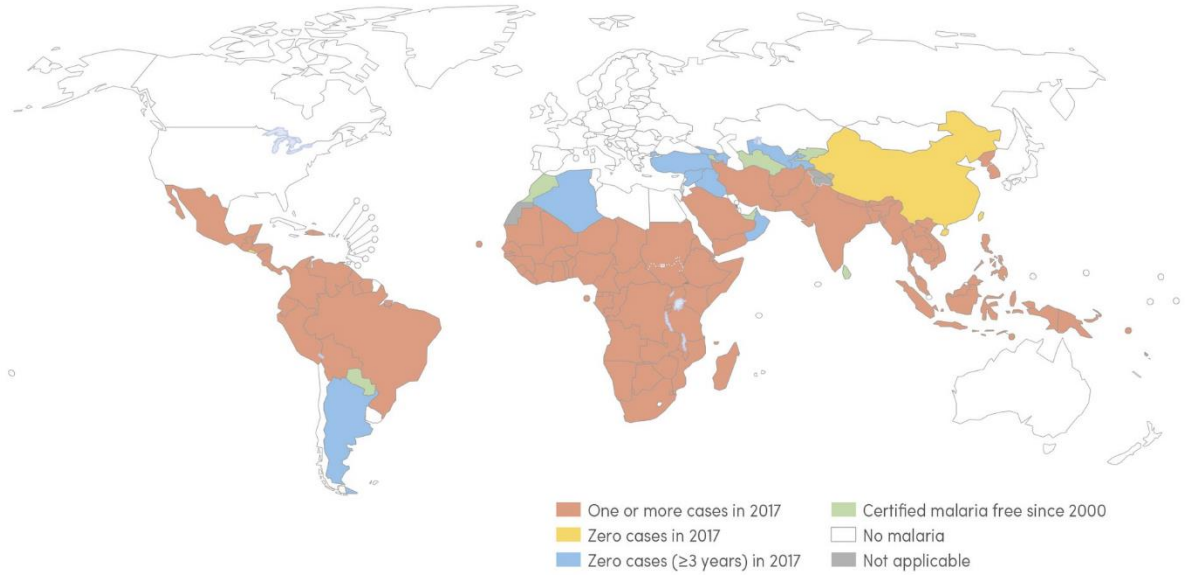


Figure 2.1: Life Cycle of the Malaria Parasite. The cycle showing the two stages, sexual and asexual, occurring in the mosquito and human, respectively. (From Epidemiology of Infectious Diseases. Available at: http://ocw.jhsph.edu. Copyright © Johns Hopkins Bloomberg School of Public Health. Creative Commons BY-NC-SA)

2.4 Malaria in South Africa

P. falciparum is the main malaria parasite found in South Africa (Maharaj *et al.*, 2019) with *Anopheles Arabiensis* being the most probable mosquito vector (Brooke *et al.*, 2013). Malaria is seasonal in South Africa and generally occurs between September and May (Maharaj *et al.*, 2019). The infected areas lie along the northern and north-eastern border of South Africa and include Limpopo, Mpumalanga and KwaZulu-Natal, shown in Figure 2.3 (Maharaj *et al.*, 2019). Robust vector controls, implemented in the early 1940's, resulted in effective control and near elimination of malaria by 1970's. However, environmental changes including rainfall and flooding resulted in an increase in malaria cases, occurring between 1972 and 1973 (Raman *et al.*, 2016). There was a major increase in malaria cases in 2016, with South Africa having the highest number of cases reported post-2000 epidemic, in 2017 (Maharaj *et al.*, 2019). In the mid-1980's, there was a decline in malaria cases due to increased surveillance, prompt treatment with chloroquine and the use of dichloro-diphenyl-trichloroethane (DDT)-based indoor residual spray (IRS) (Coetzee *et al.*, 2013). However, due to chloroquine-resistant parasite, a minor increase in malaria cases occurred in the mid- to late-1980's which was counteracted by replacing the first-line treatment, chloroquine, with sulfadoxine-pyrimethamine (Ukpe *et al.*, 2013). A malaria outbreak, with over 60 000 reported cases, between 1999 and 2000 saw the country's worst malaria epidemic to date. Factors contributing to this epidemic included favourable environmental changes (elevated temperatures and rainfall), a rise in immigration across the country's borders, drug-resistant parasites, and insecticide-resistant vectors. Since the epidemic, the immigration control, introduction of artemisinin-based combination therapy (ACT) and re-introduction of DDT for IRS operations has decreased the indigenous malaria cases to 9500 and 8700 imported cases of malaria, in 2018 (World Health Organisation, 2019).



WHO: World Health Organization.

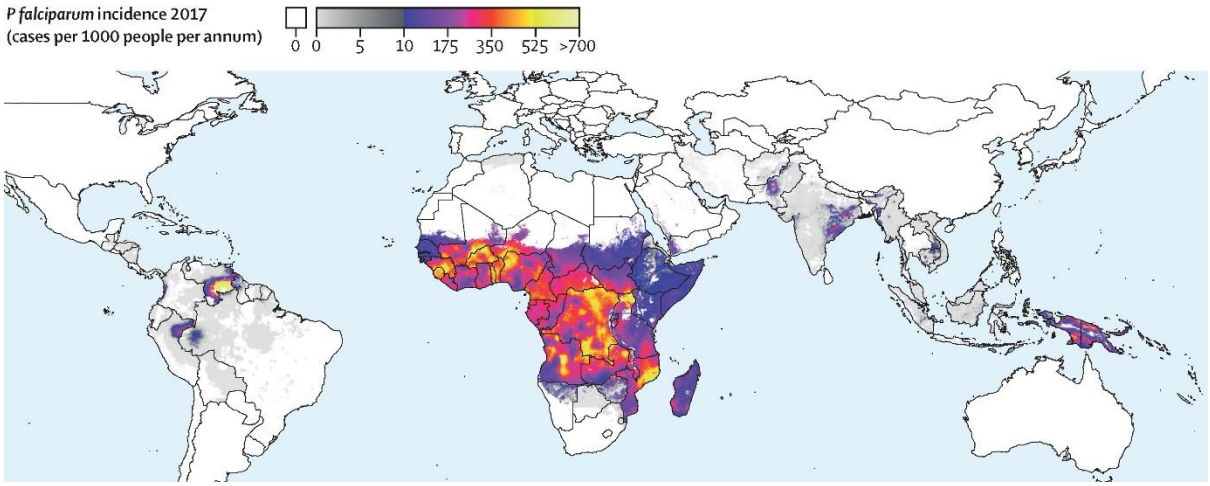


Figure 2.2: Global malaria risk. (A) A world map showing the distribution of territories of malaria transmission in 2017. The above figure shows countries or areas where malaria cases occurred (red), zero malarial cases (yellow), zero malarial cases ≥ 3 years (blue), malaria free since 2000 (green), no malaria (white) and not applicable (grey). (B) A world map showing the incident (cases per 1000 people per annum) of P. falciparum in tropical areas of Africa, South America and Asia (Weiss *et al.*, 2019).

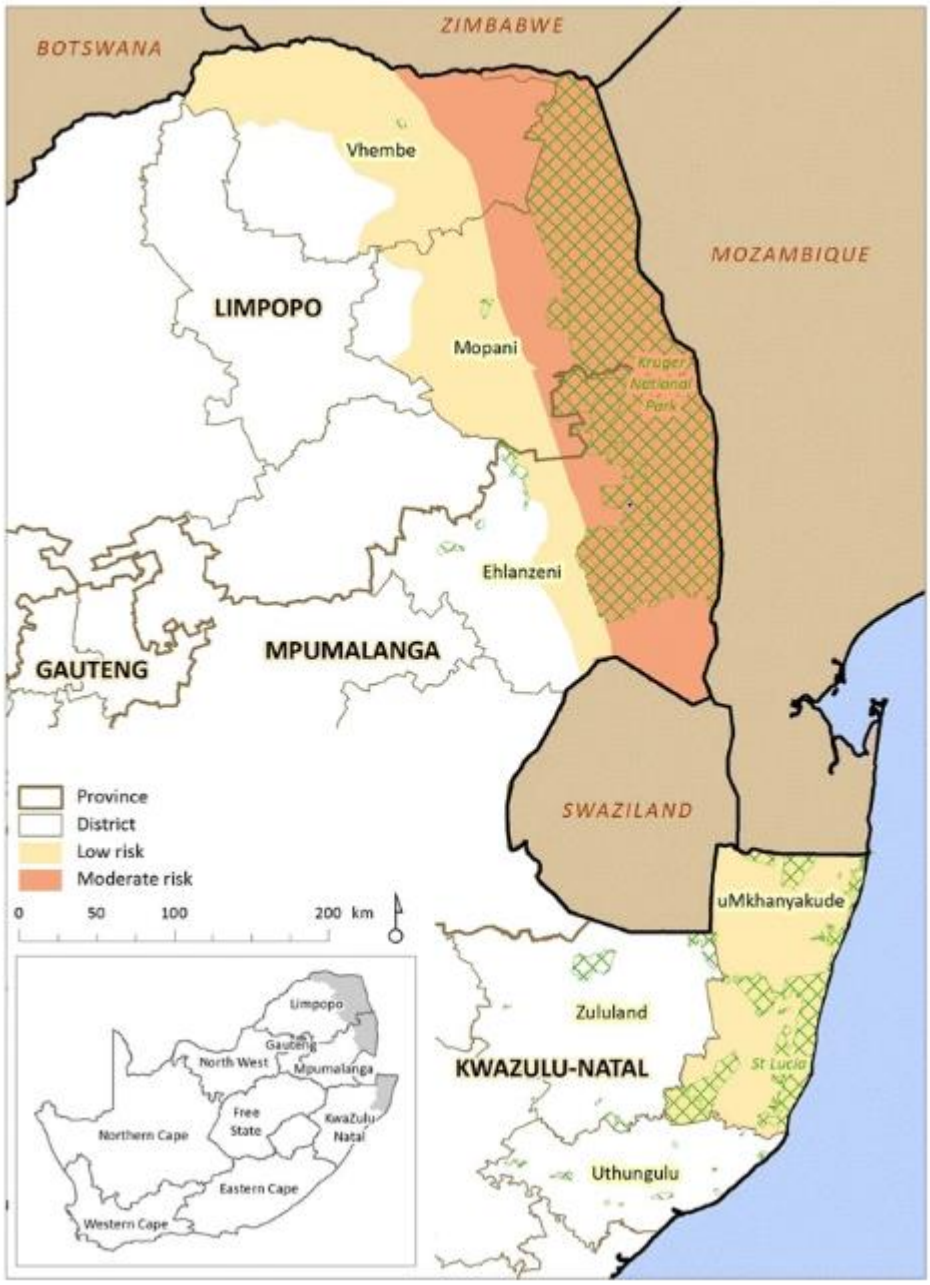


Figure 2.3: South African malaria risk map. A distribution of territories of malaria in 2016 which shows areas with both low (peach) and moderate (orange) risk (Raman *et al.*, 2016).

2.5 Treatment

There are many antimalarial drugs available. A well-known antimalarial drug chloroquine, which interferes with haem detoxification (Wells *et al.*, 2009). Sulphadoxine, pyrimethamine and proguanil are used to target folate metabolism (Wells *et al.*, 2009). Atovaquone which inhibits the mitochondrial electron transport and is also used as a prophylaxis (Wells *et al.*, 2009). Pyrimethamine and Proguanil are also used as prophylaxis for malaria (Wells *et al.*, 2009). However, the current drug for prophylaxis is a fixed-dose atovaquone–proguanil combination (Wells *et al.*, 2009). The downfall of this anti-malaria drug is that it is expensive in comparison to the others (Wells *et al.*, 2009). A major problem faced with the drugs stated above is that the parasite has developed first- and second-line resistance, with other added problems like low efficacy, safety issues, ineffective compliance and the prohibitive cost of the drugs (Kar and Kar, 2010).

A solution that was explored was combination drug treatment, which has different mechanisms of action and decreases the resistance to the drug. Artemisinin-based combination therapy (ACT) was recommended by WHO and has been found to be 95% effective when the drug in combination with it is selected carefully (Kar and Kar, 2010). However, disadvantages of this therapy include the cost of the therapy, especially for a developing country (Kar and Kar, 2010) and in recent years there has been resistance that has emerged.

Other methods such as insecticide-treated bed nets and the spraying of insecticides that are environmentally friendly which have helped decrease the incident, however, there is still a need for a new generation of anti-malarial therapeutics due to these limitations faced by current drugs for malaria including; drug resistance, low efficacy, safety issues, poor compliance and high cost, and the high mortality rate that is faced globally.

2.6 Redox implications in *Plasmodium falciparum*

All aerobic life requires sufficient coping mechanisms in order to maintain intracellular redox environments which reduce the generation of reactive oxygen species (ROS), caused by essential metabolic functions that take place (Imlay, 2003).

During the intraerythrocytic of the *P. falciparum,* the increased level in ROS is due to haemoglobin digestion which takes place in the food vacuoles of the parasite (Müller, 2004). The digestion results in the release of free haem (ferri/ferroprotoporphyrine IX (FP IX)), of which most is biomineralised to haemozoin. However, some free haem (FP IX) (Loria *et al.*, 1999; Zhang *et al.*, 1999) manages to escape the food vacuole, causing damage to the membrane of the cells, along with the generation of superoxide anions (Müller, 2004). This is a result of oxidation of the haem-iron that then reacts with H_2O_2, forming highly reactive hydroxyl radicals which causes lipid peroxidation, leading to cellular damage (Figure 2.4). Therefore, the parasite relies on systems which allow for the maintenance of the redox balance (Bunik, 2003). This balance is maintained through the ratios of reduced and oxidised pyridine nucleotides ($NAD(P)H/NAD(P)^+$), thiols such as glutathione/glutathione disulphide (GSH/GSSG) and thioredoxin/thioredoxin disulphide ($Trx(SH)^2/Trx(S)^2$), which is regulated through enzymatic functions (Bunik, 2003).

2.7 Glutathione *S*-Transferase

An important enzyme, which plays a role in the redox maintenance within the parasite is *P. falciparum* glutathione *S*-transferase (*Pf*GST). The phase II detoxification enzyme, *Pf*GST [EC: 2.5.1.18], catalyse the nucleophilic addition of GSH to a large variety of non-polar or electrophilic compounds, which neutralizing the electrophilic sites on the compound (Chasseaud, 1979; Salinas and Wong, 1999). The compound in now more water soluble and can be excreted out of the cell, allowing for the maintenance of the cell's redox environment, seen in Figure 2.5 (Müller, 2004). There are multiple forms of GST enzymes which are found across many species and generally take part in intracellular detoxification of harmful substances like mutagens, carcinogens, and other noxious chemical substances (Fritz-Wolf *et al.*, 2003). In addition, certain GSTs are known for their antioxidant role in the cell by detoxifying lipid peroxidation products (Bruns *et al.*, 1999). The enzymes are generally homo- or heterodimeric and are divided into different subclasses which are mainly Alpha (α), Mu (μ), Pi (π), Delta (σ), and Theta (θ), which is based on sequence similarity, immunological cross-reactivity and specificity towards the electrophilic second substrate (Fritz-Wolf *et al.*, 2003).

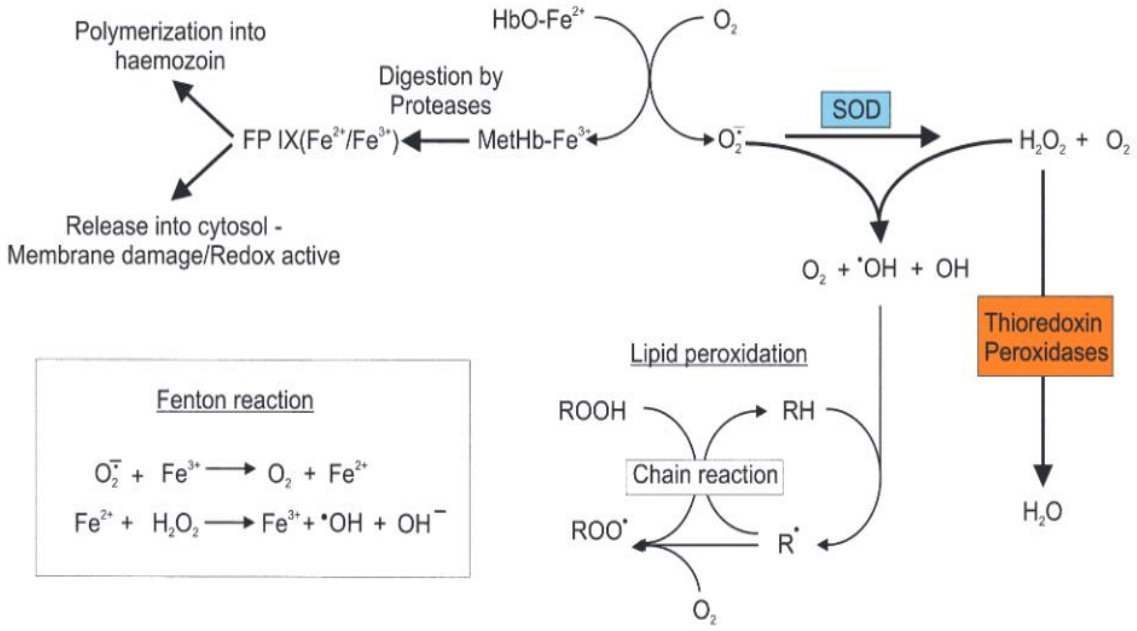


Figure 2.4: Production of ROS. A diagram showing the sources of ROS due mainly to the digestion of host cell haemoglobin, in *P. falciparum*. The release of free haem (FP IX) can cause damage to the cellular membrane and the generation of superoxide anions, which react with H_2O_2, forming hydroxyl radicals that lead to lipid peroxidation. *P. falciparum* has developed systems that maintain the intracellular redox environment (Müller, 2004).

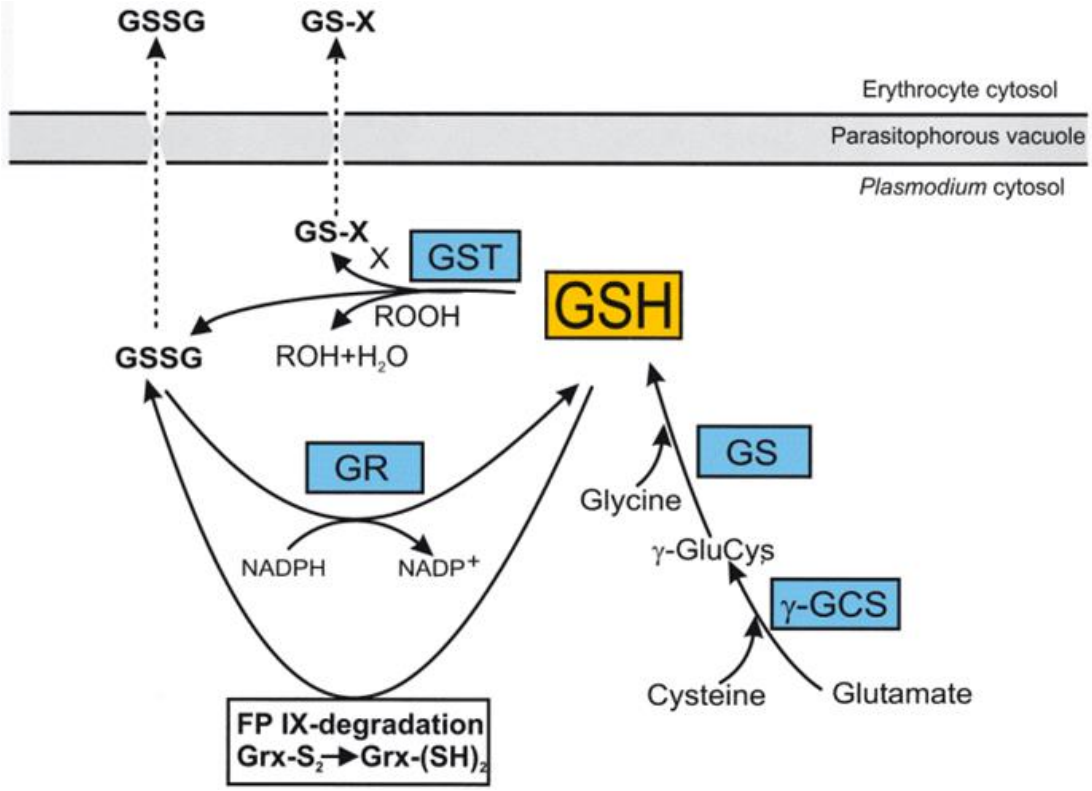


Figure 2.5: Maintenance of the redox balance, specifically looking at *Plasmodium falciparum* glutathione *S*-transferase. For the parasite to sustain a redox balance, it is important for an appropriate ratio of glutathione (GSH) and glutathione disulphide (GSSG) to be maintained (Müller, 2004). Non-enzymatic degradation of FP IX or the reduction of glutaredoxin (Grx) causes the formation of GSSG (Müller, 2004). The GSSG is either reduced back to GSH via the enzyme glutathione reductase (GR) or removed via GSSG-efflux system (Müller, 2004). Due to a loss of some of the GSSG via the GSSG-efflux system, γ-glutamylcysteine synthetase (γ-GCS) and glutathione synthetase replenish the GSH-pool by producing GSH (Müller, 2004). *Pf*GST detoxifies the parasite through its binding to electrophilic harmful compounds (X), which are removed from the parasite by a GS-X pump (Müller, 2004). It also functions as an alkyl hydroperoxide (ROOH) peroxidase. Therefore, GSSG produced from the reductive reactions of the *Pf*GST goes through the glutathione redox cycle via GR enzyme and be recycled back to GSH (Müller, 2004).

Fritz-Wolf *et al*. (2003) recombinantly expressed (using *E. coli*) and purified *Pf*GST. After determination that the enzyme was active, it was crystallised and later solved at a 1.9 Å resolution using single isomorphous replacement with anomalous scattering (Fritz-Wolf *et al.*, 2003). This was achieved through soaking native crystals in mother liquor, which contained mercury (0.1 mM MeHgCl) (Fritz-Wolf *et al.*, 2003). The space group was determined as $P2_12_12$, with the asymmetric unit containing the two monomers A and B (Fritz-Wolf *et al.*, 2003). Each monomer contains two domains (Fritz-Wolf *et al.*, 2003). Domain 1 has three α-helices that surround a four-stranded mixed β-sheet (βαβαββα) (Fritz-Wolf *et al.*, 2003). This domain is linked to domain 2, which is mostly α-helical, by a loop formed by residues TYR 83 to GLU 88 (Fritz-Wolf *et al.*, 2003). Unlike known GST's, who's enzymatically active dimer is formed using the A and B monomer, the two active dimers are formed by the monomers A and A′ and monomers B and B′ respectively (Fritz-Wolf *et al.*, 2003). The monomers A and B are related to A′ and B′ by a twofold crystallographic c-axis. However, like other GST superfamily, *Pf*GST adopts the canonical GST fold with the active site located in the cleft between the two domains of the monomer (Fritz-Wolf *et al.*, 2003). This contains the two active sites in their crystal structure, which are namely the G- and H-site.

The G-site within the *Pf*GST has a common backbone fold that is conserved in other GST enzymes (Fritz-Wolf *et al.*, 2003). This site binds GSH, its natural substrate, thereby allowing activation of *Pf*GST that binds toxic compounds, which are then removed from the parasite (Müller, 2004). During the activation process, there is a unique transition from the inactive tetramer state to the active dimeric state (Tripathi *et al.*, 2007). This is not found in any other GST class and makes *Pf*GST a potential target for malaria therapeutic strategies (Sarma *et al.*, 2003). Figure 2.6 below shows the dimeric *Pf*GST state with bound GSH. When GSH is bound, there are two hydrogen bonds which form between *Pf*GST and GSH through the backbone of *Pf*GST Chain A's VAL 59 and the backbone of the cysteine moiety of GSH (Fritz-Wolf *et al.*, 2003). The SH group of the cysteine moiety is then stabalised by a conserved tyrosine residue (*Pf*GST's chain A's TYR 9) (Fritz-Wolf *et al.*, 2003). This was in agreeance with a study done by Perbandt *et al.* (2004), using the universal inhibitor S-hexyl-GSH (GTX), who demonstrated that in addition to the TYR 9, 2 other conserved residues, GLN 71 AND ASP 105, interact with the GST/GTX when



Figure 2.6: A homodimer of *Plasmodium falciparum* glutathione *S*-transferase (3FRC). A homodimer consisting of chain A in green and chain B in blue, each chain consisting of 208 amino acids. Only one active site is shown with the GSH molecule, which is bound at the G-site, circled in red. The H-site has been shown with a black arrow. The structure was extracted from the Protein Data Bank (PDB) with its respective code and the image was generated using Pymol V1.3.

