Kinetic analysis of a recombinantly expressed *Plasmodium falciparum* dihydrofolate synthase-folylpolyglutamate synthase

By

Esmaré Human

Submitted in partial fulfillment of the requirements for the
degree *Magister Scientiae*
In the Faculty of Natural and Agricultural Sciences
Department of Biochemistry
University of Pretoria
Pretoria

November 2007

I declare that the thesis/dissertation, which I hereby submit for the degree MSc. Biochemistry at The University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

Signature:…………………….
Date:………………………….

AKNOWLEDGEMENTS

To Professor Louw, my promoter for all his advice and involvement throughout this project. For providing me with opportunities to further my development as a scientist.

To Dr Birkholtz, my co-promoter for all her support, encouragement, advice and mostly for having faith in my abilities. For helping me grow not only as a scientist but also as an individual.

To Professor Hyde and Dr Paul Sims, from the University of Manchester, for their warm welcome during may stay in Manchester. Their valuable opinions and enthusiasm with this project. Special thanks to Sarah Mitchell, Niroshni Nirmalan and the rest of the research group for making my stay in Manchester memorable.

To all the staff, lecturers and students at UP, especially to Katherine Clark for her critical reading and suggestions towards writing the dissertation.

To my parents who made it possible for me to pursue my dreams, for their unconditional love and support. To my friends for their interest and advice throughout.

To our Father in Heaven for entrusting me with this talent, **Philippians 4:13:** "I can do all things through him who strengthens me". For His love and support He bestow on us everyday, **John 3:16:** "For God so loved the world that he gave his only Son, so that everyone who believes in him may not perish but may have eternal life".

Index

Chapter 1: Introduction

Chapter 2: Heterologous expression of recombinant native and synthetic *P. falciparum* DHFS-FPGS

Chapter 3: Activity and kinetic properties of the *P. falciparum* DHFS-FPGS

Chapter 4: Concluding discussion

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

LIST OF TABLES

ABBREVIATIONS