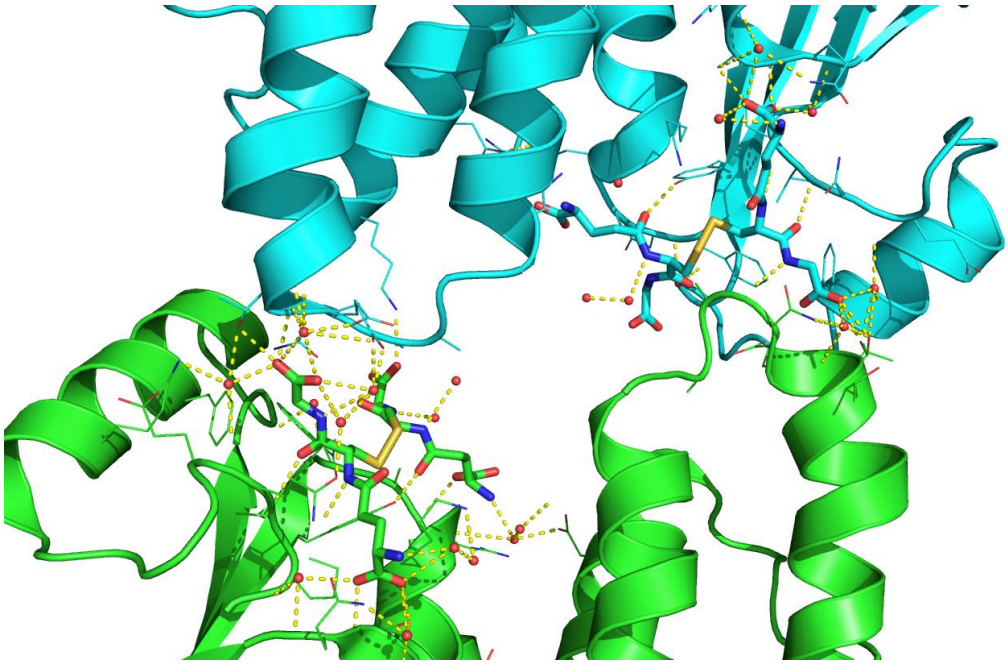


Figure 2.7: *Plasmodium falciparum* glutathione *S*-transferase bound to the universal inhibitor S-hexyl-GSH (1Q4J). The GTX molecule binds to both the G- and H-site of PfGST. The binding is facilitated by polar interactions between the protein and the tripeptide. Binding of the GTX predominantly occurs at the G-site, however, S-hexyl moiety of GTX interacts with the H-site (Perbandt et al., 2004).

it binds to the G-site (Figure 2.7). These 3 residues are also found within the mammalian Mu-class, therefore attributing to the Mu-class character of the *Pf*GST's G-site (Perbandt *et al.*, 2004). The GSH's glycine moiety is bound to *Pf*GST chain A's PHE 45 and LYS 49 (similar to π- and μ GST class). Another residues that both studies found that play's a role in the stabilisation of the GSH is LYS 15 (Perbandt *et al.*, 2004).

Contrasting the G-site, the H-site binds a variety of substrates and is usually formed by a loop that connects β1 and α1, the C-terminal part of helix α4, and the residues after helix α8 (Fritz-Wolf *et al.*, 2003). However, unlike other GST classes, *Pf*GST has a truncation of its amino acid sequence after helix α8 (Perbandt *et al.*, 2015). This forms a short turn instead of the typical Mu-loop that shields the H-site of the Mu-class GST, resulting in *Pf*GST's H-site being more solvent exposed in comparison to other GST classes (Fritz-Wolf *et al.*, 2003). Therefore, this allows for a winder range of inhibitors to bind for potential drug therapeutic strategies (Fritz-Wolf *et al.*, 2003). Perbandt *et al.* (2004) also showed that the H-site of *Pf*GST has a higher similarity to the Pi-class enzyme. Therefore, having features of both the Mu- and Pi-class GST's, *Pf*GST is considered a 'chimeric enzyme' (Perbandt *et al.*, 2004).

In the absence of GSH, the active dimers form a tetramer through the interlocking of loop 113-119, which connects two conserved α helices, that occupy the H-site of the other monomer and vice versa (Perbandt *et al.*, 2015). This loop is made up the residues TNLFKQN which are involved in both hydrogen bonding and mediating majority of the hydrophobic interactions (Perbandt *et al.*, 2015). Perbandt *et al.* (2015) showed that loop was important for the tetramerization process by using a mutated GST (*Pf*GSTΔ114-118), which lacked residues making up the loop. It was shown that in the absence of GSH, the tetramerization was 100 times slower with *Pf*GSTΔ114-118 in comparison to the wild-type *Pf*GST (Perbandt *et al.*, 2015). Therefore, the loop plays a role in the tertiarization process, however, the stabilisation of the tetramer is proposed to be the interaction between ASN 112 and LYS 117 (Perbandt *et al.*, 2015). Following the interlocking of the loop and the occupying of the H-Site, the now blocked active site region renders the enzyme inactive (Perbandt *et al.*, 2015). A study done by blah et al found that the oxidised form of GSH, glutathione disulphide (GSSG), prevented the inactivation process (Perbandt *et al.*,

2015). This was done via the stabilisation of the dimeric structure in combination of the GSSG compound behaving as a competitive inhibitor (Perbandt *et al.*, 2015).

Other studies also showed that *Pf*GST is able to interact and bind to large lipophilic molecules including bilirubin, hemin, and certain drugs (Axarli et al., 2004; Deponte and Becker, 2005; Harwaldt et al., 2002). A study done by Hiller *et al*. (2006) revealed that *Pf*GST was able to bind FP in an uncompetitive mode. It was also shown that the FP had a preference in binding to the *Pf*GST in the presence of GSH (*Pf*GST-GSH) complex (Hiller *et al*., 2006). In addition, it was shown that the artemisinin drug that is usually employed against drug resistant strains of *Pf*GST (Maitland et al., 2004) is not a substrate of *Pf*GST and does not turn over artemisinin in sufficient biologically quantities (Hiller *et al*., 2006). It was also reported that in addition to artemisinin, other malarial drugs such as primaquine and chloroquine would not lead to inhibition of the *Pf*GST enzyme in vivo (Hiller *et al*., 2006).

These current anti-malarial drugs target the parasite primarily in the blood stage (asexual stage of the life cycle), which leads to the symptoms of the disease (Delves *et al*., 2012). In Figure 2.8, adapted from plasmoDB (Le Roch *et al.,* 2004), different percentiles of glyceraldehyde-3-phosphate dehydrogenase (GAPDH), a housekeeping gene, and *Pf*GST in different life cycle stages were investigated to show the importance of the targeted enzyme *Pf*GST, which functions in the blood stage. The various stages in the life cycle analysed included the ring stage, early and late trophozoite stage, schizont stage, gametocyte II and V stage, and ookinete stage. GAPDH, a housekeeping enzyme, is used for comparison. Using the bar graphs from Figure 2.8, the transcriptome analysis showed that *Pf*GST is found at comparable levels to GAPDH, the known housekeeping enzyme. This suggests that *Pf*GST is an important enzyme during the blood stage; especially the ring stage, early trophozoite stage, and gametocyte II stage. The *Pf*GST is a better target for the development of a drug in comparison to GAPDH as the *P. falciparum* GAPDH has a sequence identity of 63.5% to the human erythrocytic GAPDH (Daubenberger *et al*, 2000). The *Pf*GST isoform only has up to 37% identity with other GST classes, therefore, suggesting that *Pf*GST is a better target for drug development as the drug can be targeted to the particular protein, and will not bind similar proteins (Fritz-Wolf *et al.*, 2003).

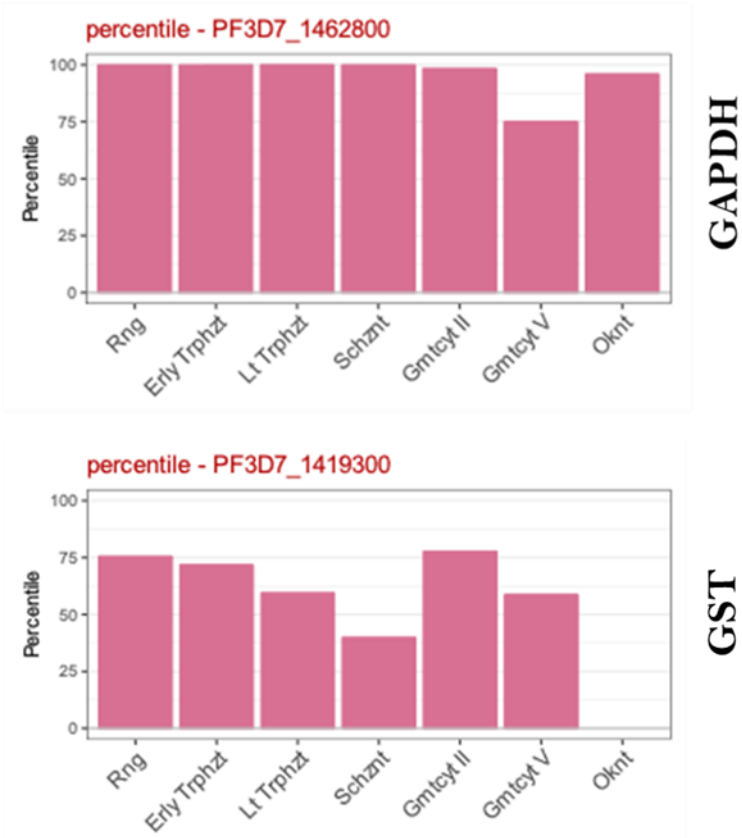


Figure 2.8: Transcriptome analysis of GST and GADPH in Plasmodium falciparum. *A* Bar Graph showing the percentile of two different enzymes; GAPDH and GST in several stages of the malaria life cycle. GAPDH (Gene ID: PF3D7_1462800) is a housekeeping gene that is used as a comparison, as it is highly expressed at all times in the life cycle. *Pf*GST (Gene ID: PF3D7_1419300) is the enzyme that is under investigation in this study as a possible target for the development of a new therapeutic. This figure was adapted from plasmaDB (Le Roch *et al.*, 2004)

2.8 Bromosulphophthalein

In this study, bromosulphophthalein (BSP) was investigated as a possible inhibitor of *Pf*GST. The Food and Drug Administration (FDA) approved compound, is a phthalein dye (IUPAC: disodium;2-hydroxy-5-[(4-oxo-3-sulfonatocyclohexa-2,5-dien-1-ylidene)-(2,3,4,5-tetrabromo-6-carboxyphenyl)methyl]benzenesulfonate) that is used in liver function testing which is based on the rate of the removal of the dye, which then allows for the measure of the liver function (Rosenthal and White, 1925). It has a molecular weight of 838 g/mol with a hydrogen bond donor and acceptor count of 2 and 10 respectively (https://pubchem.ncbi.nlm.nih.gov). In addition, there are 8 rotatable bonds that it can form. Figure 2.9 below is a two-dimensional (2D) representation of BSP. Bradley *et al.* (1945) where the first to use BSP to estimate liver blood flow in humans. Since then, BSP is used in a range of studies in both humans and animals. Some studies include understanding liver failure and management (Steiner *et al.,* 2004), nonradioactive uptake assay for rapid analysis of thyroid hormone transport function (Jayarama-Naidu *et al.,* 2015), measuring the liver blood flow resulting from elevated feed intake in lactating dairy cows (Sangsritavong *et al.,* 2002), hepatic BSP and glucose uptake were measured in hepatic venous samples in a study of hepatic arterial perfusion that decreases intrahepatic shunting and maintains glucose uptake in the rat liver (Alexander *et al.,* 2002).

It has also been shown by Kolobe *et al.* (2004), as a non-substrate anionic ligand that interacts and inhibits the human class Alpha GST with two type 1 subunits (hGSTA1-1). Using Isothermal titration calorimetry techniques, along with inhibition kinetics, it was shown that there are 2 sets of independent binding sites of which one has a strong affinity for BSP (k_d = 0.12 µM) and the other has a lower affinity for BSP (k_d = 9.1 µM). The binding of the BSP was both enthalpically and entropically

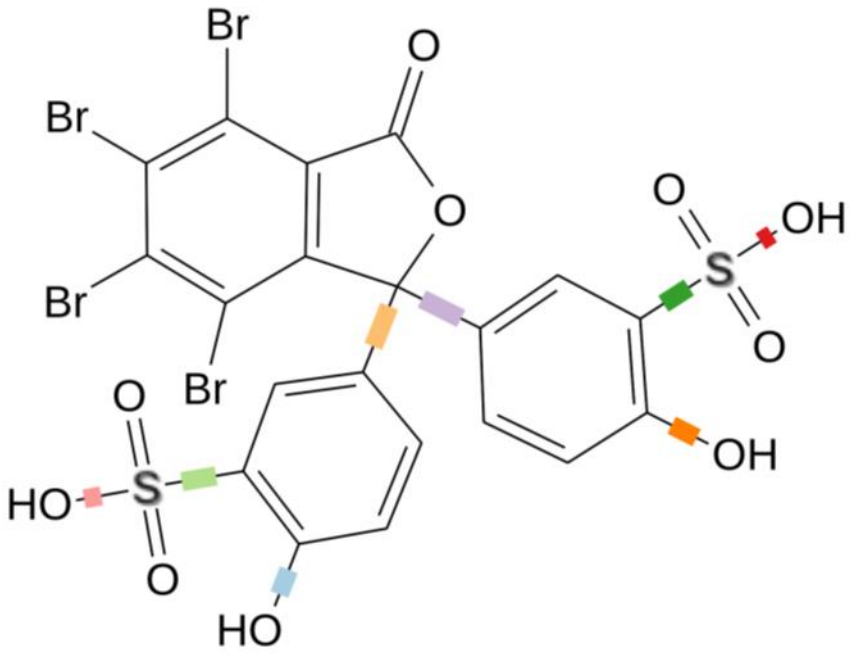


Figure 2.9: The Chemical structure of BSP. The 8 rotatable bonds are shown with the different coloured squares. BSP has 2 hydrogen bond donors and 10 hydrogen bond acceptors. This image was constructed using Desmond™ (Schrödinger Release 2017-3).

favourable. The inhibition kinetics study showed that binding of BSP to the high affinity site did not inhibit the enzymes activity. However, binding to the low affinity site caused inhibition of the enzyme's activity non-competitively, which is suggested as the active site (H-site) (Kolobe *et al.*, 2004). Therefore, due to the binding of subclass Alpha of the GST family, this study investigated the BSP's potential inhibition capacity of *Pf*GST, leading to the adaption of the BSP molecular, which targets *Pf*GST alone. BSP could then provide a scaffold for future novel inhibitors, which are targeted to only *Pf*GST.

2.9 Computational Chemistry

Molecular modelling uses algorithms based on Newtonian laws to simulate chemical events that determine and calculate the chemical properties of atoms and molecules (Palermo and De Vivo, 2014). Designing of potentially new drugs and events such as the binding of a drug to its target are all determined through diverse computational chemistry approaches, that it enables the prediction and calculation of these events (Palermo and De Vivo, 2014). Why this is a favourable avenue in drug design and discovery is that the production of new drug candidates occurs faster and at a lower cost, therefore, increasing the effectiveness of the overall drug design process (Jorgensen, 2004). Furthermore, computational chemistry is based on the essentials of molecular mechanics (MM) (Young, 2004).

Semi-empirical treatments, based on the Hartree-Fock formalism, however, making estimates and acquiring parameters from empirical data, cannot always be used effectively when the molecule is too large. Therefore, avoiding quantum mechanics (QM), the behaviour of the molecule can still be modeled through MM. MM constructs a simple algebraic expression, specifically the harmonic oscillator equation, for the determination of the total energy of a compound, which does not include the computing of a wave function or total electron density. This energy expression describes the energy associated with bond stretching, bending, rotation and intermolecular forces, specifically Van der Waals interactions and hydrogen bonding (Young, 2004). This equation can be seen below:

$$E_{total} = \sum bonds\ E_{stretch} + \sum angles\ E_{bend} + \sum dihedral\ E_{torsion} + \sum pairs\ E_{non-bonded} \quad (1)$$

It must be noted that all the constants in the above equations must be obtained from experimental data or an *ab initio* calculation. The advantages of MM include modelling of huge molecules like proteins and segments of DNA which are a necessity for computational biochemists as a primary tool (Young, 2004)

Another difference that can be seen between the QM and MM is that MM parameterise the method by constructing parameters and functions known as a force field. QM does not require the parameters as it describes the interactions through nuclei and electrons interacting with one another (Young, 2004). Force fields, which include AMBER (Assisted Model Building with Energy Refinement), CHARMm (Chemistry at HARvard Macromolecular Mechanics), GROMOS (GROningen Molecular Simulation) and OPLS (Optimised Potentials for Liquid Simulations), are functions which enables the prediction of the energy that is associated with the co-ordinates of atoms in the system (Hu *et al.*, 2003). Experimental data including X-ray diffraction, NMR and neutron spectroscopy are all used in the parameters for the force fields (González, 2011). Therefore, the following equation uses the experimental data to determine the total energy of the system:

$$E_{total} = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K\,\theta(\theta - \theta_{eq})^2 + \sum_{adihedrals} \frac{V_n}{2} [1 + \cos n\emptyset - \gamma] + \sum^{LJ} i \sum i < j\ 4\varepsilon ij [(\frac{\sigma_{ij}}{r_{ij}})^{12} - (\frac{\sigma_{ij}}{r_{ij}})^{6}] + \sum_{i}^{Elec} \sum_{i<j} \frac{qiqj}{4\pi\varepsilon r_{ij}} \quad (2)$$

Chapter 3

Expression, purification, and quantification of *Plasmodium falciparum* glutathione *S*-transferase

3.1 Introduction

To overexpress, purify and characterise a recombinant protein, an expression vector needs to be constructed and inserted into a host, usually *E. coli*, to produce the required protein. Vectors generally are comprised of the origin of replication, a selectable marker, restriction sites and multiple cloning site where the gene of interest is inserted (Wagner *et al.*, 2007). Following this, the vector is then introduced into a host through transformation and once housed in the bacterial host, the protein is overexpressed. To characterise the protein both structurally and functionally, protein purification must be completed to separate the recombinant protein from other bacterial proteins, therefore, producing pure recombinant protein. Once purification has taken place, both the quality and the quantity of the pure recombinant protein needs to be assessed to ensure that the recombinant protein is devoid of (i) nucleic acid contamination (ii) aggregation and (iii) other protein contaminants. Thus, once the protein of interest, *Pf*GST, was successfully overexpressed and purified, a UV-Vis absorbance spectrum and sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis of the protein sample was done to assess the purity and quality of the protein sample.

The *Pf*GST gene was inserted into a pET11 vector (pET11-*Pf*GST), with the addition of a his-tag. The T7 Express™ *E. coli* cells were then transformed with the pET11-*Pf*GST vector. The protein was successfully expressed and purified using Immobilised Metal Affinity Chromatography (IMAC). Quantification of the recombinant *Pf*GST was then performed in order to determine the purity and quantity of P*f*GST, using SDS-PAGE analysis and UV-Vis spectrometry.

3.2 Materials

Table: 3.1 Non-standard reagents and material, used in the different experimental methods, with their appropriate manufacturers.

Experimental Method	Material	Manufacturer
Transformation and Expression	pET-lla plasmid	GenScript, USA
	Ampicillin	Melford, UK
	T7 Express™ cells	New England Biolab, USA
	Isopropyl βD-1-thiogalactopyranoside	Melford, UK
Purification	HisTrap column	GE healthcare, Sweden
Glycine SDS-PAGE	BLUeye Prestained Protein Ladder	
	A Bio-Rad Mini Protein II® gel-casting unit	Bio-Rad

Therefore, the complementary DNA (cDNA), which encodes the *Pf*GST (UniProt code: Q8MU52), was extracted from GenBank (https://www.ncbi.nlm.nih.gov/genbank/). The messenger RNA (mRNA) sequence of the enzyme was re-coded so that it was compatible with the natural codon usage frequency in *E. coli*. Furthermore, a non-cleavable 6×his-tag was added to a site, upstream of the open reading frame (ORF). The gene was synthesised and cloned into a pET-11a vector (the pET11-*Pf*GST), which allows for the selection with ampicillin and has a T7 polymerase promoter for expression control. The vector also has an incorporated Ndel and BamHI sites (Figure 3.1) for the targeted gene insertion.

Sanger sequencing was previously done by Inqaba Biotec (Pretoria, South Africa) and it was confirmed that the presence of *Pf*GST ORF. Figure 3.1 depicts the pET11-*Pf*GST vector construct and the *Pf*GST translatable region, with a non-cleavable 6×His tag attached to the N-terminus.

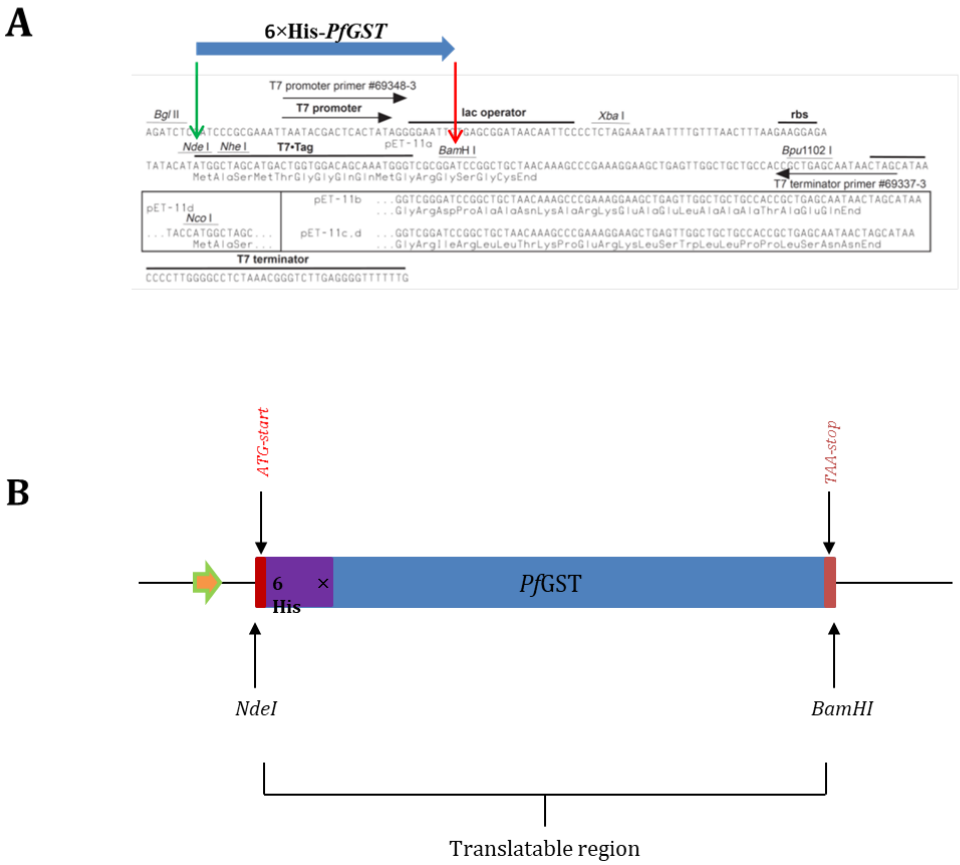


Figure 3.1: A schematic displaying the vector construct, the *Pf*GST FASTA sequence and its translatable region. (A) Illustration of the pET-11 vector construct. (B) The *Pf*GST translatable region with a non-cleavable 6×His tag, attached to the N-terminus. Ndel and BamHI sites on the N-terminus and the C-terminus were incorporated respectively, to clone the target gene into the vector.

3.3 Methods

3.3.1 Transformation and Expression of PfGST

The T7 Express™ cells and SOC media [2% (w/v) tryptone, 0.5% (w/v) yeast, 10 mM, NaCl, 2.5 mM KCl, 10 mM $MgCl_2$ and 20 mM glucose] were defrosted on ice and thereafter 1 µL of the pET11-*Pf*GST vector (0.04 ng/mL) was added to the T7 Express™ cells and incubated on ice for 30 minutes. Thereafter, the cells were heat shocked at 42°C for 45 seconds and the transformed cells were then added to the SOC media and incubated at 37°C for 1 hour (230 rpm). After incubation, 200 µL of the cells were plated on an agar plate, which contained 50 µg/mL of ampicillin. The plate was sealed and grown for 12-16 hours at 37°C. A colony was selected and cultured for 16 hours in 2 × yeast tryptone (2YT) [1% yeast (w/v), 1.6% tryptone (w/v), 0.5% NaCl (w/v)] at 30°C, 230 rpm with 50 µg/mL of Ampicillin. The glycerol stock was then prepared in a 1:1 ratio using 1 mL of the overnight. The sample was centrifuged (5000×*g*, 10 minutes, 10°C) and the pellet was re-suspended in 250 µL of supernatant. Following this, 250 µL of 80% (v/v) sterile glycerol was added and the stocks were frozen at -80°C.

The glycerol stock of *E. coli* T7 Express™ containing the *Pf*GST was cultured overnight (30°C, 230 rpm with 50 µg/mL of Ampicillin) in 2YT. The 500 mL cultures were diluted 1/50 with fresh 2YT and incubated (37°C, 230 rpm) until the OD_{600} ~0.6. The culture was subsequently induced with IPTG (0.5 mM) and incubated (8 h, 250 rpm, 30°C). Using centrifugation (5000×*g*, 15 minutes, 10°C), the cells were harvested and resuspended in resuspension buffer [50 mM Sodium phosphate buffer with 0.01% (w/v) NaN_3; pH 8.0] and frozen (-80 ºC).

3.3.2 Purification of PfGST

In this study, IMAC was used in order to isolate the recombinant protein sample. IMAC is a commonly used method for fast purification of poly-histidine affinity-tagged proteins in a single step (Porath and Olin, 1983). Agarose beads or magnetic particles are functionalised with chelating groups which allows the immobilisation of a chosen metal ion; therefore, functioning as ligands for the binding and purification of a specific protein (Porath and Olin, 1983). Two common chelating groups are

nitrilotriacetic acid (NTA) or iminodiacetic acid (IDA), which chelate divalent metal ions such as Ni^{2+}, Co^{2+}, Cu^{2+}, or Fe^{2+}.

The frozen cell suspension was thawed at room temperature (25°C) and disrupted by sonification on ice for five rounds of 20 second pulses at high intensity (Qsonica Sonicators, USA). Centrifugation (18000×*g*, 20 min, 5°C) was performed to separate the insoluble cell fraction from the soluble cell fraction. The soluble cell fraction was subsequently used in the purification process. IMAC, using a 5 mL (1.6 × 2.5 cm) HisTrap® column charged with Ni^{2+}, was equilibrated (5 mL/min) with 20 column volumes of equilibration buffer [50 mM Sodium phosphate buffer with 0.01% (w/v) NaN_3; pH 7.8], was used to purify *Pf*GST (ÄKTAprime plus, GE Healthcare Life, USA). The soluble fraction (supernatant) was injected onto the column and allowed to move through the column with the flow through being collected. Once the flow through was collected, the column was washed using 20 column volumes of the equilibration buffer, ensuring the non-specifically bound protein was washed off. The bound protein was eluted using a one-step elution [50 mM Sodium phosphate buffer, 500 mM imidazole, 0.01% NaN_3; pH 8.0]. The purification profile is shown in Figure 3.2 A. The samples of the pellet, supernatant, flow through and protein fractions were resolved on a glycine SDS-PAGE gel.

3.3.3 Quantification of PfGST

It is important to assess the quality as well as the quantity of a recombinant protein, following purification. This enables us to determine if the recombinant protein is devoid of (i) nucleic acid contamination (ii) aggregation and (iii) other protein contaminants. This is a critical step in any recombinant protein production especially if the protein of interest is known to be associated with nucleic acids and/or be involved in protein-protein interactions. Thus, a UV-Vis absorbance spectrum (260-340 nm) and SDS-PAGE analysis of the protein sample was done to assess the purity of the protein sample.

A Laemmli (Laemmli, 1970) SDS-PAGE glycine gel was set up [4% (w/v) stacking and 10% (w/v) separating] to determine if the recombinant *Pf*GST *was* overexpressed and purified successfully. This electrophoretic technique is used to analyse proteins through a polyacrylamide matrix, which allows for the separation of the proteins as

they move through the matrix. Proteins are separated based on their molecular weight, shape, and charge. However, SDS (an ionic surfactant), is used to denature the protein sample (forming a uniform rod shape) by disrupting non-covalent bonds and to mask their natural charge, based on their amino acid composition. Therefore, the protein samples will separate according to their molecular weight. During the separation process, smaller molecules migrate a lot further then the larger molecule (Maurer, 1971). Therefore, using this technique both the protein purity and the molecular mass of the protein were determined (Wilson and Walker, 2000).

The collected samples (50 µL) were mixed with reducing sample buffer (50 µL) [125 mM Tris-HCl, 4% (w/v) SDS, 10% (v/v) β-mercaptoethanol, 20% (v/v) glycerol; pH 6.8] and boiled (100°C) to ensure complete denaturation of the protein. The samples were then centrifuged (15000×*g*) to ensure the samples were mixed thoroughly and that the contents of the Eppendorf tubes were at the bottom. The samples (10 µL) were separated on a 12,5% (w/v) discontinuous gel consisting of a 3% (w/v) stacking gel [500 mM Tris-HCl, pH 6.8] and 10 % (w/v) separating gel [1.5 M Tris-HCl, pH 8.8], using a Bio-Rad Mini Protean ® electrophoresis system. A Bio-Rad Mini Protein II® gel-casting unit was assembled according to the manufacturers' instructions and comprised of a gel-casting stand with casting gaskets, a ten-well gel comb (1.5 mm thick), 81.5 mm × 101.5 mm and 72.5 mm × 101.5 mm glass plates. Once the samples were loaded and the tank buffer [250 mM Tris-HCl, 192 mM glycine, 0.1% (w/v) SDS, pH 8.3] was added, the gel was run at 120 V until the dye front was 0.1 cm from the bottom. The gel was removed from the unit, stained with Coomassie Blue R 250 staining solution [0.1% (w/v) Coomassie dye in 1:5:4 (v/v/v) acetic acid-methanol-water solution] and destained [1:5:4 (v/v/v) acetic acid-methanol-water solution]. An image of the gel was captured (Figure 3.2 B) and the molecular weight of the recombinant *Pf*GST protein band was determined using the graph of Log molecular weight of standards against the distance migrated (mm) by the protein on the gel (Figure 3.2 C).

The pure *Pf*GST protein fractions were pooled and centrifuged (10 000×*g*, 10 min) to remove any aggregates. A UV-Vis absorbance spectrum (250 to 340 nm), using a JASCO V-630 spectrophotometer, was done to check that the protein sample is devoid of nucleic acid contaminations and aggregation (Figure 3.3). This is

characterised by a trough at 260 nm, a peak at 280 nm and zero absorbance at 340 nm, respectively. A 260/280 ratio was used to detect if there was any nucleic acid contamination. A ratio of ~ 0.57 is indicative of a near 100% pure protein sample. On the other hand, a 260/280 ratio of 1 and above indicates a decreased purity of protein sample due to nucleic acid contamination (Glasel, 1995).

The concentration of *Pf*GST was determined spectroscopically at 280 nm through the application of Beer-Lambert law:

$$A = \varepsilon c l \tag{3}$$

where A is the absorbance of the sample, ε is the molar extinction coefficient ($M^{-1}cm^{-1}$), c is the concentration of the sample in solution (M) and l is the path length of the light passing through the solution (cm). The ε for the *Pf*GST is described using the formulae by Perkins (1986), which works by predicting the contributions of the residues that absorb at 280 nm. These residues are namely tryptophan, tyrosine, and cysteine residues to the molar extinction coefficient of a protein:

$$\varepsilon = 5550\ \Sigma\ Trp + 1340\ \Sigma\ Tyr + 150\ \Sigma\ Cys \tag{4}$$

Based on the *Pf*GST sequence, the theoretical ε, calculated by the ProtParam tool (https://web.expasy.org/protparam/), implemented in Expasy (Gasteiger *et al.*, 2005) was 29 340 $M^{-1}cm^{-1}$. The theoretical ε was used to estimate the *Pf*GST concentration.

To determine the proteins concentration, 3 absorbance readings were taken at 260 nm, 280 nm, and 340 nm for a doubling dilution series of the recombinant protein. In order to account for any aggregates, the absorbance reading at 340 nm was subtracted from absorbance reading at 280 nm. These absorbance readings were then plotted against the dilutions and the concentration was determined (Supplementary Figure 1).

3.4 Results

3.4.1 Expression and Purification of PfGST

The supernatant, after sonification and centrifugation of the soluble cell fraction, was subsequently used in the purification process (described in Section 3.3.2). The *Pf*GST was eluted in a one-step process (Figure 3.2 A) and 2 mL fractions were collected. The first peak represents the flow through. This occurs when the supernatant sample is loaded and injected onto the column, causing an increase in the absorbance reading at 280 nm. The column was then washed using the equilibration buffer, which caused a decline in absorbance. Following the wash step, the protein of interest was eluted, causing an increase in the absorbance and the formation of the second peak.

A 12.5% (w/v) glycine SDS-PAGE gel was set up using samples of the pellet, supernatant, flow through and eluted protein fractions, as described in Section 3.3.3 (Figure 3.2 B). This technique was used to determine the size and purity of the *Pf*GST. A standard curve was constructed using the BLUeye Prestained Protein ladder standards to determine the approximate molecular weight of the purified *Pf*GST sample (Figure 3.2 C). The approximate molecular weight of the protein is 24.3 kDa and is considered electrophoretically pure with a single band (Figure 3.2 B, Lane 5-10). The dimer and tetramer can also be observed on the SDS-PAGE gel.

3.4.2 Quantification of PfGST

A UV-Vis absorbance spectrum (250 to 340 nm) of the *Pf*GST, using a JASCO V-630 spectrophotometer, was done to check that the protein sample is devoid of nucleic acid contaminations and aggregation (Figure 3.3). There was a characteristic peak of 2.58 A.U. at 280 nm and a flat region with an absorbance reading of 0.035 A.U. at 340 nm, which indicative of a protein sample that is devoid of aggregates. The 260/280 ratio of 0.9 falls between 0.57 and 1, suggesting that the protein sample is considered pure of nucleic acid contamination.

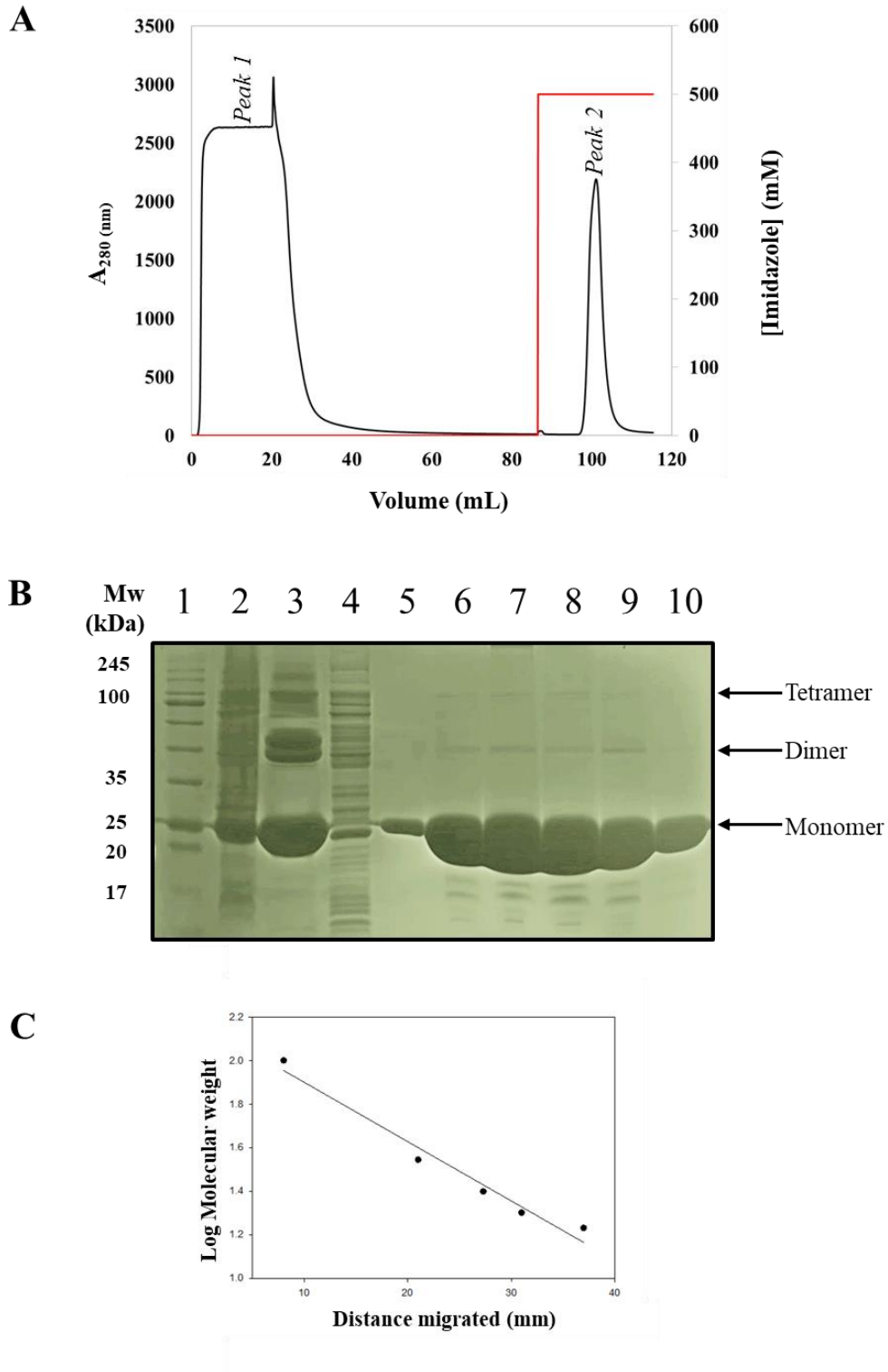


Figure 3.2: Expression and purification analysis of *Pf*GST. (A) The IMAC elution profile displays the absorbance at 280 nm of the effluent. The first peak represents the flow through. The second peak is the eluted *Pf*GST, using a 50 mM Sodium phosphate monobasic buffer, containing 500 mM imidazole and 0.01% NaN_3, at a pH 8.0 (one-step elution shown with the red line). (B) A 12.5% glycine SDS-PAGE gel analysis showing the samples collected from the purification process. Lane 1: Molecular weight marker, Lane 2: Pellet, Lane 3: Supernatant, Lane 4: Flow through, Lane 5-10: Elution fractions. (C) The standard curve constructed using the SDS-Page gel, and fitted to a straight line, $y = -0.272x + 2.17$, with an R^2 value of 0.97. The molecular weight of *Pf*GST was determined as 24.28 kDa.

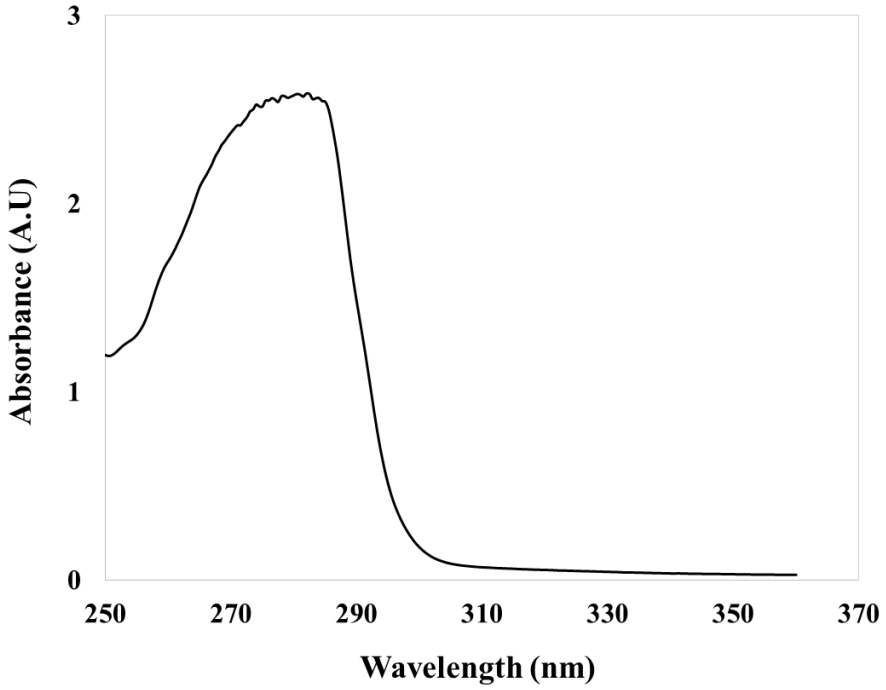


Figure 3.3: Absorbance Spectrum of *Pf*GST. A UV spectrum (250 to 340 nm) of the *Pf*GST shows a characteristic peak at 280 nm suggestion the sample is protein. No peak observed at 260 and 340 nm suggest that there is little to no nucleic acid contamination and aggregates, respectively.

3.5 Discussion

In order to express a target protein for characterisation, both structural and functional, the target sequence must be inserted into a vector. The pET11-*Pf*GST vector was constructed successfully, using incorporated Ndel and BamHI sites and transformed into T7 Express^{TM} cells. The transformation process allowed for the vector to be inserted into the host where the targeted *Pf*GST gene was overexpressed. The protein was overexpressed in the soluble fraction (Figure 3.2 B, Lane 3). Once the protein was overexpressed successfully, purification was done using a HisTrap column, charged with Ni^{2+}. The protein has a non-cleavable 6×His-tag on the N-terminus which allowed for the binding to a Ni^{2+} column whilst the contaminants were eluted. Following this, the protein was eluted through a one-step elution, with minimal contamination observed on the SDS-PAGE gel (Figure 3.2 B, Lane 5-10). Using the standard curve and SDS-PAGE gel, the monomeric size of the protein was determined as 24.3 kDa, which corresponded with literature values of 24.8 (Harwaldt *et al.*, 2002). The wet biomass was 11.85 g from 1 L of the cultured media and the protein did not require concentrating due to the high yields that ranged between 20-40 mg/mg in 12 mL fractions. The protein was also considered stable as there was little to no aggregation at high concentrations. The UV-Vis spectrum of the protein indicated that the protein sample was pure. The lack of nucleic acids was confirmed using the 260/280 ratio. The ratio of 0.9 falls between 0.57 and 1, which suggests that the protein sample is pure of nucleic acid contamination. Therefore, the protein was considered pure and of good quality for further experiments.

Chapter 4

Catalytic and ligandin activity of *Plasmodium falciparum* glutathione *S*-transferase in conjunction with X-ray crystallography

4.1 Introduction

The activity of protein can be monitored through functional characterisation techniques which may include enzyme kinetics assays, the half maximal inhibitory concentration (IC_{50}), extrinsic fluorescence spectroscopy and isothermal titration calorimetry. These techniques use known inhibitors or substrates to determine the functionality of the protein by determining the activity and kinetic parameters of the protein, whether or not the activity of the enzyme is affected using a non-substrate ligand BSP and how its binding affects other substrate bound molecules such as 1-anilinonaphthalene-8-sulfonate (ANS). The thermodynamic parameters of the *Pf*GST and BSP binding event will also be determined. In addition, crystallography studies were performed to identify the location at which BSP is binding to *Pf*GST.

4.2 Materials

Table 4.1: Non-standard reagents and materials, used in the different experimental methods.

Experimental method	Material	Manufacturer
Specific activity assay, kinetics assay and IC_{50}	1-chloro-2,4-dinitrobenzene (CDNB)	Sigma Aldrich
Extrinsic ANS Fluorescence spectroscopy	ANS	Sigma Aldrich
Crystallography	The Index™ crystallization screening kit (HR2-144)	Hampton Research
	Crystal Clear Tape (HR4-506)	Hampton Research
	Paratone	Hampton Research
	CrystalCap™Spine HT 0.05-0.1mm CryoLoop	Hampton Research
	Dino-Eye eyepiece camera	Dino-Lite Digital Microscope
ITC	Dimethyl sulfoxide (DMSO)	Sigma Aldrich
	Tris(2-carboxyethyl) phosphine (TCEP)	Inqaba

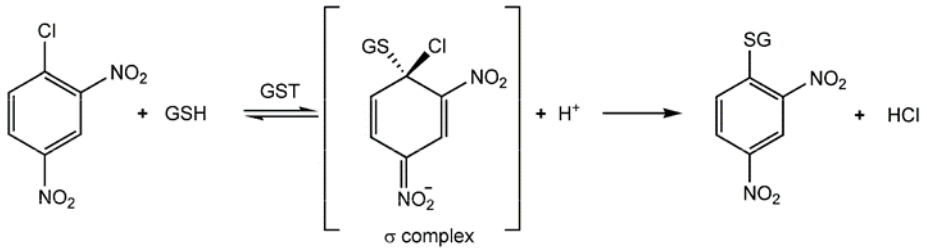
Note: for all above experiments, a 310 mM BSP stock was used and diluted accordingly for the different experiments.

4.3 Methods

4.3.1 Functional characterisation

4.3.1.1 Specific activity of PfGST, in the presence and absence of BSP

Enzyme assays are methods that allow for the measurement of enzymatic activity, which is an index of biological activity for a given protein with catalytic function. Therefore, in order to monitor the enzymatic activity of *Pf*GST in the presence and absence of BSP, a GSH-CDNB conjugation enzyme activity assay (Habig *et al.*, 1974), was performed whereby the specific activity of *Pf*GST was determined by monitoring the formation of 1-(*S*-glutathionyl)-2,4-dinitrobenzene (ε_{340} = 9600 $M^{-1}.cm^{-1}$) at 340 nm using spectrophotometry (Jasco V-600 spectrometer) with a steady state kinetics method (Equation 5). Various samples of *Pf*GST [0-864 nM] were prepared in the assay buffer [100 mM sodium phosphate, 1 mM NaEDTA, 0.1% (w/v) NaN_3, 5 mM GSH; pH 6.5] and incubated for an hour (37 °C). A 30 mM stock solution of CDNB was prepared in 99% ethanol (v/v) and diluted to a final concentration of 1mM CDNB in a total assay volume of 3 mL in order to initiate the conjugation reaction with GSH. The reaction of GST is shown below:



(5)

Linear progress curves were obtained for the various concentrations of *Pf*GST by measuring the absorbance, measured at 340 nm, over a period of 1 minute (37°C). This was done in triplicate and the data from each reaction were corrected for by using the sample that contained no enzyme. The initial reaction velocity of the complex formed ($\mu mol.min^{-1}$) was determined using the Beer-Lambert equation (see Section 3.3.3) with the extinction coefficient of CDNB at 340 nm (ε_{340} = 9600 $M^{-1}.cm^{-1}$). The initial velocity was plotted as a function of the mass of enzyme (mg) to produce a straight-line which was fitted to a linear equation, using Sigma Plot 12.0. The gradient of the straight-line (1st order kinetics) was used to determine the specific activity of the enzyme ($\mu mol.min^{-1}.mg^{-1}$). This experiment was repeated with the

addition of 50 µM BSP (25°C). This was to determine whether BSP was able to inhibit the activity of *Pf*GST.

Following this, the Michaelis constant (K_M) of *Pf*GST was determined using the initial velocity at varied concentrations [0-50 mM] of GSH (20°C). This was repeated using varied concentrations [0-50 mM] of CDNB. Other conditions including pH, temperature and enzyme concentration were kept constant. Samples were set up from a 50 mM GSH stock (3× dilution from original stock) with a fixed concentration of 1 µM of *Pf*GST in enzyme assay buffer. The kinetics assay followed the same set up as the enzyme assay stated in this section above. This was done in triplicates and the reactions were corrected using the no enzyme sample. The initial reaction velocity of the complex formed (µmol.min^{-1}) was plotted against the concentration of the GSH or CDNB (mM), which was then fitted using a rectangular hyperbola fit, using Sigma Plot 12.0, based on the Michaelis-Menten model, using Sigma Plot 12.0:

$$V_0 = \frac{V_{max}[S]}{K_M + [S]} \quad (6)$$

Where V_0 is the initial velocity, V_{max} is the maximum reaction velocity, [S] is the substrate concentration, K_M (Michaelis constant) is the substrate concentration when the reaction velocity is half of the Vmax.

4.3.1.2 IC_{50} estimation of BSP inhibition of PfGST.

The IC_{50} quantifies the amount of an inhibitor that is needed to inhibit half (50%) of a given biological process, or in this case the enzymatic activity of *Pf*GST (Neubig *et al.*, 2003). Once it was established that BSP does inhibit the activity of *Pf*GST (see Section 4.4.1), an IC_{50} of *Pf*GST was performed using varied concentrations of BSP (2× dilution from an original stock of 7.5 mM BSP), and the concentration of *Pf*GST was fixed at 3 µM in enzyme assay buffer. This was done in triplicate (37°C) and the reactions were corrected for by using the sample that contained no enzyme. A progress curve was constructed to determine the rate of the reaction (as described in Section 4.3.1) with the different concentrations of BSP. The percent (%) activity of the enzyme was determined using the protein sample with no BSP as the maximum activity (100% activity point). This was then used to divide the other point, which

were then converted to a percentage. The % activity was then plotted as a function of the log of the different concentrations of BSP, using Sigma Plot 12.0, to a non-linear 4 parameter logistic regression curve with equation:

$$f = min_y + \frac{\max_y - min_y}{1+(\frac{x}{EC_{50}})^{-Hillslope}} \quad (7)$$

where the bottom of the curve (of the % activity) is represented as min_y and the top of the curve (of the % activity) is represented as max_y, the half-maximal effective concentration is represented as EC_{50} and its equivalent definition in the case of inhibitions curves is the IC_{50}. The slope of the curve at the midpoint of the graph is represented as the Hillslope.

4.3.1.3 Extrinsic ANS fluorescence spectroscopy

Unlike intrinsic protein fluorescence, which uses the natural fluorescent properties of aromatic amino acids (Ghisaidoobe and Chung, 2014), extrinsic dyes can be covalently or non-covalently attached to proteins either through the ε-amino group of lysine, the α-amino group of the N-terminus or the thiol group of cysteine residue (Hawe *et al.*, 2008). In this study, ANS was used as an extrinsic probe in order to determine hydrophobic pockets (Hawe *et al.*, 2008) as it becomes highly fluorescent in apolar organic environments (Stryer, 1965). The extrinsic ANS fluorescence was monitored by initially incubating 5 μM of *Pf*GST with the addition of a 200 μM ANS solution in the enzyme assay buffer, in the dark (20°C). The concentration of ANS stock solution was determined using the Beer-Lambert equation (see Section 3.3.3) with the extinction coefficient of ANS at 350 nm (ε_{350} = 4950 $M^{-1}.cm^{-1}$) (Weber and Young, 1964). The samples were excited at 390 nm and the emission spectra were measured between 400 and 600 nm. The data was collected and averaged. In addition, the emission spectra for free ANS was subtracted from the corresponding ANS-protein spectra in the data analysis; thereby correcting for the contribution of unbound ANS. This experiment was repeated with the addition of 200 μM BSP. The samples were incubated for the same period (20°C). The data was collected and averaged. In addition, the emission spectra for free ANS, BSP and DMSO was subtracted from the corresponding ANS-protein-BSP spectra in the data analysis. Therefore, correcting

for the contribution of unbound ANS, BSP and DMSO. A quartz cuvette with a 10 mm path length was used for this experiment. The excitation and emission bandwith were 5 and 2,5 nm respectively with a scanning speed of 200 nm/min. In addition, UV-vis spectroscopy was used to evaluate the inner filter effect of BSP (Supplementary Figure 2). There was minimal absorbance between 400-600 nm (wavelength values of the emission spectra), therefore, there is no inner filter effect of BSP.

4.3.1.4 Thermodynamics of the PfGST-BSP interaction

The thermodynamic parameters for the interaction between *Pf*GST and BSP were determined using ITC. ITC is a quantitative technique that can determine the change in the enthalpy ($\Delta H°$), the binding stoichiometry (n), the change in entropy ($\Delta S°$) and the dissociation constant (K_d). ITC was performed using a Nano-ITC instrument (TA Instruments, Delaware, USA). The ITC machine consists of two cells that are enclosed in an adiabatic jacket (O'Brien *et al.*, 2002). The investigated compounds are placed in the sample cell, while the reference cell is used as a control and contains the buffer in which the sample is dissolved (O'Brien *et al.*, 2002). Therefore, ITC was used to determine the thermodynamic parameters for the interaction between *Pf*GST and BSP in the presence GSH. The ΔH°, n, ΔS° and the K_d were measured, whereas the ΔG° was estimated using the equation below:

$$\Delta G° = \Delta H° - T\Delta S° \quad (8)$$

where ΔG° is the change in Gibbs free energy, ΔH° is the change in enthalpy, T is the absolute temperature in kelvin and ΔS° is the change in entropy.

The protein sample was dialysed for 16 hours and centrifuged (10 000×*g*, 10 minutes) to remove any aggregates. The 7 mM BSP solution was prepared from a BSP stock (310 mM BSP in 100 % (v/v) DMSO), using the dialysate buffer. The amount of DMSO in the 7 mM BSP sample was calculated and added to the protein sample in order to avoid any buffer mismatch between the ligand and protein sample. Importantly, the amount of DMSO that was added was kept at 2% (v/v), therefore, avoiding perturbation of the protein structure and function. A 7 mM BSP solution was titrated against 50 µM *Pf*GST [50 mM Sodium phosphate monobasic, 150 mM NaCl,

3 µM TCEP, 0.01% (w/v) NaN_3, 5 mM GSH: pH 7.2]. A total of 300 µL of 50 µM protein sample was loaded into the sample cell and 250 µL of 7 mM BSP solution was loaded into the syringe. Once the system had equilibrated (stirring 75 RPM) at 25°C, computer-controlled titrations were done using 5 µL injections with a total of 30 injections over 6000 seconds of BSP solution into the sample cell, which contain the protein. The heat of dilution was accounted for by repeating the above experiment by titrating BSP into buffer (which did not contain protein). The experimental conditions were kept identical. Nano-ITC NanoAnalyze software was used to determine the thermodynamic parameters and a sequential two site model was used to fit the raw ITC data.

4.3.2 Structural characterisation

4.3.2.1 Crystallisation trials, X-ray diffraction and data collection

The crystallisation trials of the *apo Pf*GST conformation and *Pf*GST in complex with GSH and/or BSP were performed using the Hampton Research Index™ reagent kit (HR2-144). This kit is a 96 well buffer system that allows for the crystallisation of globular proteins and protein complexes. The protein was prepared at an initial concentration of 4.6 mg/ml [2 mM sodium, 0.01% (w/v) NaN_3, 15 mM NaCl, 5 mM GSH, 10 µM DTT; pH 7.8] with the addition of 50 µM BSP. Following this, a 96-well plate was prepared by filling each well 40 µL reservoir solution. The protein stock was mixed in a ratio of 1:2 (final drop of 0.75 µL), 2:1 (final drop of 0.75 µL) and 1:1 ratio (final drop of 1µL) with the Index™ screening kit, using the Oryx 8 robot (Douglas Instruments, United Kingdom). The plate was then sealed with Crystal Clear Tape and kept at 20°C and were screened over 2 weeks. The protein crystals conditions were recorded, and images were taken using a microscope mounted eyepiece camera (Dino-Eye eyepiece camera, Dino-Lite). The protein crystals were harvested using CrystalCap^{TM}Spine HT 0.05-0.1mm CryoLoop (using a Paratone cryoprotectant). Following this, the protein crystals were flash frozen in liquid nitrogen. The X-ray data was then collected using synchrotron (I03 beamline) at the Diamond Light Source (DLS), located at the Harwell Science and Innovation Campus in Oxfordshire. A beamsize of 60×50 µm was used with 0,100 s exposure. There were 2400 images captured with 100% transmission. The resolution at which the data was captured was 1.75 Å with a wavelength of 0.92 Å.

4.4 Results

4.4.1 Enzyme kinetics of PfGST in the presence and absence of BSP

The specific activity of *Pf*GST in the absence and presence of 50 µM BSP was determined using a standard GSH-CDNB conjugation assay (as described in Section 4.3.1, *Determination of the specific activity of PfGST in the presence and absence of BSP*) (Habig *et al.*, 1974).The initial reaction velocity of the complex formed (µmol.min^{-1}) was plotted vs the function of enzyme (mg) to produce a straight-line which was then fitted to a linear equation (Figure 4.1). The slope of the linear progress curve corresponded with the specific activity of the protein. The specific activity of *Pf*GST in the absence and presence of BSP was 0.84 µmol/min/mg and 0.34 µmol/min/mg, respectively. There was an overall 2.4 fold decrease in the enzyme's specific activity in the presence of BSP. Harwaldt *et al.* (2002) reported a lower *Pf*GST specific activity of 0.2 µmol/min/mg. Steady state kinetics was used in order to determine the K_M of *Pf*GST (as described in Section 4.3.1, *Determination of the specific activity of PfGST in the presence and absence of BSP*)). The K_M of *Pf*GST using varied concentrations of GSH was 5,41 mM (Figure 4.2). However, Harwaldt *et al.* (2002) reported a much lower K_M of 0,16 mM. The K_M of *Pf*GST with CDNB could not be determined as the reaction did not reach saturation.

4.4.2 IC_{50} for the interaction between PfGST and BSP

The IC_{50} of *Pf*GST was determined using varied concentrations of the non-substrate binding BSP with a fixed concentration of *Pf*GST (as described in Section 4.3.1, *IC_{50} estimation of BSP inhibition of PfGST.*). This curve was fitted to a non-linear 4 parameter logistic regression curve (Equation 6) and the IC_{50} was determined as 537 µM (Figure 4.3), however, Kolobe *et al*. (2004), reported a much lower IC_{50} value of 7µM in the case of BSP binding to human class Alpha GST with two type 1 subunits (hGSTA1-1).

4.4.3 Extrinsic fluorescence spectroscopy of PfGST in the presence and absence of BSP

Extrinsic ANS fluorescence was performed by incubating 200 µM ANS with 13.5 µM *Pf*GST in the presence and absence of 200 µM BSP (20°C). The samples were excited

at 390 nm and the emission spectra, between 400 and 600 nm, was collected and averaged (Figure 4.4). Free ANS shows a maximum emission at 516 nm. ANS bound to *Pf*GST in the absence of BSP shows a blue shift from 516 to 483 nm and a 2.5 fold increase in the fluorescence intensity, in comparison to the free ANS. ANS bound to *Pf*GST in the presence of BSP shows a blue shift from 516 to 483 nm and a 3-fold decrease in fluorescence intensity, in comparison to the free ANS. The spectral blue shifts are in accordance with other studies done by Stryer *et al*. (1965). Free ANS was in accordance with studies done by Möller and Denicola (2002). The decrease in the fluorescence intensity when the protein is in the presence of BSP can be explained by the study done by Mogensen *et al.* (2002). They showed that the decrease in fluorescent intensity can be a result of ANS displacement by a ligand.

4.4.4 Thermodynamics of PfGST-BSP interaction

The thermodynamic parameters of *Pf*GST binding to BSP were determined using ITC. Figure 4.5 shows the titration of *Pf*GST with BSP in the presence of GSH (25°C). A sequential two-site model was shown to result in the best fit for the raw data and suggested that the *Pf*GST has two types of co-operative binding sites for BSP which are both exergonic. This correlated with the findings reported by Kolobe *et al.* (2004), which showed that hGSTA1-1 comprised a high-affinity binding site and a low-affinity binding site for BSP. The first binding site had a low affinity for the BSP molecule. However, the second binding site had a much greater affinity for BSP. The first binding event was enthalpically favourable (exothermic) and entropically unfavourable. The second binding event was enthalpically unfavourable and entropically favourable. The stoichiometry was n=2.

4.4.5 Crystallisation and X-ray diffraction data collection of the PfGST-BSP complex

The crystallisation trials of the *apo Pf*GST conformation and *Pf*GST in complex with GSH and/or BSP were performed using the Hampton Research Index™ reagent kit (HR2-144). A beamsize of 60×50 µm was used, in order to diffract the crystal (Figure 4.6), with 0.100 sec exposure. A total of 2400 images were captured with 100% transmission. The resolution at which the data was captured was 1.75 Å with a wavelength of 0.92 Å. The crystal had a space group of P622, which is a unique space

group in which the *Pf*GST crystallised, in comparison to the published structures. The space groups which are most common is $P2_12_12$ and $P2_12_12_1$ (PDB codes: 10KT, 3FR6, 3FR9, 3FRC, 42XG). Another space group seen is C121 and $C222_1$ (PDB codes: 52KH and 3IGM). However, due to insufficient data collection with an overall completion of 89.9% (98.7% completion of the inner shell and 69.3% completion of the outer shell), the size of the crystal and the space group in which the protein had crystallised, the crystal structure could not be solved.

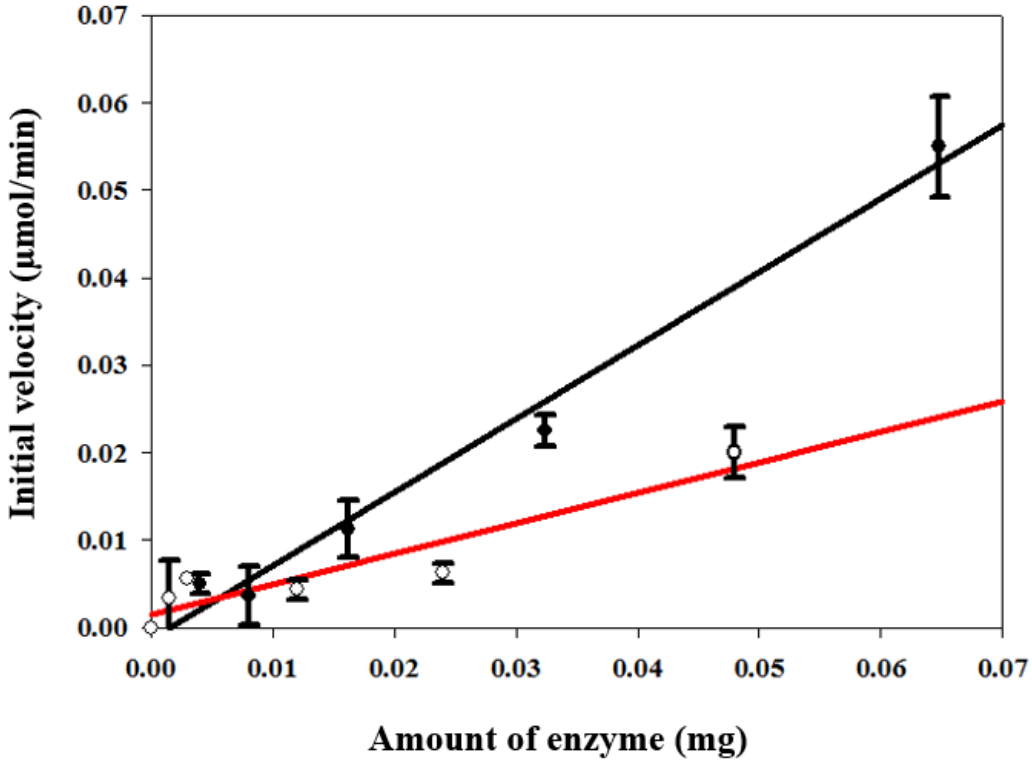


Figure 4.1: The specific activity of *Pf*GST in the presence and absence of BSP. The gradient of the straight-line fit indicated the specific activity. In the absence of BSP, the standard curve was fitted to a straight line, $y = 0.84x - 0.0013$, with an R^2 value of 0.99 (black). In the presence of BSP, standard curve was fitted to a straight line, $y = 0.35x + 0.0015$, with an R^2 value of 0.87 (red). This experiment was performed in triplicate and the error bars showed the standard deviation of the mean. All reactions were corrected using a non-enzymatic control. This figure was constructed in Sigma Plot 12.0.

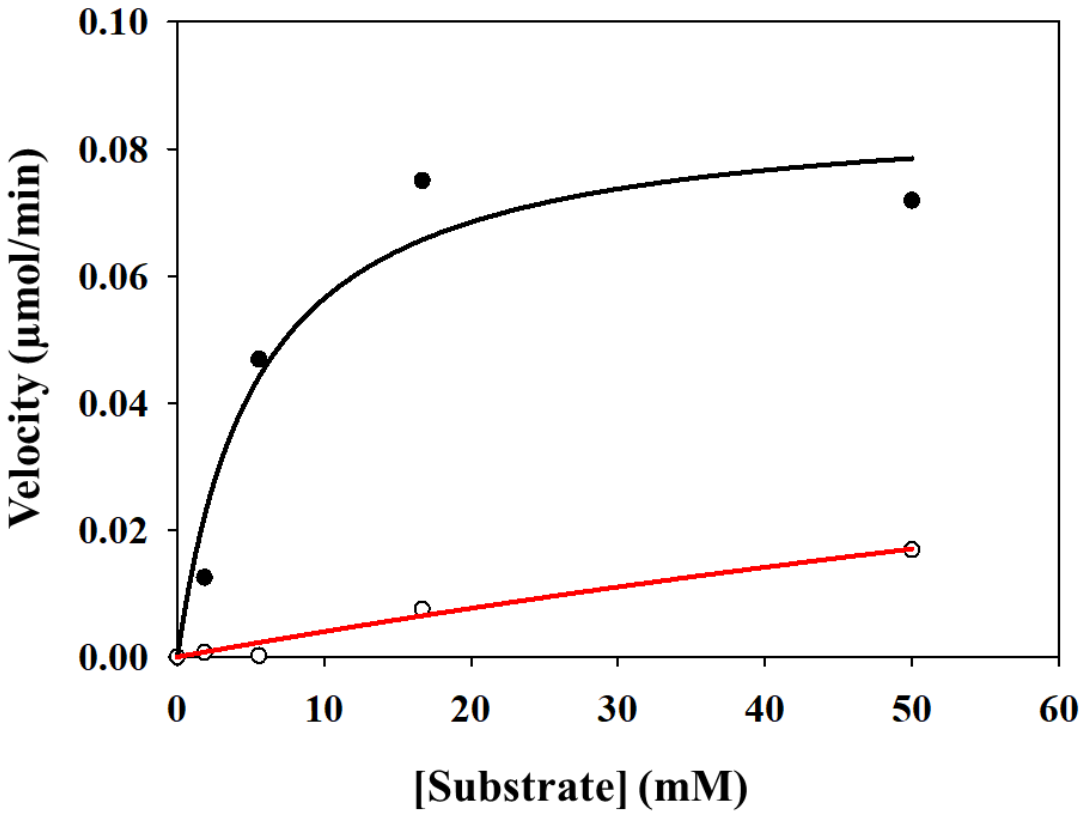


Figure 4.2: Steady State kinetics plots for *Pf*GST. A rectangular hyperbola, based on the Michaelis-Menton model, was fitted to the data of the enzyme activity as a function of the substrate concentration. The *Pf*GST was fixed at 1 µM. The equation for varied GSH was $V_0 = 0.087[S] / 5.4108 + [S]$ *with a* R^2 of 0.95. The equation of the varied CDNB could not be determined as the experiment did not reach saturation, therefore, this fit cannot be used. This experiment was performed in triplicates. All reactions were corrected using a non-enzymatic control. This figure was constructed in Sigma Plot 12.0.

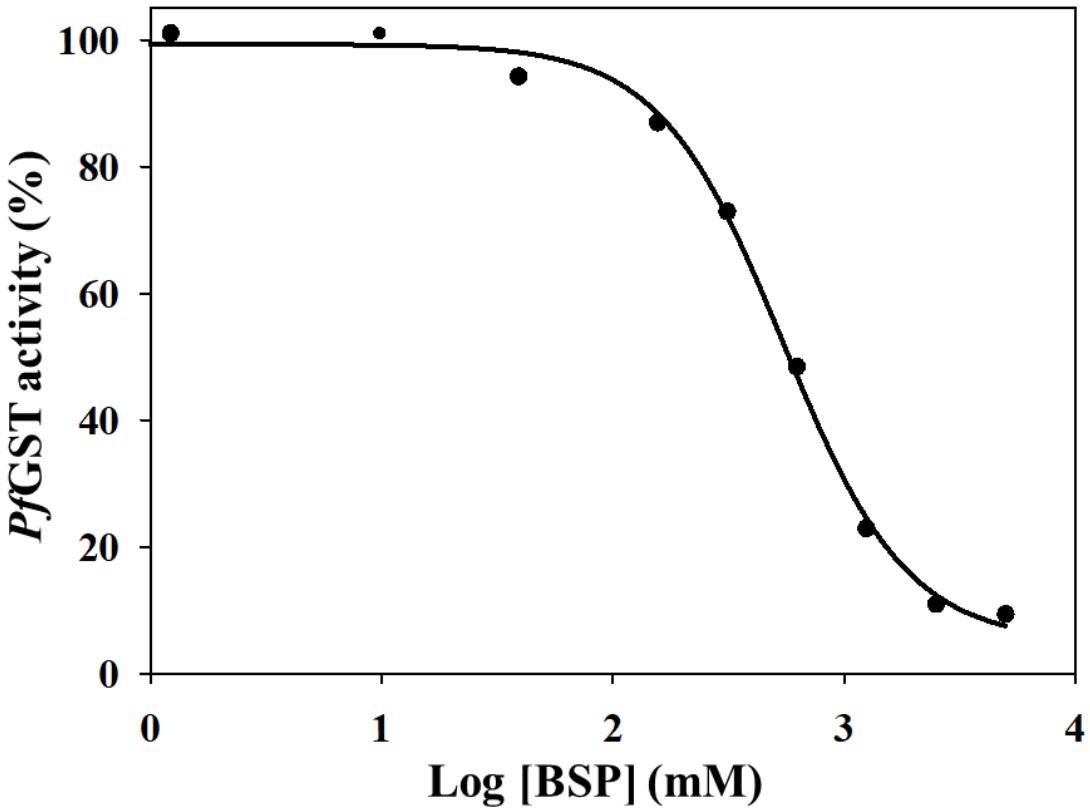


Figure 4.3: Determining the IC_{50} in the presence of BSP. A non-linear 4 parameter logistic regression curve was fitted to the data. This experiment was performed in triplicates. All reactions were corrected using a non-enzymatic. This figure was constructed in Sigma Plot 12.0.

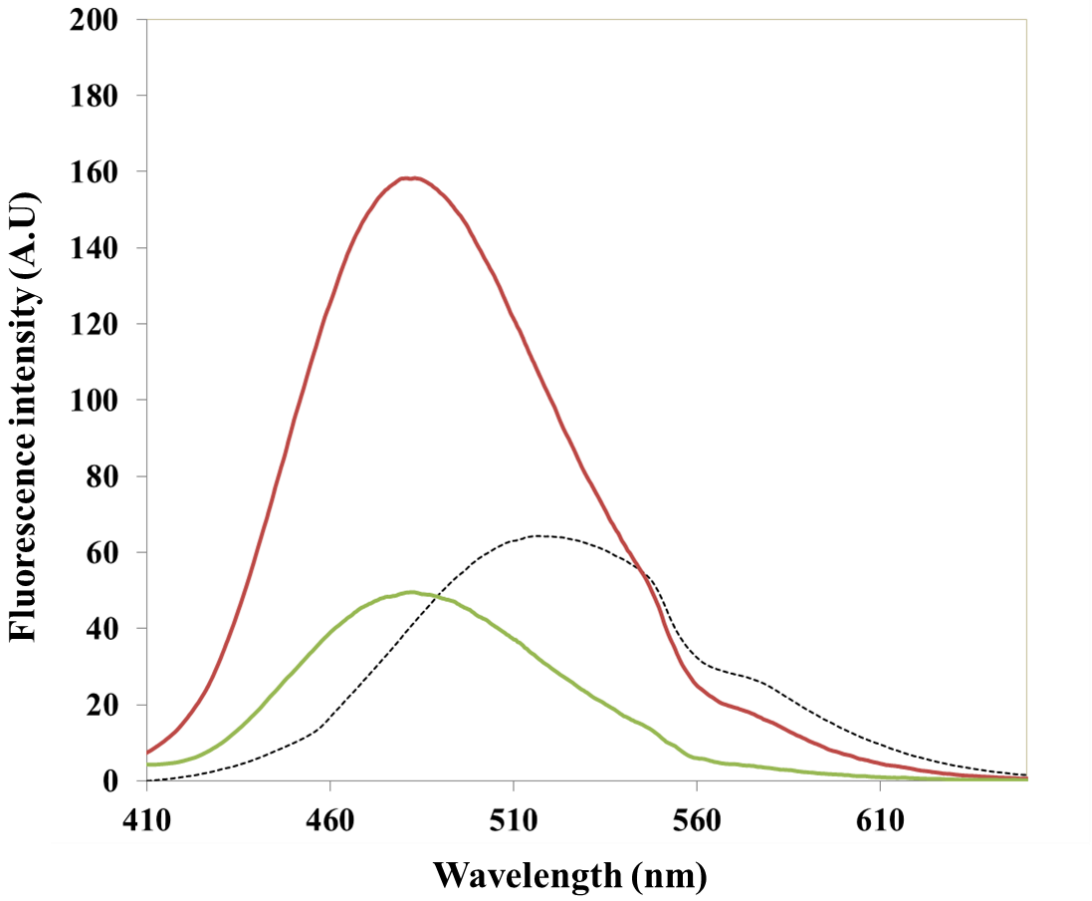


Figure 4.4: ANS fluorescence emission spectra of *Pf*GST in the presence and absence of BSP. ANS, bound to *Pf*GST in the presence (green) and absence of BSP (red), was excited at 390 nm and the emission spectra were recorder from 400-700 nm. The free ANS and BSP emission spectra were subtracted from the corresponding ANS-protein spectra in the data analysis, therefore, correcting for the contribution of unbound ANS and BSP. The ANS spectrum (dash-line) was included.

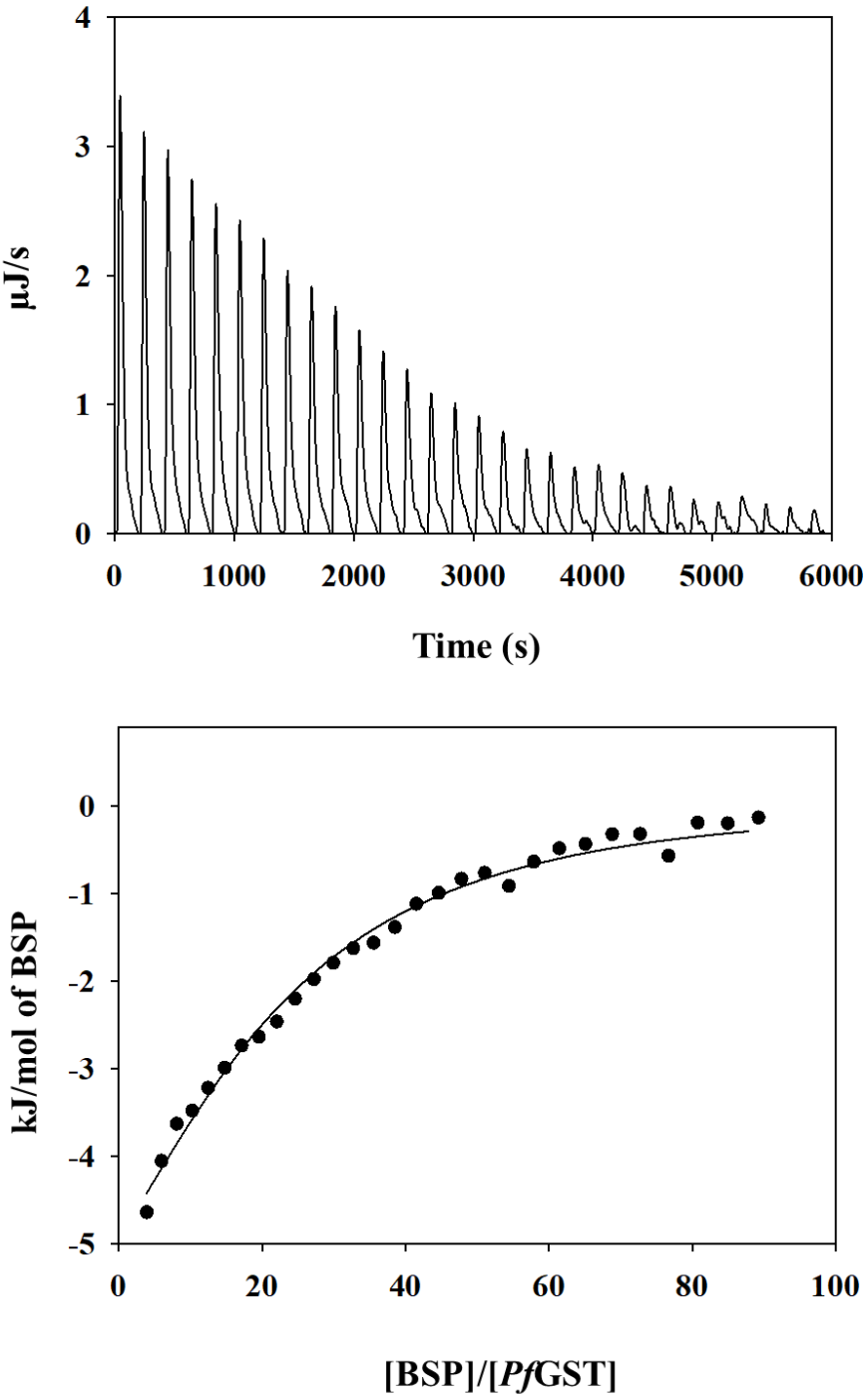
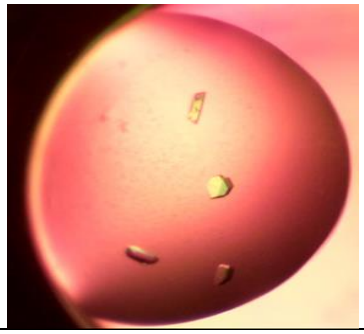
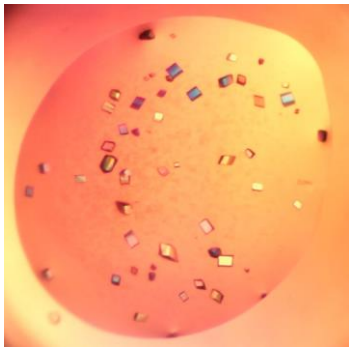
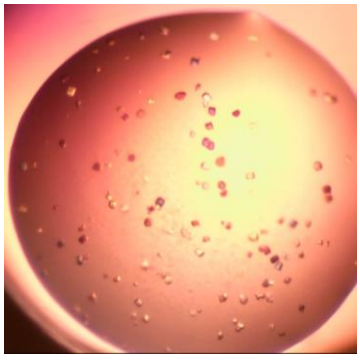


Figure 4.5: The titration profile of *Pf*GST with BSP in the presence and absence of GSH. (A) 7 mM BSP solution was titrated against 50 µM *Pf*GST, in the presence of 5 mM GSH (25°C). Once equilibrated (stirring 75 rpm), BSP was injected (5 µL injections with a total of 30 injections over 6000 sec) into the sample cell containing the *Pf*GST-GSH. The interaction yielded a stoichiometry of n=2 with the first binding site being entropically driven and a K_d value in the millimolar range. The second binding site is entropically driven with a K_d value in the micromolar range. The data has been corrected for by overlaying the heats generated from the titration of BSP into *Pf*GST with the heats generated from the titration of BSP into buffer.

Table 4.2: The thermodynamic of the *Pf*GST-BSP interaction, in the presence of GSH, based on a sequential two site model.

Variable	Value
Kd_1 (M)	9.94E-2
Kd_2 (M)	2.92E-4
ΔH_1 (kJ/mol)	-4787
ΔH_2 (kJ/mol)	4751
ΔS_1 (J/mol·K)	-1.60E4
ΔS_2 (J/mol·K)	1.60E4
ΔG_1 (kJ/mol)	-19
ΔG_2 (kJ/mol)	-17

Table 4.3: Crystallisation hits for the *apo* and complex forms of *Pf*GST in combination with GSH and BSP

Complex	Condition	Crystal Images
*Pf*GST	0.1 M HEPES pH 7.5, 2.0 M Ammonium sulfate	
*Pf*GST-GSH	0.1 M Tris pH 8.5, 25% w/v Polyethylene glycol 3,350	
*Pf*GST- BSP	0.2 M Ammonium citrate tribasic pH 7.0, 20% w/v Polyethylene glycol 3,350	
*Pf*GST-GSH-BSP	0.2 M Sodium chloride, 0.1 M Tris pH 8.5, 25% w/v Polyethylene glycol 3,350	

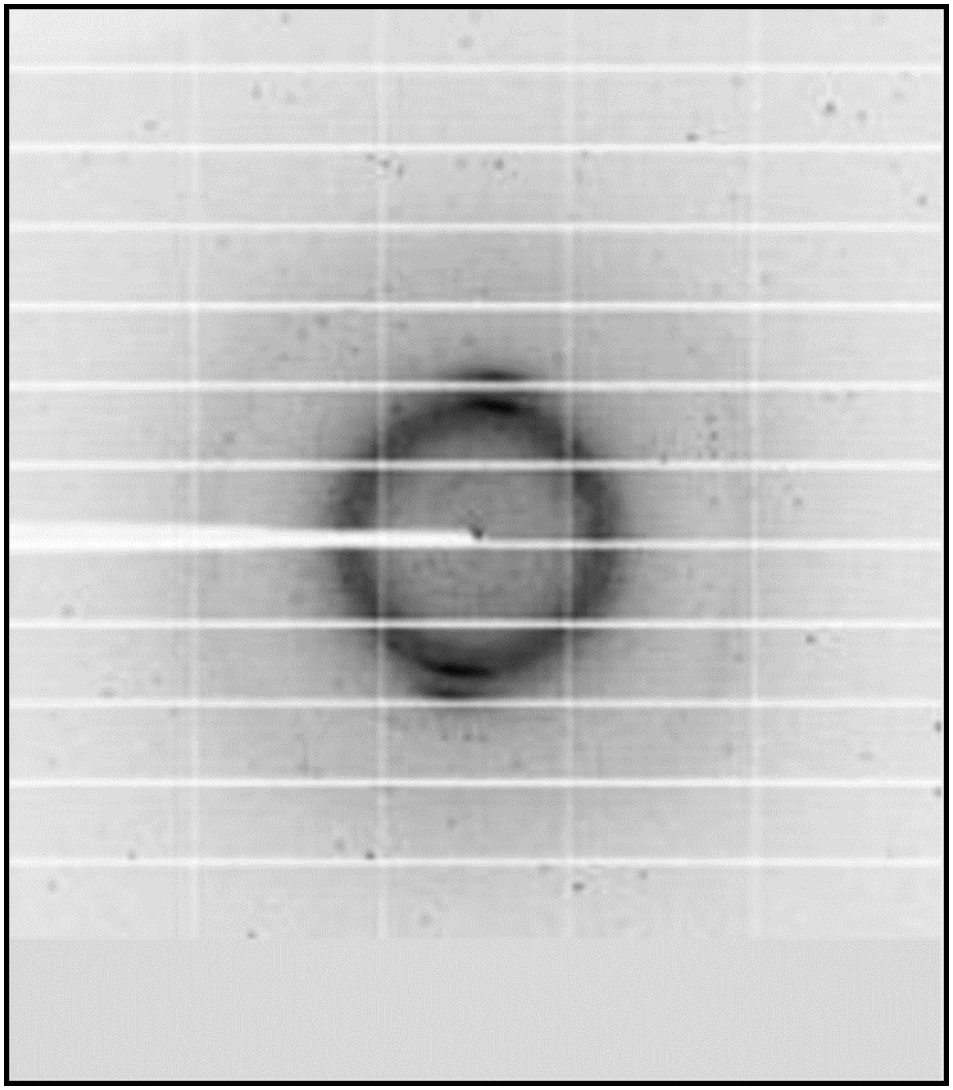


Figure 4.6: X-ray diffraction pattern of *Pf*GST-GSH-BSP. The data was collected at a resolution of 1.75 Å. The space group of the crystal was P 622. The A, B and C lengths were 110.67 Å, 110.67 Å and 74,97 Å and the angles α, β and γ were 90°, 90° and 120°.

4.5 Discussion

Functional characterisation of *Pf*GST (in the presence and absence of BSP) was performed using an ultraviolet chromogenic substrate, CDNB, which becomes conjugated with GSH in order to form 1-(*S*-glutathionyl)-2,4-dinitrobenzene, that can be measured at 340 nm. It must be noted that the concentrations of enzyme used for all the enzyme assays was higher than the likes of GSTA1 and GSTP1-1 (Kolobe and Sayed, 2004; Sayed, 2001) because this enzyme was observed to have low GSH-CDNB activity (Supplementary Figure 3). Therefore, the decrease in sensitivity of the assay was compensated by an increase in the concentration of enzyme. The specific activity of *Pf*GST first determined to assess the functionality of the enzyme (Figure 4.1). The specific activity value of 0.84 $\mu mol.min^{-1}.mg^{-1}$ (37°C) was slightly higher than the specific activity of 0.2 $\mu mol.min^{-1}.mg^{-1}$ (25°C) reported by Hardwaldt *et al*. (2002).

The difference in the specific activity observed between both studies could be due to the temperature differences at which the activity assay was performed. Enzyme activity is temperature dependent; increasing with a rise in temperature until it reaches the temperature that it denatures, at which the activity will then decrease. The host temperature upon infection is 37°C, therefore, an increase in the specific activity of the enzyme at 37°C in comparison to 25°C will be observed as the assay temperature matches the physiological temperature at which the enzyme naturally functions (Bisswanger, 2014). In addition, the most prominent symptom of malaria is a fever. During the asexual stage where the mature schizonts rupture, the temperature of the body reaches a temperature of 41°C (Baron, 1996). It was reported that the *P. falciparum* parasite was able to resist high temperatures associated with malarial infection, for short periods of time (Baron, 1996). Therefore, the enzyme naturally works at higher temperatures and it can be suggested that there is an increase in specific activity of the enzyme at a higher temperature (Rojas and Wasserman, 1993). Following this, a specific activity assay was done in the presence of BSP. There was a 2.4 fold decrease in the overall enzyme's specific activity, therefore, BSP did bind and inhibit the specific activity of *Pf*GST.

Given that there was inhibition of *Pf*GST specific activity in the presence of BSP, an IC_{50} was performed in order to investigate the concentration at which BSP inhibited

50% of the *Pf*GST's activity. The IC_{50} was determined as 537 µM (Figure 4.3). A study done by Kolobe *et al*. (2002), showed a much lower IC_{50} value of 7 µM for the binding of BSP with hGSTA1-1. However, this could be a result in using a much higher *Pf*GST concentration to compensate for the lower GSH-CDNB assay sensitivity (Sayed, 2001). In addition, the difference in the IC_{50} values can be accounted for by the fact that the protein in this study is different to the one investigated by Kolobe *et al*. (2002). Both the IC_{50} and specific activity assay in the presence of BSP, showed that there was overall inhibition of *Pf*GST by BSP. Both experiments also provided the concentration of BSP that were used further in the structural studies using crystallography.

Following this, the K_M was determined for both GSH and CDNB, using varied concentrations of the respective substrates. A study done by Harwaldt *et al.* (2002), reported a much lower K_M of 0,16 mM in comparison to the 5,41 mM that was determined. This could be the result of doing the kinetics assay at a lower temperature (Bisswanger, 2014). The K_M value for varied concentrations of CDNB could not be determined accurately as the reaction did not reach saturation. Harwaldt *et al.* (2002), estimated a K_M value of >2 mM, however, due to the strong absorption nature of CDNB, concentrations higher than 1.5 mM cannot be tested. In this study, a higher concentration of protein was used due to the decrease in the protein's sensitivity to GSH-CDNB assay, which resulted in a higher concentration of substrate CDNB used. Therefore, due to the strong absorption nature of CDNB stated above, we could not determine the K_M accurately.

Fluorescence spectroscopy, using the extrinsic fluorophore ANS, has been used to probe binding sites of *Pf*GST and in addition monitor the binding of the ligand, BSP, to *Pf*GST. Changes in the surroundings of the fluorophore causes changes in the fluorescent intensity and shifts in the maximum wavelength emission (Stryer *et al*., 1965). When ANS binds to *Pf*GST, in the absence of BSP, there is increase in the fluorescent intensity and blue shift in the maximum emission wavelength. This is characteristic of a decrease in the polarity of the environment of the fluorophore (Stryer *et al*., 1965) and indicates that ANS is binding to surface hydrophobic patches/pockets of the *Pf*GST. A study done by Kolobe *et al*. (2004), suggested that ANS most likely bind at the H-site of the hGSTA1-1. The H-site is an all-alpha

helical structure found on the C-terminal domain (domain II), which allows for the recognition and binding of a variety of hydrophobic substrate. Therefore, it can be inferred that the ANS molecule is binding the hydrophobic H-site of *Pf*GST, therefore, causing an increase in the fluorescence intensity. ANS in the absence of *Pf*GST and BSP, shows a decrease in the fluorescent intensity and an overall red shift. This is due to increase in the polarity of the environment around the fluorophore i.e. the solvent (Möller and Denicola, 2002).

Interestingly, ANS is also used to evaluate the interaction between the protein and ligand by looking at the displacement of the dye due to the ligand competing for the same binding site. In the presence of *Pf*GST complexed with BSP, the fluorescence intensity decreased significantly. This could be due to BSP displacing ANS from the hydrophobic patch/pocket, therefore, decreasing the fluorescent intensity. In the *De et al.* (2005) study, ANS was displaced from albumin due to the binding of the ligand SDS, therefore, causing a decrease in the fluorescence intensity. Mogensen *et al.* (2002) also showed that there was a decrease in fluorescence signal upon the binding of a variety of ligands. The non-substrate is known to bind to the GST's H-site (Oakley *et al.* (1999) and Sluis-Cremer *et al.* (1998) showed that BSP binds non-competitively with ANS when binding to human class Alpha GST with two type 1 subunits (hGSTA1-1). Therefore, this could be the reason for the displacement of ANS and therefore decrease in the fluorescence intensity. It must be noted that all ANS experiments in the presence of GSH. This is because *Pf*GST is in its active dimeric form when GSH is bound and the dimeric form is the form in which majority of the *Pf*GST exist physiologically (Tripathi *et al.*, 2007).

A sequential two-site model was shown to be the best model for *Pf*GST complexed to BSP in the presence GSH (Figure 4.5). This suggests that *Pf*GST has two types of co-operative and/or allosteric binding sites for BSP. This correlated with Kolobe *et al.* (2004) study on hGSTA1-1, which showed a high-affinity and a low-affinity binding site for BSP. In the presence of GSH (dimeric form of *Pf*GST), the first binding site had a low affinity for BSP. There is no crystal structure for *Pf*GST bound to BSP, however, the binding sites could be inferred using other studies and literature. The low affinity site was proposed to be BSP binding to the H-site. The H-site is known to be promiscuous and can bind a variety of hydrophobic molecules with a low affinity

(Valenzuela-Chavira *et al.*, 2017). For both BSP binding events, the binding is exergonic, therefore, the binding of BSP to *Pf*GST was spontaneous (Bronowska, 2011). The first binding site was enthalpically driven with an unfavourable entropy. The unfavourable entropy could be due to the fact that when GSH binds to the G-site it causes the exposure of hydrophobic residues in the H-site, therefore, leading to an unfavourable entropy (Valenzuela-Chavira *et al.*, 2017). This has been observed in other GST isoenzymes, one of which is human GSTP1-1 (Pi-class enzyme) which is known to have a similar H-site to *Pf*GST (Harwaldt *et al.*, 2002). A study by Barratt *et al*. (2006), found that not all binding between a protein and a ligand is entropically driven, but rather enthalpically driven. The mouse major urinary protein (MUP) in the study had a strongly hydrophobic binding pocket which has the ability to bind a variety of hydrophobic ligands (Barratt *et al.*, 2006), like *Pf*GST. MUP shows sub-optimal solvation of the binding pocket in its *apo* conformation, therefore, only a few waters remained in the binding pocket before the ligand bound to it (Barratt *et al.*, 2006). The ITC data showed that the binding was enthalpically driven which was due to desolvation of the ligand with little desolvation of the protein, which is referred to as the non-classical hydrophobic effect (Bronowska, 2011). Therefore, the enthalpically driven entropically unfavourable binding of *Pf*GST to BSP at the first site, could be due to the exposure of the hydrophobic residues at the H-site along (Valenzuela-Chavira *et al.*, 2017) with the non-classical hydrophobic effect shown in Barratt *et al.* (2006) study of the MUP study.

The second binding site has a much greater affinity for BSP than the first binding site. Due to the co-operative binding of *Pf*GST to BSP, it is proposed that there is some change in the 3D conformation of the protein when BSP binds the H-site. A study done by Cutler *et al.* (1999), showed that there was a change in the proteins conformational upon binding of CDNB to the H-site. Therefore, when BSP binds to the first site, proposed to be the H-site, there was a change in *Pf*GST's conformation which then lead to the second BSP molecule binding to *Pf*GST. In order to predict where this binding site is, molecular docking and dynamic studies were performed (see Section *5.3*). Molecular docking allows for the prediction of the position and orientation of a ligand to its receptor protein. The molecular docking studies showed that BSP was bound to the dimer interface, also known as the L-site given that McTigue *et al.* (1995) and Ji *et al.* (1996) showed that there was ligand binding at the

dimer interface (L-site) in both *Schistosoma japonica* and *squid* GST. Using the molecular docking studies, it can be inferred that the second binding event of BSP to *Pf*GST could be located at the L-site. The L-site in *Pf*GST, is a cavity found between the two monomeric subunits and is solvent exposed (Figure 5.2). Upon the binding of BSP to the cavity, there is burial of water-accessible surface area in the cavity at the L-site which results in the release of confined or interfacial water molecules to the bulk solvent (Bronowska, 2011). Therefore, the binding is entropically favourable which agrees with the ITC data, suggesting that the second binding event is entropically driven. The binding of the BSP is enthalpically unfavourable which could be due to the disruption and loss of mainly hydrogen bonds and van der Waals interactions between the protein and solvent and the ligand and solvent (Fischer and Verma, 1999).

In order to determine the position and orientation of BSP when binding *Pf*GST, crystal trials were set up using the Hampton Research Index™ reagent kit (HR2-144). The protein crystal was diffracted (Figure 4.6), however, the data collected during the diffraction was not completed. Data completion is important and if it is not done accurately or there is a diffraction power decay, the collection of reflections will not be complete; thereby decrease the quality and interpretability of the data that contribute to the electron-density maps, which are calculated using the Fourier transforms. In this study, there was an overall 89.9% completion of the data, with a 98.7% inner shell completion and 69.3% outer shell completion. The outer shell has the highest resolution data, therefore, with the lack of completion of the data in the outer shell, the structure could not be solved. However, incomplete data sets are still used in order to solve a structure by using symmetry. However, this is dependent on the space group of the protein, along with internal symmetry within the protein. In this study, the space group in which the protein had crystallised, specifically the hexagonal crystal system, rendered us from completing the data as symmetry could not be applied in order to complete the set of data. Therefore, the structure could not be solved. The incomplete data collection was as a result of the small crystal size. If a small crystal is exposed to the X-ray beam for long periods of time, the beam could potentially damage the crystal, rendering the data collected as unusable. Therefore, in order to avoid damage to the small protein crystal, the exposure time to the beam was decreased, therefore, less data was collected.

Chapter 5

Molecular modelling of interactions between *Pf*GST and BSP

5.1 Introduction

Molecular docking and dynamics uses algorithms based on Newtonian laws to determine the chemical properties of atoms and molecules in order to simulate chemical events such as the binding of a ligand to a protein (Palermo and De Vivo, 2014). In order to provide insight into the structural interactions underlying the binding and inhibition of *Pf*GST by BSP, molecular docking and molecular dynamic (MD) simulations were performed using the Induced Fit Docking (IFD) algorithm (Schrödinger Release 2017-3) and Desmond algorithm (Schrödinger Release 2017-3) respectively. Once the structures were extracted and prepared, *Pf*GST and BSP were docked. Following this, MD simulations were performed on the *apo* form of *Pf*GST and the top ranked pose of BSP, from IFD, with *Pf*GST.

5.2 Materials

All molecular modelling studies were performed using a desk top computer equipped with Intel Core i7-5960X (3.3GHz, 20M Cache, 16x Cores) Overclocked up to 4.3GHz Per Core Extreme Edition CPU with the first Haswell-E Eight-Core 16 Threads CPU processor, DDR4 Intel X99 SATA 6Gb/s USB 3.0 ATX Motherboard Nvidia GTX 4GB 256bit (2048 Cuda cores) DDR5 overclocked edition graphics card. The CPU was equipped with Series H100i Extreme Performance Liquid CPU Cooler [DDR4] 16GB Quad Channel (4x4GB) DDR4 2400MHz High Performance Gaming RAM.

Associated software used in this study includes SWISS-MODEL^{TM} (UniProt code: Q8MU52) which was used in homology modelling of *Pf*GST. A ProCheck analysis, using EMBL-EBI database, was performed on the protein structure produced from the homology modelling, after energy minimisation of the protein structures was done. The BSP ligand was constructed and underwent initial energy minimisation using Maestro Elements 10.1 (Schrödinger, 2015). Molecular docking was performed using multiple algorithms implemented in the Schrödinger Suite 2017-3 (Schrödinger, LLC, New York, NY, 2017). Protein Preparation Wizard (Schrödinger Release 2017-3)

(Sastry *et al.*, 2013) was used to amend structural problems derived from the crystallographic structure, thereby creating dependable all-atom protein 3D models for molecular docking and molecular dynamic (MD) simulation. Once the protein structure was evaluated, MacroModel algorithm (Schrödinger Release 2017-3) were used to prepare the ligands and perform energy minimisation on the ligands prior to molecular docking. The molecular docking was carried out using the Induced Fit Docking algorithm (Schrödinger Release 2017-3) (Farid *et al.*, 2006). Molecular dynamic (MD) simulations were performed on the *apo* and top-performing ligand and the simulated interaction diagram (SID^{TM}) diagrams were produced using the Desmond^{TM} algorithm (Schrödinger Release 2017-3).

The homology model based on the crystallographic protein structure, 3FRC, was utilised in both the ligand docking and MD simulations. 3FRC, classified as transferase, was first deposited on the 8th of January 2009, and released on the 26th of June 2009 by Savvides & Karplus. The protein is found in *Plasmodium falciparum* and has been expressed in *Escherichia coli*. The structure has a resolution of 2.0 Å, with a R-Value free and work of 0.27 and 0.22. The homology model produced is the most accurate active dimer conformation of *Pf*GST, therefore, the structure is considered as good and reliable to be used in the study.

5.3 Methods

5.3.1 Protein structure preparation and model validation

In order to perform ligand docking and molecular dynamic studies, the *Pf*GST enzyme was modeled using SWISS-MODEL^{TM} (Biasini *et al.*, 2014) (https://swissmodel.expasy.org/). SWISS-MODEL^{TM} is an automated system that predicts the 3D structure of a protein based on its amino acid sequence by means of homology modeling techniques (Schrödinger Release 2017-3 & Sastry, *et al*., 2013). The FASTA sequence for the primary sequence of *Pf*GST was obtained from Uniport (UniProt code: Q8MU52) and used as the query sequence. The queried FASTA sequence of the protein was then uploaded into the SWISS-MODEL^{TM} server were the sequence underwent sequence alignment using BLAST (Basic Local Alignment Search Tool), in order to identify template sequence that best matches the query sequence. Thereafter, based on the alignment between the query and the template sequence of a known protein sequence, the homology model was built. The template

that accurately represented the active dimer of *Pf*GST was the protein crystal structure with the PDB code 3FRC. This crystallographic structure had missing amino acids due to the flexible loop region in chain A (residue numbers: 144-150) and in chain B (residue numbers: 36-38 and 144-148). However, SWISS-MODEL™ allows for the completion of these flexible regions, producing a sound model, which was then used for induced-fit molecular docking (IFMD).

5.3.2 Protein and ligand structure preparation for molecular docking

BSP was constructed and validated using Maestro Elements 11.2 (Schrödinger, 2015). Maestro provides an intuitive and full-featured building tool that allows for the construction of molecular models of any type, in this case, the construction of BSP (Schrödinger, 2015). The constructed structure of BSP underwent structural cleaning and initial energy minimisation (OPLS 2005) on each atom (Schrödinger Release 2017-3).

*Pf*GST underwent protein pre-processing and energy minimisation using the Protein Preparation Wizard™ tool which ensured structural correctness at the onset of the project, producing high-confidence structures for modelling applications (Sastry *et al.*, 2013). The protein was initially pre-processed which includes assigning bond orders, the addition of hydrogens, deleting waters beyond 5Å from heteroatom groups and using Epik™ to generate heteroatom states at a pH of 7. These steps allowed for any structural problems to be fixed, producing sound structures. The protein structure was then reviewed, modified if required and lastly refined. Refinement of the protein included optimisation which assigned hydrogen bonds using Propka™ at a pH of 7. The final step then included a restrained minimisation, by using the OPLS 2005 force field. Following this, the ligands were prepared and energy minimisation was performed using MacroModel™ with 500 cycles of steepest descend and 5000 cycles of a conjugated gradient, using an OPLS 2005 force field.

Subsequent to the preparation and energy minimisation of the protein and ligand, BSP was docked to *Pf*GST using the induce fit ligand protocol. This protocol starts by docking the ligands with Glide™ which uses reduced van der Waals radii and an increased Coulomb-van der Waals cut-off, therefore, producing multiple poses. This

protocol briefly removes very flexible side chains during the docking stage. Once a pose was produced, a Prime structure prediction was used to re-orientate nearby side chains, to accommodate the ligand. The residues and ligand underwent energy minimisation and the ligands were docked into the corresponding low energy protein structure and the complexes are then ranked according to GlideScore™ (Sherman *et al.*, 2006). Therefore, the ligands were docked in *Pf*GST using the above protocol for Schrödinger Suite 2017-3 Induced Fit Docking, with an OPLS 2005 force field. The empirical scoring function, Glidescore, gives an approximate binding free energy of the ligand when complexed to the protein and has been optimised for docking accuracy and binding affinity prediction. Contributions from force field (which includes electrostatic and van der Waals forces) along with terms that reward or penalise interactions, allow for the ranking of the different poses. Therefore, the more negative the glide score, the higher the binding affinity of the ligand for the protein. Another important scoring function is the Glide emodel, who's scoring relies heavily on the force field component (Electrostatic and van der Waals energies) (Friesner *et al.*, 2004). Therefore, it is reliable in comparing conformers, playing a major role in pose selection, and determining the best pose. Accuracy is ensured by Glide™ , which has a superior scoring function, and Prime™, which has an advanced conformational refinement (Sherman *et al.*, 2006).

5.3.3 MD simulation of PfGST in the absence and presence of BSP

MD simulations for the *apo*- and BSP-bound conformation of *Pf*GST were performed using Desmond™. MD simulations is a central computational tool used to monitor the dynamic nature of protein molecules which is often an issue within drug discovery (Shivakumar, *et al.*, 2010). MD simulations are performed on a time scale that provides parallel scalability, simulation throughput and scientific accuracy (Shivakumar, *et al.*, 2010). The *apo*- and BSP-bound conformation of *Pf*GST were investigated using MD simulations to understand the stability of the *apo Pf*GST in relation to the stability of the ligand's interactions with *Pf*GST, under physiological conditions (25°C, pH 7) over 25 ns. The GSH ligand was not included in the simulation as the model used was already in its dimeric form and including the GSH peptide poses a challenge in running the simulation. The *apo* conformation of *Pf*GST was pre-processed and optimised (see Section 5.3.2). The *Pf*GST in complex to BSP

previously underwent both pre-processing and optimisation. Following this, both systems underwent solvation and the addition of ions via the System Builder, which then built a simulation box around the protein to define the region were MD simulation will occur. This occurred through the wrapping of a pre-equilibrated and repeated water box around the protein (Shivakumar, *et al*., 2010) but any overlapping waters were removed. The system was set up using the TIP3 explicit solvent model in a cubic simulation box with the 20 Å distance of the buffer in all directions of the protein. The system was neutralised, and the ionic strength was set to 0.15 M using sodium ions (Na^+) and chloride ions (Cl^-), to mimic physiological conditions. The simulations were run in the isothermal isobaric (temperature and pressure) ensembles (NPT), and the temperature was kept constant at 300 K and the pressure at 1.01 bar. Default settings were used for all the other options. The results of molecular dynamics simulation were analysed using SID™ (Simulation Interaction Diagram), implemented in Desmond™ (Schrödinger Release 2017-3). This included the root mean square deviation (RMSD) of the Cα, backbone and sidechains of the protein and the root mean square fluctuation (RMSF) of the protein's sidechains.

The RMSD measures the average change in the displacement of a selection of atoms for a precise frame in comparison to the reference frame. This is calculated for each frame in the trajectory, therefore, the RMSD for frame x is

$$RMSD_X = \sqrt{\frac{1}{N}\ \sum_{i=1}^{N}\left(r'_i(t_x)\right) - \left(r_i\left(t_{ref}\right)\right)^2} \qquad (9)$$

where N is the number of atoms in the atom selection; t_{ref} is the reference time (typically the first frame is used as the reference and it is regarded as time t=0) and r' is the position of the selected atoms in frame x after superimposing on the reference frame, where frame x is recorded at time t_x. This procedure is then repeated for each frame in the simulation trajectory (Schrödinger Release 2017-3).

By evaluating the RMSD of the protein, either in its apo conformation or when complexed with the ligand, the structural conformation throughout the simulation can be assessed and indicate if the system has equilibrated. Changes less than 3 Å (within a hydrogen bonding distance), are considered acceptable for small globular proteins (Schrödinger Release 2017-3). If the change occurred are larger than this range, the

protein underwent large conformational changes and equilibration was not reached. during the simulation. It is also important that your simulation converges. If the fluctuations increased or decreased on average at the end of the simulation, the system is considered as not equilibrated and the simulation should be done for a longer time. In this simulation,

The RMSF is useful for characterizing local changes along the protein chain. The RMSF for residue i is:

$$RMSF_i = \sqrt{\frac{1}{T}\ \sum_{t=1}^{N} < \left(r'_i(t)\right) - \left(r_i\left(t_{ref}\right)\right)^2 >} \qquad (10)$$

where T is the trajectory time over which the RMSF is calculated, t_{ref} is the reference time, r_i is the position of residue i; r' is the position of atoms in residue i after superposition on the reference, and the angle brackets indicate that the average of the square distance which is taken over the selection of atoms in the residue (Schrödinger Release 2017-3).

The contacts made by the ligand on the protein was also analysed using a protein-ligand contacts bar graph. The protein-ligand interactions are categorised as either hydrogen bonds, ionic bonds, hydrophobic interactions, and water bridges. The stacked bar charts are normalised over the course of the trajectory, where the interaction fraction suggests the percentage of the simulation time that the specific interaction is maintained (Schrödinger Release 2017-3).

5.4 Results

5.4.1 Construction and validating the homology modelling of PfGST

Homology modelling was done in order to produced is the most accurate active dimer conformation of *Pf*GST, which was based on the *Pf*GST crystal structure (PDB code: 3FRC). The model produced, was used in both ligand docking and MD simulations. The original crystal structure *Pf*GST (PDB code: 3FRC) and the newly built homology model were aligned (Figure 5.1) and the RMSD value of the Cα chain were determined as 0.20 Å. The homology model was then validated by performing a ProCheck analysis in order to assess the stereochemical properties of the original crystallographic structure and the newly built homology model. (Table 5.1), which included the Ramachandran plots for both models (Supplementary Figure 4). There was a slight increase of 1% observed in the most favoured regions of the homology modelled *Pf*GST, in comparison to the original structure. The additional allowed regions decreased of 0.6% for the homology modeled *Pf*GST, in comparison to the original *Pf*GST structure. The generously allowed regions showed an increase of 0.2% for the homology modelled *Pf*GST, in comparison to the original *Pf*GST structure. Finally, there was a decrease of 0.5% for the homology modeled *Pf*GST, in comparison to the original *Pf*GST structure. Therefore, there were no significant changes in the stereochemical properties between the original crystallographic structure and the newly built homology model.

5.4.2 Induce Fit™ ligand docking, using Glide™ and Prime™, of PfGST and BSP

BSP, after structural cleaning and energy minimisation (using Maestro 11.2), was docked into the homology modelled *Pf*GST in order to determine the binding affinity and energy associated with the binding event. All 9 poses were shown to interact at the L-site of the *Pf*GST protein. The respective docking scores obtained from the induced fit (Table 5.2) indicated very slight differences between the 9 poses produced, especially between the top 3 poses. A ribbon structure and surface representation of *Pf*GST structure in complex with the 9 different poses of BSP confirmed that BSP docked at the L-site (Figure 5.2) and that the ligands aligned with one another, suggesting that BSP did not undergo significant changes in its structural conformation when at the L-site. Details of hydrogen bonding, pi stacking and electrostatic interactions between the BSP and *Pf*GST are found in the 2D interaction diagrams (Figure 5.3 B), with its respective 3D image (Figure 5.3 A).



Figure 5.1: Ribbon representation of the alignment between *the* original *Pf*GST structure (green) and modelled *Pf*GST (blue). The crystallographic structure of *Pf*GST has missing amino acids due to flexible loop regions in chain A (residue numbers: 144-150) and in chain B (residue numbers: 36-38 and 144-148). The red circles show where the missing amino acids in 3FRC were modelled. This image was created using Pymol V1.3.

Table 5.1: Analysis of the Ramachandran plot stereochemical properties of the *Pf*GST model before and after homology modelling

Parameters	Before (%)	After (%)
Most favoured regions	88.2	89.2
Additional allowed regions	9.6	9.0
Generously allowed regions	1.4	1.6
Disallowed regions	0.8	0.3

Table 5.2: A summary of the docking scores produced from IFMD of PfGST with BSP

Ranked Pose	Glide score	Glide vdw	Glide emodel	Glide energy	glide einternal	IFD scores
	kcal.mol^{-1}					
1	-9.08	-51.22	-122.35	-83.42	7.62	-968.67
2	-9.40	-56.87	-123.04	-82.22	6.72	-968.20
3	-9.11	-57.96	-118.72	-80.92	7.12	-967.67
4	-8.27	-48.25	-115,09	-77.23	4.75	-967.15
5	-8.32	-45.74	-107.45	-73.67	7.15	-967.13
6	-7.55	-53.91	-107.28	-72.49	3.99	-966.72
7	-7.67	-44.59	-97.68	-68.23	4.37	-965.95
8	-7.17	-54.15	-103.49	-72.50	4.43	-965.83
9	-7.50	-49.85	-108.86	-73.86	3.61	-965.65

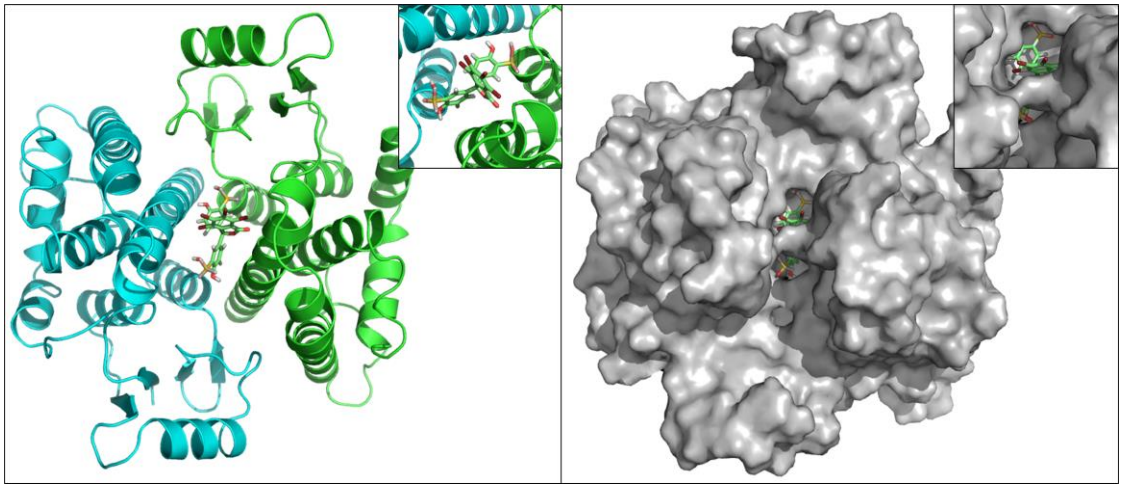


Figure 5.2: *PF*GST-BSP complex. Diagrams showing the (A) ribbon structure and (B) surface representation of the *Pf*GST structure in complex with the 9 different poses of BSP, along with a magnified image of the ligand at the L-site. Both diagrams show that BSP binds at the L-site of the protein which is a cavity formed between both subunits. There was little difference in the poses of BSP and all 9 poses align with one another. This suggested that there are minimal movement of the ligand at the L-site. Both images were created using Pymol V1.3.

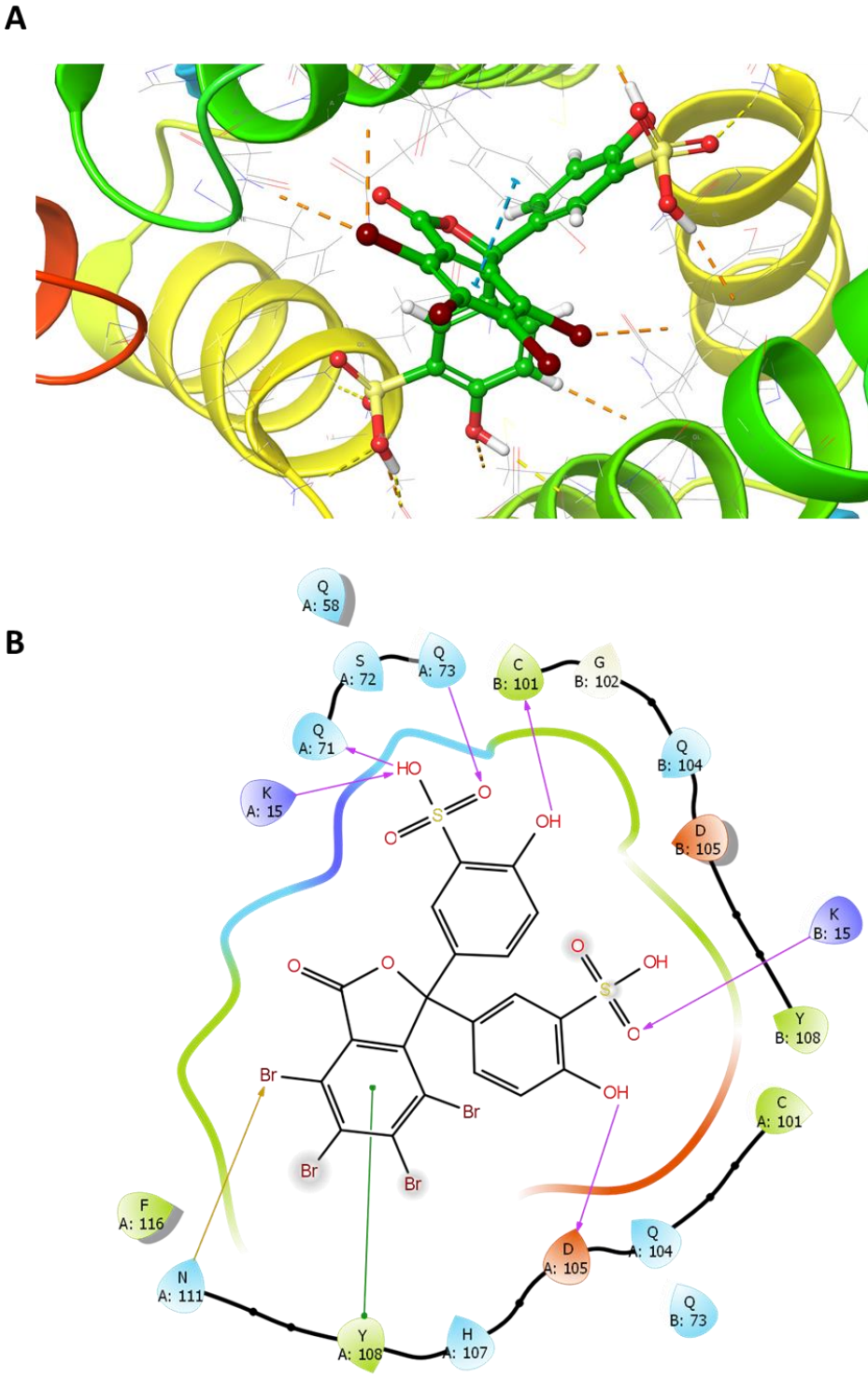


Figure 5.3: The ligand-protein interactions that occurred during IFMD of *Pf*GST and BSP. (A) A 3D diagram showing the top pose ligand which interacted with the *Pf*GST. (B) The 2D diagram shows the top pose for the interaction between BSP and *Pf*GST through hydrogen bonding (purple arrow), π-π stacking and weak hydrogen bonds (yellow arrow). The negatively charged residues are shown in red, positively charged residues are shown in purple, polar residues are in light blue, hydrophobic residues are in green and glycine residue is white. Both images were created using Maestro 11.2 (Schrödinger, 2015).

5.4.3 MD simulation of the apo- and BSP-bound conformation of PfGST

MD simulation studies were completed with the top pose BSP ligand, from IFMD, which was complexed to *Pf*GST. This was repeated with the *apo* conformation of *Pf*GST. The results from the simulation, done over 25 ns, were examined using SID^{TM}. The RMSD for both systems were determined for the protein's Cα chain, the backbone, and the sidechains. The RMSD of the protein's Cα chain (Figure 5.4 A) and backbone (Figure 5.4 B) for the *apo* conformation of *Pf*GST increased slightly at the beginning of the simulation until stabilisation occurred at ~5 ns. There was an increase in the RMSD of the protein C^{α} chain and backbone initially at the beginning of the simulation for *Pf*GST-BSP complex. Stabilisation occurred for both the protein's Cα and backbone's RMSD until ~ 19 ns, where an increase of ~1 Å occurred. The RMSD then decreased at 23 ns, for both the protein's Cα and backbone until the end of the simulation. The RMSD for the sidechain's (Figure 5.4 C) for both the *apo*- and BSP-*Pf*GST conformations increased at the beginning of the simulation. Stabilisation of the sidechain's RMSD occurred from ~5 ns until the end of the simulation for the apo conformation of *Pf*GST. For the *Pf*GST-BSP complex, there was a continual increase in the RMSD until ~23 ns, followed by a decrease until the end of the simulation. In the presence of BSP, there was a definite increase in the protein's RMSD, in terms of the protein's Cα, backbone and sidechains. For both, systems, the fluctuation occurred within the allowed range of 1-3 Å, for globular proteins (Schrödinger Release 2017-3). Therefore, both systems are considered as equilibrated.

The ligand RMSD was also calculated, which showed how stable the ligand is with respect to the protein and the binding pocket (Figure 5.4 D). There was an increase in the RMSD for both the ligand fitted onto the protein and ligand fitted onto the ligand. The ligand fitted on the protein showed an increase in the RMSD until stabilisation occurred at ~22.2 ns. The ligand fitted to the ligand showed an increase in the RMSD until stabilisation occurred at ~10 ns. Both RMSD's suggest that there was a definite increase in the protein's dynamics, when BSP was bound to the protein.

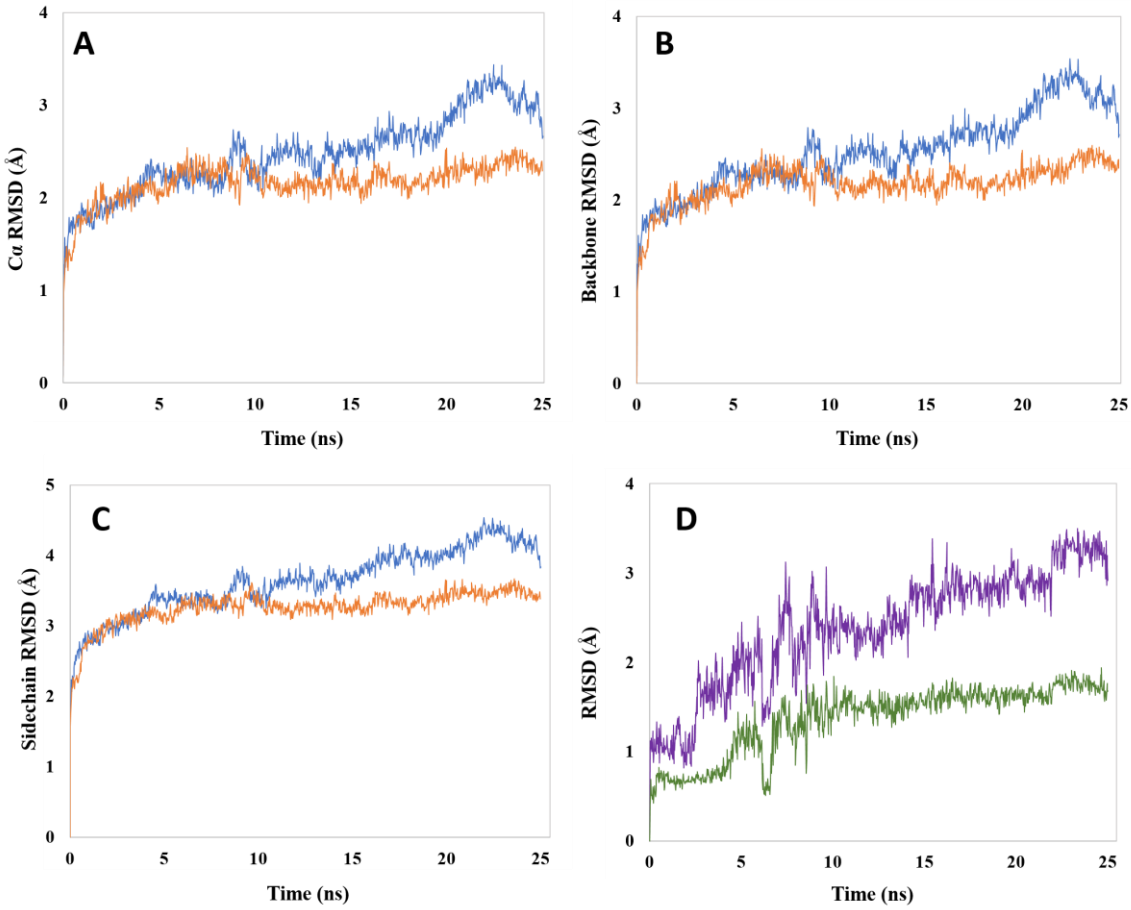


Figure 5.4: RMSD of the apo *Pf*GST vs the *Pf*GST in complex with BSP. A line graph showing the RMSD of the protein (A) Cα chain, (B) backbone and the (C) sidechain of *Pf*GST both in its *apo* conformation (orange) and when bound to BSP (blue). Fluctuations, both in the apo and complexed *Pf*GST, for all 3 graphs occurred less than 2 Å. Therefore, the system was considered to be equilibrated. (D) A line graph showing the RMSD of the BSP ligand when bound to the protein (purple) when the protein-ligand complex is first aligned on the protein backbone of the reference and the RMSD of the ligand that is aligned and measured on its reference ligand conformation (green). All the data was extracted from SID™ from the Desmond™ (Schrödinger Release 2017-3).

The RMSF was determined, over 25 ns, for each residue's sidechain in the proteins *apo* conformation and when it is complexed to BSP (Figure 5.5). The RMSF of the protein was used to characterise local changes that occurred along the protein chain considering that peaks represent areas of the protein which fluctuated the most during the simulation. The two highly flexible loop regions of *Pf*GST (in both chain A and B), shown in Figure 5.5 with arrows, displayed a decrease in their RMSF when BSP was bound, in comparison to the *apo* protein conformation. There was also stabilisation of residues 105-118 in chain A due to the ligand contacts that were made with these residues. Interestingly, residues 235-278 (residue numbers 29-72 in chain B), which make up the G-site, showed an increase in their RMSF, when BSP was bound to the protein. In addition, there are slight fluctuations that naturally occur at the N- and C-terminal. This is expected as these regions naturally fluctuate more as they are not bound to other neighboring residues and do not form more structured elements such as alpha helices and beta strand, which are more rigid than the unstructured part of the protein (Schrödinger Release 2017-3).

Protein-ligand contacts occurred during the 25 ns MD simulation. The bar graph (Figure 5.6) represents specific residues found in *Pf*GST and the specific contacts they made with the ligand such as hydrogen bonds (green), hydrophobic bonds (purple) and water bridges (blue). The fraction of interactions of various residues during MD simulations were determined to analyse how the ligand interacted with the protein in the complex during the simulation (Figure 5.6). The binding site residues that made contact with the protein were the same as the binding site residues illustrated in the 2D interaction diagrams (Figure 5.3 B) and were mainly residue number LYS 15, GLN 71, GLN 73, CYS 101, GLN 104, ASP 105, TYR 108 and ASN 111 on chain A and residue numbers LYS 15, GLN 73, CYS 101, GLN 104, ASP105 and TYR108 on chain B.

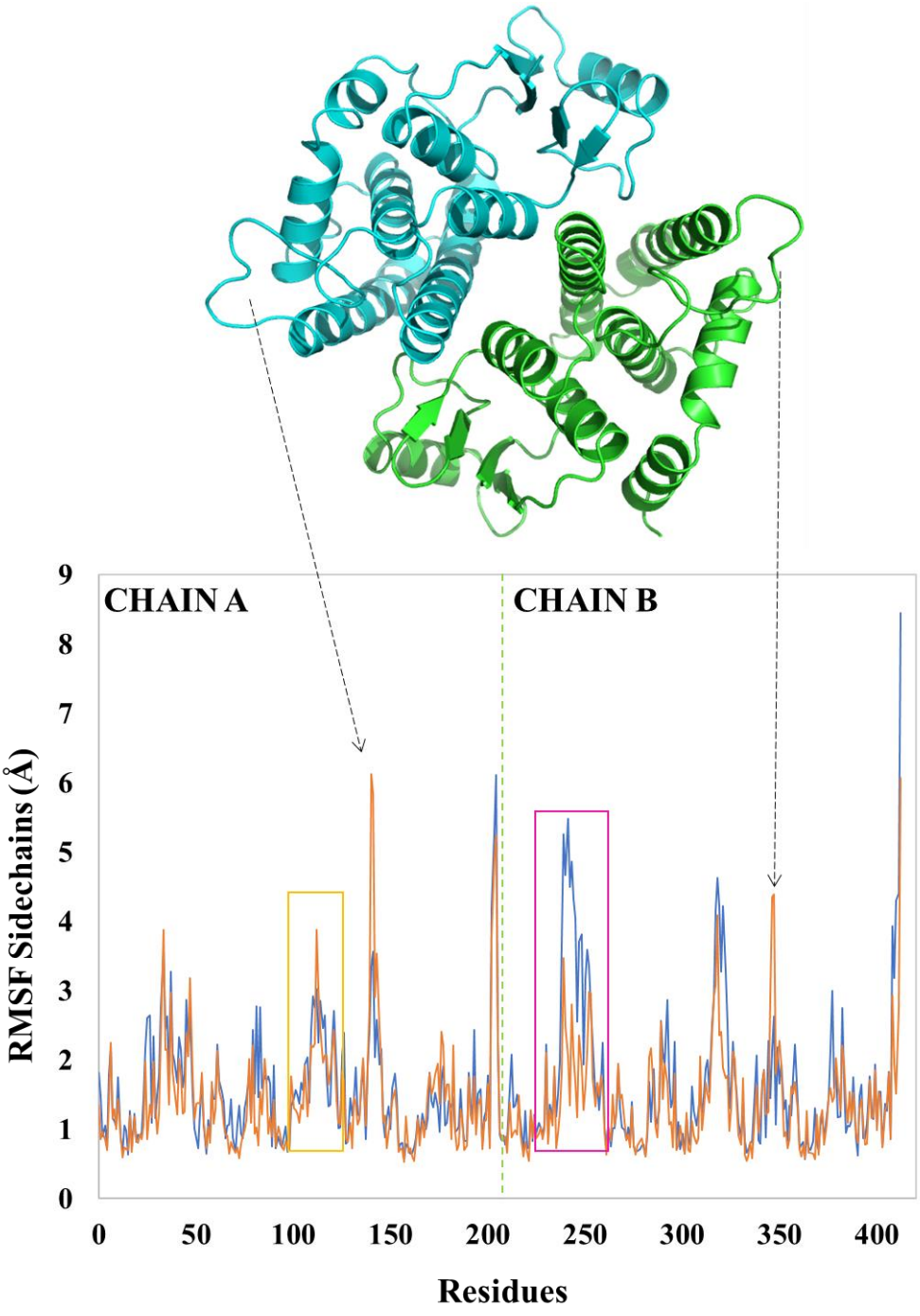


Figure 5.5: RMSF of the residues side chain when the protein is complexed with and without BSP, over a 25 ns simulation. The apo-*Pf*GST and *Pf*GST-BSP complex are represented as orange and blue lines, respectively. The *Pf*GST's structure, which includes chain A (cyan) and B (green), has been used in addition to the RMSF plot. The two highly flexible loop regions of *Pf*GST shown with arrows, displayed a decrease in their RMSF when BSP was bound, in comparison to the *apo* protein conformation. Stabilisation of residues 105-118 in chain A also occurred and is shown with an orange box. An increase in the RMSF occurred at residues 235-278 (chain B) when BSP was bound to the protein and is shown with the pink box. In addition, there are slight fluctuations that naturally occurred at the N- and C-terminal. All the data was extracted from SID™ from the Desmond™ (Schrödinger Release 2017-3).

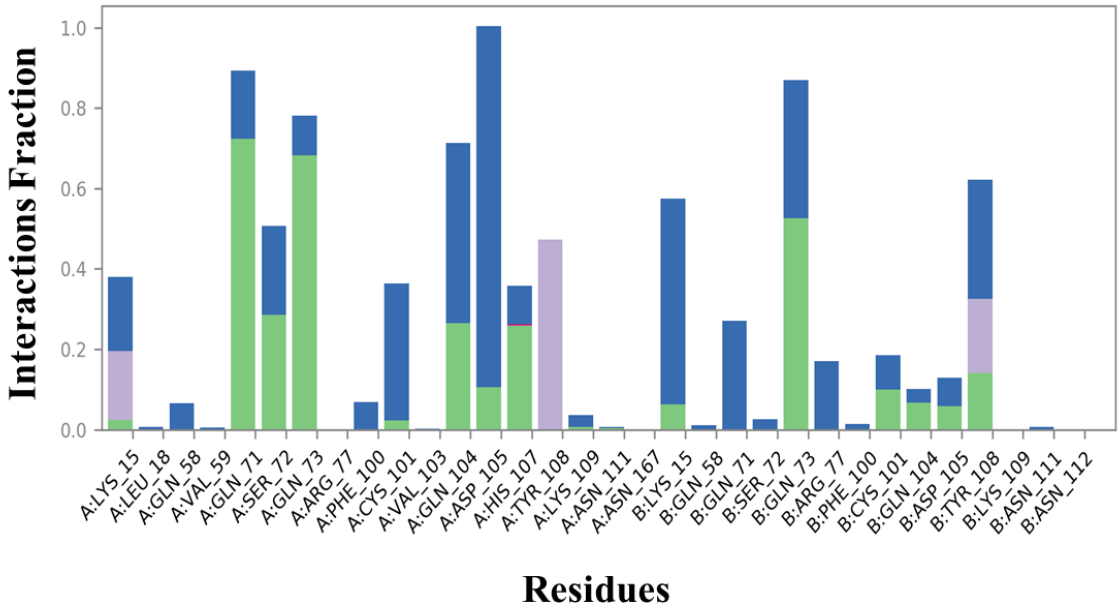


Figure 5.6: Interaction fractions of various residues during molecular dynamic simulations. The protein-ligand interactions are categorised which are mainly hydrogen bonds (green), hydrophobic (purple) and water bridges (blue). The stacked bar charts are normalised over the course of the trajectory, where the interaction fraction suggests the percentage of the simulation time the specific interaction is maintained. All the data was extracted from SID™ from the Desmond™ (Schrödinger Release 2017-3).

5.6 Discussion

For the computational study to be done, the active dimeric *Pf*GST structure was produced through homology modeling, using SWISS-MODEL™. The crystallographic structure 3FRC was identified as the most accurate template sequence for the active dimer, using SWISS-MODEL™. The 3FRC PDB structure, however, had 3 missing loop regions, which would later lead to problems in MD simulation (Sarma, *et al.*, 2003). Therefore, the homology modeling allowed for the completion of the missing loop regions. The homology model produced was used for the docking of BSP. An alignment of the Cα of both the original *Pf*GST model and the homology modelled *Pf*GST, produced a RMSD Cα of 0.20 Å. This suggests that *Pf*GST did not considerably change during the homology modelling process. This was further confirmed by performing a ProCheck (Biasini *et al.*, 2014) analysis (Table 3.1). Any changes in the stereochemical properties before after homology modelling, resulted from the addition of the 3 flexible loops, which could have caused changes in the sidechain of the selected residues and consequently, slight structural changes occurred. These changes were overall negligible, therefore, homology modelled *Pf*GST was considered structurally sound for further use in the molecular modelling study.

The homology-modelled *Pf*GST underwent protein preparation and energy minimisation, therefore, amending structural problems derived from the crystallographic structure, to create a dependable all-atom 3D protein model for IFMD and MD simulations (Schrödinger Release 2017-3 & Sastry, *et al.*, 2013). Crystallographic structures are rigid and have contacts, therefore, they do not represent the dynamic nature of a protein (Sastry, *et al.*, 2013). Protein energy minimisation was carried out to produce a protein that was energetically favourable. This now energetically favourable protein conformation interacted with the BSP, both in induce fit ligand docking and MD simulation.

The docking of BSP and *Pf*GST, using the Induce Fit™ docking protocol, was performed. The docking was scored using Glidescore, Glide van der Waals forces, Glide emodel, Glide energy, Glide energy internal and IFD scores, who's units are all kcal.mol^{-1}. During the preparation of the ligand docking, the protein was incapsulated

in a box (see Section 5.3.2), which included the G-site, the H-site, and the L-site. The program then scanned the protein and docked all the ligands at the L-site of the protein. This could suggest that the strong affinity site observed in ITC (see section 4.4.4) is the L-site and not the H-site. Interestingly, there were slight differences in the scores of the first 3 poses, which may be a result of the ligand binding within the proteins cavity, at the L-site. The ligands movements are limited within the cavity (Fons, 2017), therefore, no major changes in the ligands position were observed (Figure 5.3). The top ranked pose had negative scores for both its Glidescore and its Glide emodel score. This suggested that there was tight binding of BSP to the L-site of the protein. The Glidescore was lower than -10 kcal.mol^{-1}, which suggests good binding of the ligand to the protein (Schrödinger Release 2017-3). Residues within the cavity (L-site) of the protein in which the ligand is binding were charged and polar. There were hydrogen bond interactions between residues LYS 15, GLN 71, SER 72 and GLN 73 from chain A and the oxygen atoms from the SO_3H group of BSP. The interaction between the O group of the SO_3H of the BSP was also observed for chain B's residue LYS 15. In addition, hydrogen bond interactions between chain A's residue ASP 105 and chain B's residue CYS 101 with the OH group of the BSP was observed. There was also π-π stacking between chain A's residue TYR 106 and the BSP'S benzene ring. Therefore, this confirmed that there is binding of the BSP molecule at the L-site of the protein.

However, IFMD still shows a static look at how a ligand interacts and binds to a protein. Therefore, MD simulation studies were performed in order to understand how the *Pf*GST-BSP complex could behave within a physiological environment and how the binding of the BSP could cause a change in how the *apo* conformation of the protein behaves. Therefore, giving a better and more realistic explanation of the results produced from the docking studies (Kakarala & Jamil, 2015). The results of the simulation were examined using SID^{TM}, implemented in Desmond. The RMSD of the proteins' Cα chain, backbone and sidechains were determined for both the apo conformation of the protein and for the protein-ligand complex. The RMSD (Figure 5.4 A-C), which measures the average change in the displacement of a selection of atoms for a precise frame in comparison to the reference frame (Kakarala & Jamil, 2015) for the protein bound to the ligand was determined. Both the apo conformation of the protein and the protein-ligand complex showed fluctuations between 1-3 Å. If

fluctuations at the end of the simulation were greater than 1-3 Å, it would indicate that the protein structure was undergoing a large conformational change during the simulation or that the system did not equilibrate, suggesting that the simulation time period may not have been long enough for a rigorous analysis (Schrödinger Release 2017-3). There were notably more fluctuations observed for the protein-ligand complex in comparison to the apo conformation of the protein. Therefore, the binding of BSP causes an increase in the overall dynamics of *Pf*GST.

The RMSD of the ligand with respect to the protein (Figure 5.4 D) showed fluctuations that remain between 1-3 Å. This suggested that the ligand had not diffused away from its initial binding site during the simulation and that the ligand was stable (Schrödinger Release 2017-3). However, the protein is dynamic due to the increased RMSD observed for the protein with respect to the ligand. The ligand in respect to the ligand, also falling into the 1-3 Å range, shows that there was internal fluctuation of the atoms in the ligand and slight changes in the ligand structure from its reference frame during the simulation (Schrödinger Release 2017-3). This is expected due to the increase in RMSD of *Pf*GST, when BSP is present.

The RMSF of the protein complex to the ligand (Figure 5.5) is used to characterise local changes that occur along the protein chain (Schrödinger Release 2017-3). The two highly flexible loop regions of *Pf*GST, in both chains displayed a decrease in their RMSF when BSP was bound, in comparison to the *apo* protein conformation. This could be due to conformational changes in the protein when the ligand binds and interacts with the backbone of the protein, which causes the loop regions to become less flexible (Zavodszky *et al.,* 2005*)* This was unexpected as both loops are not near the binding site for BSP (Figure 5.2), however, the loop's residues are close to the residues that are interacting with the ligand, therefore, stabilisation of neighbouring residues which are involved in the in the ligand binding event may then lead to decreased fluctuation in close proximity residues. In addition, there was more stabilisation which occur in the loop on chain A in comparison to the loop in chain B. This could be because there are more contacts made in chain A with the BSP, then in chain B. Therefore, by stabilising the neighboring residues, due to the increase contacts of the ligand made with protein, there will be more stabilisation occurring within neighboring amino acids (Raff *et al.,* 2002). There was also stabilisation of

residues 105-118 in chain A. Stabilisation is due to the ligand contacts that are directly made with these residues (Raff *et al.*, 2002). Interestingly, there was an increase in RMSF of the sidechain residues 235-278 (residue numbers 29-72 in chain B) when BSP was bound to the protein, in comparison to the apo conformation of the protein. These residues make up the G-site of the protein and are important for the enzymatic activity of *Pf*GST, as it allows conjugation of GSH to a variety of electrophilic substrates, which produces a less toxic and more hydrophilic product that can be removed from the parasite, therefore, protecting the parasite against cytotoxic and genotoxic compounds. The increased fluctuations of the G-site due to BSP binding to the L-site may be the reason for the inhibition of the enzyme's activity seen experimentally (see Section 4.4.2) as GSH will not be able to bind with a strong affinity to the G-site and carry out its appropriate function.

Protein-ligand contacts occurred during the 25 ns MD simulation. The binding site residues that made contact with the protein were similar to contacts that were illustrated in the 2D interaction diagrams (Figure 5.3 A). Majority of the interaction of the ligand with the cavity were hydrogen bonds and water bridges. The ligand is quite hydrophobic as it has 3 benzene rings, however the O and OH molecules on its SO_3H groups that are attached to the benzene rings, allowed for hydrogen bonding, with a distance of 2.5 Å between the donor and acceptor atoms, to occur between the hydrophobic BSP and the hydrophilic residues in the cavity. Hydrogen bonds play a significant role in drug design due to their influence on specificity, metabolization and absorption of the drug. The cavity is also solvent exposed; therefore, water bridges will form between the protein residues and the polar atoms of the BSP, with a distance of 2.8 Å between the donor and acceptor atoms. The main hydrophobic interaction, specifically a π-π interaction/ stacking, occurred between the aromatic rings of TYR 108, of the protein, and BSP. A study done by Oakely *et al.* (1997) found that Ethacrynic acid (EA), a known inhibitor of alpha, mu, and pi class GSTs (Ploemen *et al.*, 1990), binds to the H-site of GSTP1. The H-site is also the site in which CDNB is known to bind HGSTP1. Studies demonstrated that EA acts as a reversible competitive inhibitor towards CDNB. However, it binds noncompetitive toward GSH (Ploemen *et al.*, 1990; Phillips and Mantle, 1991). EA binds within a hydrophobic pocket which contains TYR 7, PHE 8, PRO 9, AND VAL 10 and the aliphatic portions of ARG 13, VAL 35, ILE 104, AND TYR 108. TYR 108 acts as an

electrophilic participant in the Michael addition (nucleophilic addition to an α,β-unsaturated carbonyl compound) of glutathione to ethacrynic acid (Oakley *et al.*, 1997). EA-GSH complex's aromatic ring, which optimises the packing contribution by sitting at a less optimum angle, shows direct parallel π-π stacking with the aromatic ring of TYR 108 (Oakley *et al.*, 1997). Therefore, the π-π stacking between the inhibitor and the TYR 108 is critical and plays a role in the binding of the EA to GSH, which in turn inhibits the activity of GSTP1. Following this, it is hypothesized that the TYR 108 plays an important role in the inhibition of *Pf*GST enzyme activity (Section 4.3.1) with BSP, through the π-π stacking interaction.

Chapter 6

Overall discussion, conclusion, and future studies

Malaria can be considered as an emerging disease due to the increased resistance to antimalarials and growing geographical distribution of *Plasmodium falciparum*, therefore, posing a threat to individual's health and welfare (Deponte and Becker, 2005). Due limitations faced by current drugs for malaria which include; drug resistance in both first- and second-line of drug therapy, low efficacy, safety issues, poor compliance and the high cost of artemisinin-based combination therapy (ACT), there is a need to develop new drugs that are directed to novel targets (Deponte and Becker, 2005). These novel drugs, along with their targets, can be used in combination with existing antimalarial therapies. *Pf*GST has become a potential novel target due to its solvent exposed H-site, allowing for the binding of amphiphilic compounds, which differs to its human subclass counterparts (Fritz-Wolf *et al.*, 2003), along with its increased activity observed in chloroquine-resistant *Plasmodium falciparum* parasites (Dubois *et al.*, 1995; Srivastava *et al.*, 1999). *Pf*GST has also been found at comparable levels to GAPDH, a known housekeeping protein in the parasite, suggesting that *Pf*GST is an important enzyme during the blood stage of the parasite. Therefore, the loss of the *Pf*GST, could cause a build-up of cytotoxic species in the parasite, like free FP IX, which is a result of the parasite's digestion of hemoglobin in the blood stage. This ultimately leads to the elimination of the parasite. A study done by Kolobe *et al*. (2004), showed that BSP was able to interact and inhibit hGSTA1-1. However, in this study, BSP was investigate as a potential inhibitor of *Pf*GST.

Firstly, within any protein biochemistry study, the rate limiting or bottle neck step in the study is the expression and purification of a protein. The *Pf*GST vector was synthesised with the addition of a non-cleavable 6×his-tag. Both expression and purification of the protein was successful, yielding a high concentration of purified protein. This is ideal for both structural and functional studies of the protein. The purified *Pf*GST was also shown to be active, using GSH-CDNB conjugation enzyme activity assay (Habig *et al.*, 1974). In addition, in the presence of BSP, the activity of the protein decreased, proving that the BSP binds and inhibits *Pf*GST. One limitation

experienced here is the GSH-CDNB assay's sensitivity, which was compensated by using higher *Pf*GST concentrations (Sayed, 2001). This affected downstream experiments, like the IC_{50}, which could not be compared to studies done by Kolobe *et al*, (2004). Therefore, future studies could include using more sensitive enzyme assays, allowing for comparable concentrations of *Pf*GST to be used or doing a comparative study, based on Kolobe *et al*, (2004) study, however with comparable concentrations used in this study.

The protein-ligand interaction was evaluated using, extrinsic ANS fluorescence. The fluorescence intensity decreased significantly when *Pf*GST complexed with BSP, in comparison to the *apo-Pf*GST. This could be a result of BSP displacing ANS from the hydrophobic patch/pocket, Therefore, decreasing the fluorescent intensity.

ITC was done to further understand the binding that occurred between *Pf*GST and BSP. A sequential two-site model was shown to be the best model for *Pf*GST complexed to BSP in the presence GSH (Figure 4.5), suggesting *Pf*GST has two types of co-operative and/or allosteric binding sites for BSP. This correlated with Kolobe *et al*. (2004) study on hGSTA1-1, which showed a high-affinity and a low-affinity binding site for BSP. The dimeric *Pf*GST's first binding site had a low affinity for BSP. Due to the lack of a crystal structure for *Pf*GST bound to BSP, the binding sites were inferred using other studies and literature. The low affinity site was proposed to be BSP binding to the H-site and the high affinity site was proposed to be BSP binding in the L-site, which was supported by docking studies that were done. BSP binding to the H-site, a hydrophobic pocket, could be causing the displacement of the ANS by BSP, therefore causing an overall decrease in fluorescent intensity observed. Therefore, future studies could include surface plasmon resonance (SPR) and hydrogen deuterium exchange mass spectrometry (HDX-MS).

Due to the lack of a crystal structure of *Pf*GST bound to BSP, crystallography trials were set up. A crystal of the complex was diffracted, however due to collection and data limitations, the structure could not be solved. In future, known crystallization conditions identified in this study can be used in order to produce larger crystals, by using hanging drop technique. This technique allows for bigger drops to be dispended, forming larger crystals which may aid in the diffraction quality of the

crystal. With an increase in the quality and size of the crystal, more data could be collected during diffraction, therefore, producing a solved structure.

The *in-silico* study of *Pf*GST and its interactions with BSP, using IFMD and MD simulations, showed that the BSP ligand interacted with the L-site of *Pf*GST. The docking revealed that the π-π stacking between the aromatic rings of PfGST and BSP could play a very important role in the overall inhibition of *Pf*GST as a study done by Oakely *et al.* (1997) showed that TYR 108 is critical and plays a role in the binding of the EA to GSH, which in turn inhibits the activity of GSTP1. There was also a definite increase in the dynamics of *Pf*GST, when BSP was bound in comparison to the apo-*Pf*GST conformation observed in the study. The dynamics could be a result of the increased fluctuations which occurred in residues which make up the G-site of *Pf*GST. The increased fluctuations of the G-site due to BSP binding to the L-site may be the reason for the inhibition of the enzyme's activity seen experimentally (see section 4.3.1) as GSH will not be able to bind with a strong affinity to the G-site and carry out its appropriate function. Therefore, future studies by looking at the interaction of BSP at the H-site of *Pf*GST and how it compares to the L-site interactions. Also, the addition of GSH at the G-site could be done in order to determine how this affects the *Pf*GST-BSP complex at the H- and L-site, during the simulation.

Following this study, showing the interaction and inhibition of *Pf*GST with BSP, novel inhibitor's could be developed based on the BSP molecule. By using the BSP molecule as a scaffold, chemical modifications of the structure could allow for the targeting of *Pf*GST, and not human isoforms of the enzyme. Screening of the novel inhibitors, prior to laboratory testing, could be performed using molecular modeling and dynamic simulations. This allows for the could allow for the cytotoxic effects in the parasite and the inhibition of hemozoin , leading to deleterious effects and elimination of the parasites. Therefore this provides a new therapeutic option in the fight against malaria.

References

Autino, B., Noris, A., Russo, R., Castelli, F. (2012). Epidemiology of malaria in endemic areas. *Mediterranean journal of hematology and infectious diseases* **4** (1).

Alexander, B., Rogers, C. and Naftalin, R. (2002). Hepatic arterial perfusion decreases intrahepatic shunting and maintains glucose uptake in the rat liver. *Pflügers Archiv*, **444** (1-2), 291-298.

Axarli, I.A., Rigden, D.J. and Labrou, N.E. (2004). Characterization of the ligandin site of maize glutathione *S*-transferase I. *Biochemical Journal*, **382** (3), 885-893.

Barillas-Mury, C., Kumar, S. (2005). *Plasmodium*–mosquito interactions: a tale of dangerous liaisons. *Cellular microbiology* **7** (11), 1539-1545.

Baron, S. (1996). Alphaviruses (Togaviridae) and Flaviviruses (Flaviviridae)--Medical Microbiology. University of Texas Medical Branch at Galveston.

Barratt, E., Bronowska, A., Vondrášek, J., Černý, J., Bingham, R., Phillips, S., Homans, S. W. (2006). Thermodynamic penalty arising from burial of a ligand polar group within a hydrophobic pocket of a protein receptor. *Journal of molecular biology* **362** (5), 994-1003.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T. G., Bertoni, M., Bordoli, L. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic acids research* **42** (W1), W252-W258.

Bisswanger, H. (2014). Enzyme assays. *Perspectives in Science* **1** (1-6), 41-55.

Bradley, S.E., Ingelfinger, F.J., Bradley, G.P. and Curry, J.J. (1945). The estimation of hepatic blood flow in man. *The Journal of clinical investigation*, **24**(6), 890-897.

Bronowska, A. K., (2011). Thermodynamics of ligand-protein interactions: implications for molecular design, *Thermodynamics-Interaction Studies-Solids, Liquids and Gases*. IntechOpen.

Brooke, B., Koekemoer, I., Kruger, P., Urbach, J., Misiani, E., Coetzee, M. (2013). Malaria vector control in South Africa. *South African medical journal* **103** (10), 784-788.

Bruns, C. M., Hubatsch, I., RidderstroÈm, M., Mannervik, B., Tainer, J. A. (1999). *Human* glutathione transferase A4-4 crystal structures and mutagenesis reveal the basis of high catalytic efficiency with toxic lipid peroxidation products. *Journal of molecular biology* **288** (3), 427-439.

Bunik, V. I. (2003). 2-Oxo acid dehydrogenase complexes in redox regulation: Role of the lipoate residues and thioredoxin. *European Journal of Biochemistry* **270** (6), 1036-1042.

Chasseaud, L. (1979). The role of glutathione and glutathione S-transferases in the metabolism of chemical carcinogens and other electrophilic agents, *Advances in cancer research*. Elsevier, 175-274.

Coetzee, M., Kruger, P., Hunt, R., Durrheim, D., Urbach, J., Hansford, C. (2013). Malaria in South Africa: 110 years of learning to control the disease. *South African medical journal* **103** (10), 770-778.

Conrad, M.D. and Rosenthal, P.J. (2019). Antimalarial drug resistance in Africa: the calm before the storm?. *The Lancet Infectious Diseases*, **19**(10), 338-351.

Cox, F. E. (2010). History of the discovery of the malaria parasites and their vectors. *Parasites & vectors* **3** (1), 5.

Cutler, H.G. and Cutler, S.J. (1999). *Biologically active natural products: pharmaceuticals*. CRC Press.

Daubenberger, C.A., Pöltl-Frank, F., Jiang, G., Lipp, J., Certa, U. and Pluschke, G. (2000). Identification and recombinant expression of glyceraldehyde-3-phosphate dehydrogenase of *Plasmodium falciparum*. *Gene*, **246(**1-2), 255-264.
De, S., Girigoswami, A. and Das, S. (2005). Fluorescence probing of albumin–surfactant interaction. *Journal of colloid and interface science*, **285** (2), 562-573.
Delves, M., Plouffe, D., Scheurer, C., Meister, S., Wittlin, S., Winzeler, E. A., Sinden, R. E., Leroy, D. (2012). The activities of current antimalarial drugs on the life cycle stages of *Plasmodium*: a comparative study with human and rodent parasites. *PLoS medicine* **9** (2).
Deponte, M., Becker, K. (2005). Glutathione S-transferase from malarial parasites: structural and functional aspects. *Methods in enzymology* **401**, 241-253.
Dubois, V. L., Platel, D. F., Pauly, G., Tribouleyduret, J. (1995). *Plasmodium berghei*: implication of intracellular glutathione and its related enzyme in chloroquine resistance in vivo. *Experimental parasitology* **81** (1), 117-124.
Famin, O., Krugliak, M., Ginsburg, H. (1999). Kinetics of inhibition of glutathione-mediated degradation of ferriprotoporphyrin IX by antimalarial drugs. *Biochemical pharmacology* **58** (1), 59-68.
Farid, R., Day, T., Friesner, R. A., Pearlstein, R. A. (2006). New insights about HERG blockade obtained from protein modeling, potential energy mapping, and docking studies. *Bioorganic & medicinal chemistry* **14** (9), 3160-3173.
Fischer, S. and Verma, C.S. (1999). Binding of buried structural water increases the flexibility of proteins. *Proceedings of the National Academy of Sciences*, **96** (17), 9613-9615.
Ferguson, H. M., Read, A. F. (2004). Mosquito appetite for blood is stimulated by Plasmodium chabaudi infections in themselves and their vertebrate hosts. *Malaria journal* **3** (1), 12.
Florens, L., Washburn, M. P., Raine, J. D., Anthony, R. M., Grainger, M., Haynes, J. D., Moch, J. K., Muster, N., Sacci, J. B., Tabb, D. L. (2002). A proteomic view of the *Plasmodium falciparum* life cycle. *nature* **419** (6906), 520-526.
Fons N. (2017). Textbook of Drug Design and Discovery, Fifth Edition. *The Yale Journal of Biology and Medicine*, **90** (1), 160.
Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K. (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of medicinal chemistry* **47** (7), 1739-1749.
Fritz-Wolf, K., Becker, A., Rahlfs, S., Harwaldt, P., Schirmer, R. H., Kabsch, W., Becker, K. (2003). X-ray structure of glutathione S-transferase from the malarial parasite *Plasmodium falciparum*. *Proceedings of the National Academy of Sciences* **100** (24), 13821-13826.
Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server, *The proteomics protocols handbook*. Springer, 571-607.
Ghisaidoobe, A. B., Chung, S. J. (2014). Intrinsic tryptophan fluorescence in the detection and analysis of proteins: a focus on Förster resonance energy transfer techniques. *International journal of molecular sciences* **15** (12), 22518-22538.
Glasel, J. (1995). Validity of nucleic acid purities monitored by 260nm/280nm absorbance ratios. *Biotechniques* **18** (1), 62-63.
González, M. (2011). Force fields and molecular dynamics simulations. *École thématique de la Société Française de la Neutronique* **12**, 169-200.

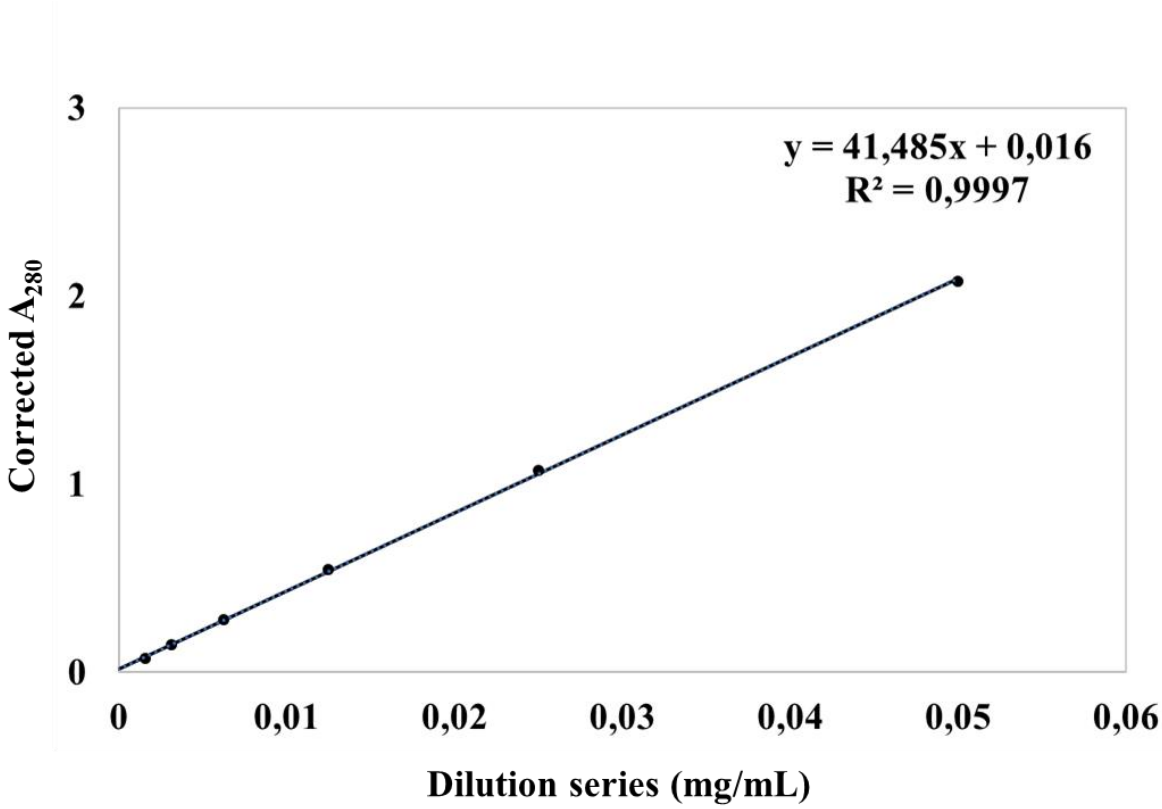
Habig, W. H., Pabst, M. J., Jakoby, W. B. (1974). Glutathione S-transferases the first enzymatic step in mercapturic acid formation. *Journal of Biological Chemistry* **249** (22), 7130-7139.
Harwaldt, P., Rahlfs, S., Becker, K. (2002). Glutathione S-transferase of the malarial parasite *Plasmodium falciparum*: characterization of a potential drug target. *Biological chemistry* **383** (5), 821-830.
Hawe, A., Sutter, M., Jiskoot, W. (2008). Extrinsic fluorescent dyes as tools for protein characterization. *Pharmaceutical research* **25** (7), 1487-1499.
Hiller, N., Fritz-Wolf, K., Deponte, M., Wende, W., Zimmermann, H. and Becker, K. (2006). Plasmodium falciparum glutathione S-transferase—Structural and mechanistic studies on ligand binding and enzyme inhibition.*Protein science*, **15**(2), 281-289.
Hu, H., Elstner, M., Hermans, J. (2003). Comparison of a QM/MM force field and molecular mechanics force fields in simulations of alanine and glycine "dipeptides" in water in relation to the problem of modeling the unfolded peptide backbone in solution. *Proteins: Structure, Function, and Bioinformatics* **50** (3), 451-463.
Imlay, J. A. (2003). Pathways of oxidative damage. *Annual Reviews in Microbiology* **57** (1), 395-418.
Jayarama-Naidu, R., Johannes, J., Meyer, F., Wirth, E.K., Schomburg, L., Köhrle, J. and Renko, K. (2015). A nonradioactive uptake assay for rapid analysis of thyroid hormone transporter function. *Endocrinology*, **156**(7), 2739-2745.
Ji, X., Zhang, P., Armstrong, R. N., Gilliland, G. L. (1992). The three-dimensional structure of a glutathione S-transferase from the Mu gene class. Structural analysis of the binary complex of isoenzyme 3-3 and glutathione at 2.2-. ANG. resolution. *Biochemistry* **31** (42), 10169-10184.
Jorgensen, W. L. (2004). The many roles of computation in drug discovery. *Science* **303** (5665), 1813-1818.
Kar, S., Kar, S. (2010). Control of malaria. *Nature Reviews: Drug Discovery* **9**, 511-512.
Kolobe, D., Sayed, Y., Dirr, H. W. (2004). Characterization of bromosulphophthalein binding to *human* glutathione S-transferase A1-1: thermodynamics and inhibition kinetics. *Biochemical Journal* **382** (2), 703-709.
Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *nature* **227** (5259), 680.
Le Roch, K.G., Johnson, J.R., Florens, L., Zhou, Y., Santrosyan, A., Grainger, M., Yan, S.F., Williamson, K.C., Holder, A.A., Carucci, D.J. and Yates, J.R. (2004). Global analysis of transcript and protein levels across the *Plasmodium falciparum* life cycle. *Genome research* **14**(11), 2308-2318.
Liebau, E., Bergmann, B., Campbell, A. M., Teesdale-Spittle, P., Brophy, P. M., Luersen, K., Walter, R. D. (2002). The glutathione *S*-transferase from *Plasmodium falciparum*. *Molecular and biochemical parasitology* **124** (1), 85-90.
Loria, P., Miller, S., Foley, M., TIlley, L. (1999). Inhibition of the peroxidative degradation of haem as the basis of action of chloroquine and other quinoline antimalarials. *Biochemical Journal* **339** (2), 363-370.
Maharaj, R., Seocharan, I., Qwabe, B., Mkhabela, M., Kissoon, S., Lakan, V. (2019). Decadal epidemiology of malaria in KwaZulu-Natal, a province in South Africa targeting elimination. *Malaria journal* **18** (1), 368.
Maurer, H. (1971). Methods of analytical disc electrophoresis. *Disc Electrophoresis and Related Techniques of Polyacrylamide Gel Electrophoresis*, 32-82.

McTigue, M.A., Williams, D.R. and Tainer, J.A. (1995). Crystal Structures of a Schistosomal Drug and Vaccine Target: Glutathione *S*-transferase from Schistosoma japonica and its Complex with the Leading Antischistomal Drug Praziquantel. *Journal of Molecular Biology* **246** (1), 21-27.
Mogensen, J.E., Wimmer, R., Larsen, J.N., Spangfort, M.D. and Otzen, D.E. (2002). The major birch allergen, Bet v 1, shows affinity for a broad spectrum of physiological ligands. *Journal of Biological Chemistry*, **277** (26), 23684-23692.
Möller, M., Denicola, A. (2002). Study of protein-ligand binding by fluorescence. *Biochemistry and Molecular Biology Education* **30** (5), 309-312.
Müller, S. (2004). Redox and antioxidant systems of the malaria parasite Plasmodium falciparum. *Molecular microbiology* **53** (5), 1291-1305
Neubig, R. R., Spedding, M., Kenakin, T., Christopoulos, A. (2003). International Union of Pharmacology Committee on Receptor Nomenclature and Drug Classification. XXXVIII. Update on terms and symbols in quantitative pharmacology. *Pharmacological Reviews* **55** (4), 597-606.
O'Brien, R., Ladbury, J. E., Chowdry, B. Z., (2002). Isothermal titration calorimetry of biomolecules. Oxford University Press, Oxford.
Oakley, A. J., Rossjohn, J., Lo Bello, M., Caccuri, A. M., Federici, G., Parker, M. W. (1997). The three-dimensional structure of the *human* Pi class glutathione transferase P1-1 in complex with the inhibitor ethacrynic acid and its glutathione conjugate. *Biochemistry* **36** (3), 576-585.
Palermo, G., De Vivo, M. (2014). Computational Chemistry for Drug Discovery. *Encyclopedia of nanotechnology*, 1-15.
Porath, J., Olin, B. (1983). Immobilized metal affinity adsorption and immobilized metal affinity chromatography of biomaterials. Serum protein affinities for gel-immobilized iron and nickel ions. *Biochemistry* **22** (7), 1621-1630.
Perbandt, M., Burmeister, C., Walter, R.D., Betzel, C. and Liebau, E. (2004). Native and inhibited structure of a Mu class-related glutathione *S*-transferase from *Plasmodium falciparum*. *Journal of Biological Chemistry*, **279**(2), 1336-1342.
Perbandt, M., Eberle, R., Fischer-Riepe, L., Cang, H., Liebau, E., Betzel, C. (2015). High resolution structures of *Plasmodium falciparum* GST complexes provide novel insights into the dimer–tetramer transition and a novel ligand-binding site. *Journal of structural biology* **191** (3), 365-375.
Raff, M., Alberts, B., Lewis, J., Johnson, A. and Roberts, K. (2002). Molecular Biology of the Cell 4th edition.
Raman, J., Morris, N., Frean, J., Brooke, B., Blumberg, L., Kruger, P., Mabusa, A., Raswiswi, E., Shandukani, B., Misani, E. (2016). Reviewing South Africa's malaria elimination strategy (2012–2018): progress, challenges and priorities. *Malaria journal* **15** (1), 438.
Rosenthal, S., White, E. (1925). *The Journal of the American Medical Association* **84**, 1112.
Rojas, M. O., Wasserman, M. (1993). Effect of low temperature on the in vitro growth of *Plasmodium falciparum*. *Journal of Eukaryotic Microbiology* **40** (2), 149-152.
Salinas, A. E., Wong, M. G. (1999). Glutathione S-transferases-a review. *Current medicinal chemistry* **6** (4), 279-310.
Salina, N.D., Tang, W.K. and Tolia, N.H. (2019). Blood-Stage Malaria Parasite Antigens: Structure, Function, and Vaccine Potential. *Journal of molecular biology* **431** (21), 4259-4280

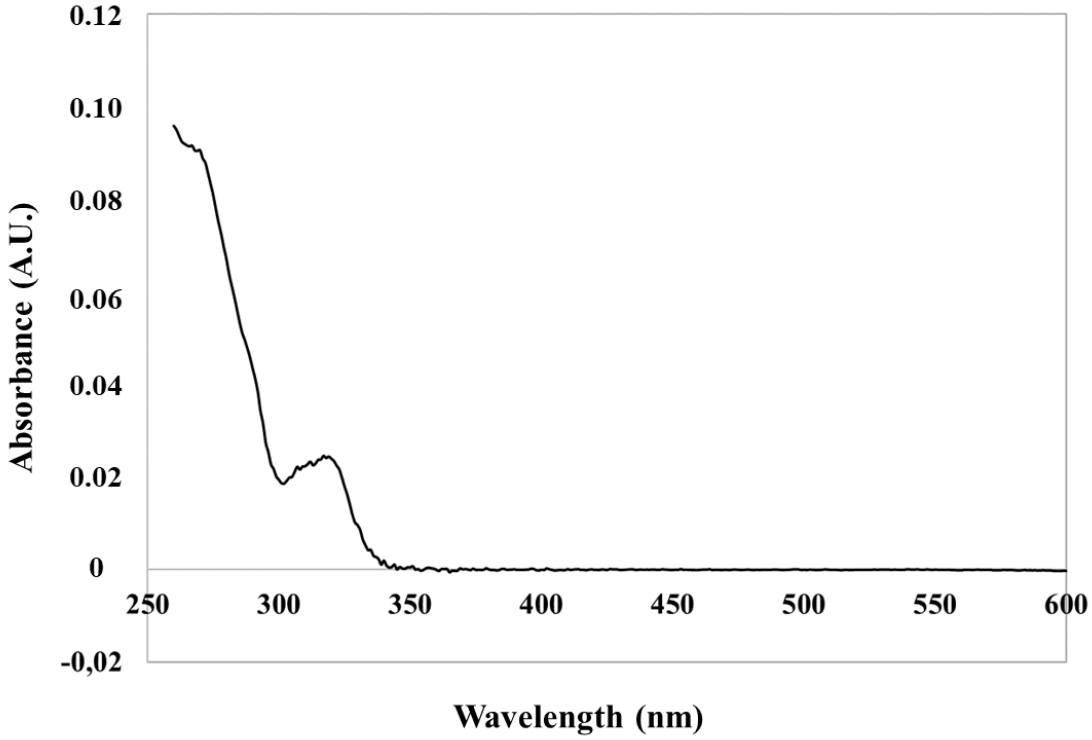
Sangsritavong, S., Combs, D.K., Sartori, R., Armentano, L.E. and Wiltbank, M.C. (2002). High feed intake increases liver blood flow and metabolism of progesterone and estradiol-17β in dairy cattle. *Journal of dairy science*, **85** (11), 2831-2842.
Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R., Sherman, W. (2013). Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of computer-aided molecular design* **27** (3), 221-234.
Sarma, G., Savvides, S., Becker, K., Schirmer, M., Schirmer, R., Karplus, P. (2003). Glutathione reductase of the malarial parasite *Plasmodium falciparum*: crystal structure and inhibitor development. *Journal of molecular biology* **328** (4), 893-907.
Savvides, S. N., Karplus, P. A. (1996). Kinetics and crystallographic analysis of human glutathione reductase in complex with a xanthene inhibitor. *Journal of Biological Chemistry* **271** (14), 8101-8107.
Sayed, Y. (2001). Biochemical and thermodynamic characterisation of ligand-binding to class alpha glutathione transferase A1-1.
Schirmer, R. H., Müller, J. G., Krauth-Siegel, R. L. (1995). Disulfide-reductase inhibitors as chemotherapeutic agents: the design of drugs for trypanosomiasis and malaria. *Angewandte Chemie International Edition in English* **34** (2), 141-154.
Schantz-Dunn, J., Nour, N. M. (2009). Malaria and pregnancy: a global health perspective. *Reviews in obstetrics and gynecology* **2** (3), 186.
Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A., Farid, R. (2006). Novel procedure for modeling ligand/receptor induced fit effects. *Journal of medicinal chemistry* **49** (2), 534-553.
Schrödinger Release 2017-3: Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2017. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2017.
Schrödinger Suite 2017-3: Induced Fit Docking protocol; Glide, Schrödinger, LLC, New York, NY, 2017; Prime, Schrödinger, LLC, New York, NY, 2017.
Schrödinger Release 2017-3: MacroModel, Schrödinger, LLC, New York, NY, 2017.
Schrödinger Suite 2017-3: Protein Preparation Wizard; Epik, Schrödinger, LLC, New York, NY, 2017; Impact, Schrödinger, LLC, New York, NY, 2017; Prime, Schrödinger, LLC, New York, NY, 2017.
Sinning, I., Kleywegt, G. J., Cowan, S. W., Reinemer, P., Dirr, H. W., Huber, R., Gilliland, G. L., Armstrong, R. N., Ji, X., Board, P. G. (1993). Structure determination and refinement of human alpha class glutathione transferase A1-1, and a comparison with the Mu and Pi class enzymes. *Journal of molecular biology* **232** (1), 192-212.
Sluis-Cremer, N., Wallace, L., Burke, J., Stevens, J. and Dirr, H. (1998) Aflatoxin B1 and sulphobromophthalein binding to the dimeric *human* glutathione S-transferase A1-1: a fluorescence spectroscopic analysis. *European Journal of Biochemistry*. **257**, 434–442
Srivastava, P., Puri, S., Kamboj, K., Pandey, V. (1999). Glutathione-S-transferase activity in malarial parasites. *Tropical Medicine & International Health* **4** (4), 251-254.
Steiner, C., Sen, S., Stange, J., Williams, R. and Jalan, R. (2004). Binding of bilirubin and bromosulphthalein to albumin: implications for understanding the pathophysiology of liver failure and its management. *Liver transplantation*, **10** (12), 1531-1538.
Stryer, L. (1965). The interaction of a naphthalene dye with apomyoglobin and apohemoglobin: a fluorescent probe of non-polar binding sites. *Journal of molecular biology* **13** (2), 482-495.

Tripathi, T., Rahlfs, S., Becker, K., Bhakuni, V. (2007). Glutathione mediated regulation of oligomeric structure and functional activity of Plasmodium falciparum glutathione S-transferase. *BMC structural biology* **7** (1), 67.
Ukpe, I. S., Moonasar, D., Raman, J., Barnes, K., Baker, L., Blumberg, L. (2013). Case management of malaria: treatment and chemoprophylaxis. *South African medical journal* **103** (10), 793-798.
Valenzuela-Chavira, I., Contreras-Vergara, C. A., Arvizu-Flores, A. A., Serrano-Posada, H., Lopez-Zavala, A. A., García-Orozco, K. D., Hernandez-Paredes, J., Rudiño-Piñera, E., Stojanoff, V., Sotelo-Mundo, R. R. (2017). Insights into ligand binding to a glutathione S-transferase from mango: Structure, thermodynamics and kinetics. *Biochimie* **135**, 35-45.
Wagner, E. K., Hewlett, M. J., Bloom, D. C., Camerini, D. (2007). Basic Virology. Wiley-Blackwell, New Jersey.
Weber, G., Young, L. B. (1964). Fragmentation of bovine serum albumin by pepsin I. The origin of the acid expansion of the albumin molecule. *Journal of Biological Chemistry* **239** (5), 1415-1423.
Weiss, D. J., Lucas, T. C., Nguyen, M., Nandi, A. K., Bisanzio, D., Battle, K. E., Cameron, E., Twohig, K. A., Pfeffer, D. A., Rozier, J. A. (2019). Mapping the global prevalence, incidence, and mortality of Plasmodium falciparum, 2000–17: a spatial and temporal modelling study. *The Lancet* **394** (10195), 322-331.
Wells, T. N., Alonso, P. L., Gutteridge, W. E. (2009). New medicines to improve control and contribute to the eradication of malaria. *Nature reviews Drug discovery* **8** (11), 879-891.
Wilson, K., Walker, J. (2000). Principles and techniques of practical biochemistry. Cambridge University Press.
Wlodawer, A., Minor, W., Dauter, Z. and Jaskolski, M. (2008). Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *The FEBS journal*, **275** (1), 1-21.
World Malaria Report 2018. Geneva: World Health Organization; 2018. Licence: CC BY-NC-SA 3.0 IGO.
Yadav, D.K., Kumar, S., Teli, M.K., Yadav, R. and Chaudhary, S. (2019). Molecular targets for malarial chemotherapy: a review. *Current topics in medicinal chemistry*, **19**(10), 861-873.
Young, D. (2004). Computational chemistry: a practical guide for applying techniques to real world problems. John Wiley & Sons.
Zavodszky, M.I. and Kuhn, L.A., 2005. Side-chain flexibility in protein–ligand binding: the minimal rotation hypothesis. *Protein Science*, ***14*** (4), 1104-1114.
Zhang, J., Krugliak, M., Ginsburg, H. (1999). The fate of ferriprotorphyrin IX in malaria infected erythrocytes in conjunction with the mode of action of antimalarial drugs. *Molecular and biochemical parasitology* **99** (1), 129-141.

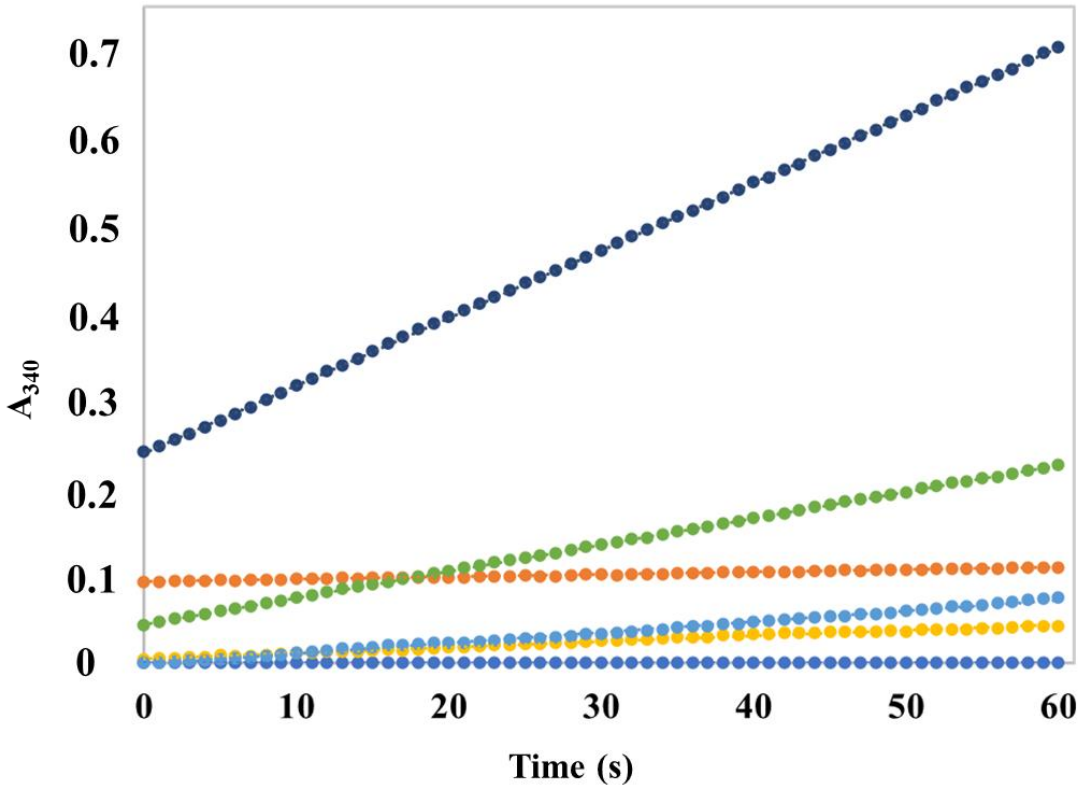
Supplementary



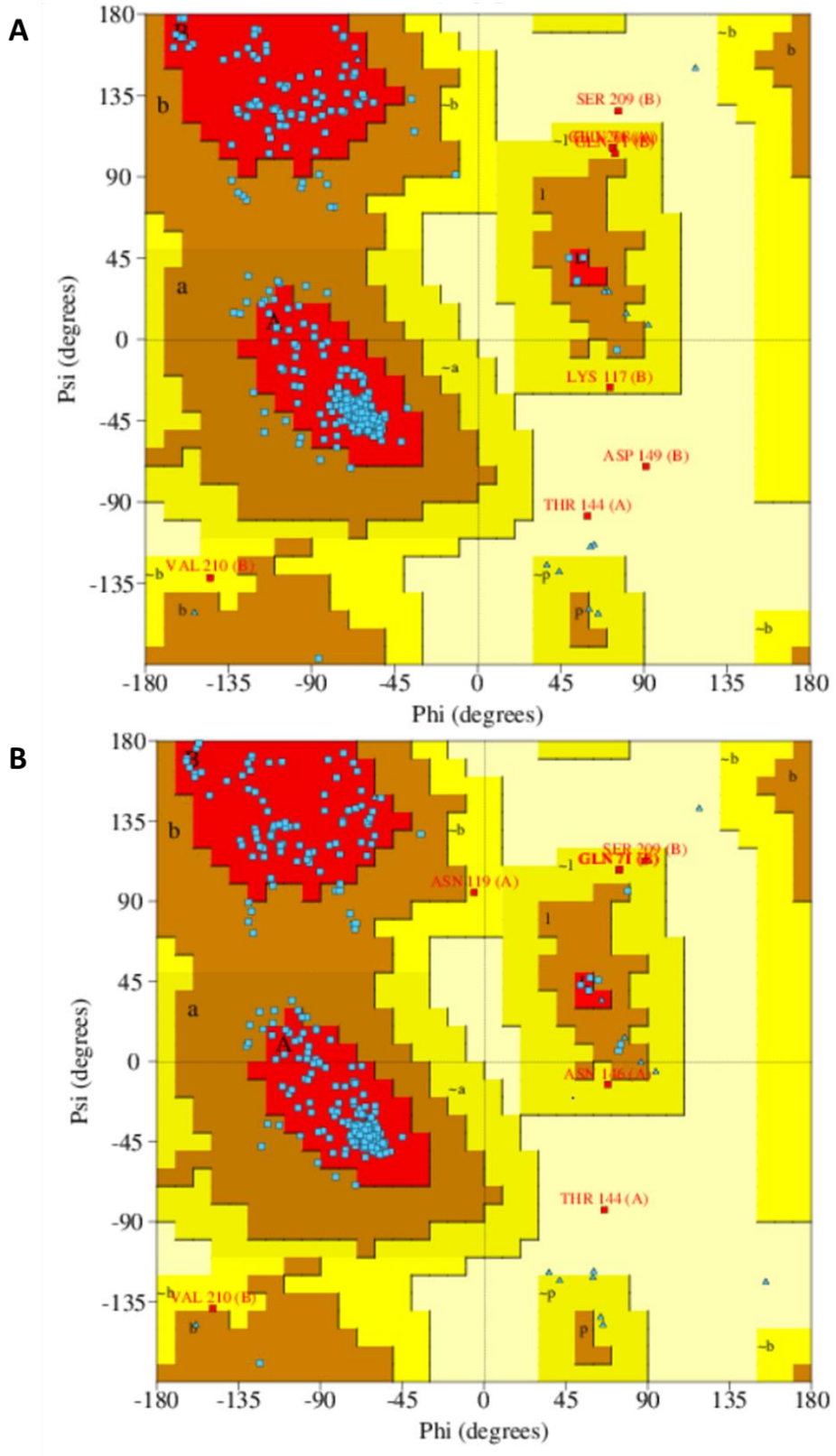
Supplementary Figure 1: Determination of protein concentration using a dilution series. 3 absorbance readings were taken at 260 nm, 280 nm, and 340 nm for a doubling dilution series of the recombinant protein. In order to account for any aggregates, the absorbance reading at 340 nm was subtracted from absorbance reading at 280 nm. These absorbance readings were then plotted against the dilutions and the concentration was determined



Supplementary Figure 2: UV-vis spectroscopy to evaluate the inner filter effect of BSP. A UV-Vis absorbance spectrum (260-600) was done for BSP. There was absorbance experienced for BSP between 260-350 nm, however, BSP had little to no absorbance between 350-600.



Supplementary Figure 3: Progress curves for the different *Pf*GST concentrations. The different concentrations are 0 (dark blue), 27 (orange), 108 (yellow), 216 (light blue), 432 (green) and 864 nM (purple). Saturation is still not reached at 864 nM.



Supplementary Figure 4: Ramachandran plots of *Pf*GST before and after homology modelling. This image was created using EMBL-EBI database's Procheck analysis.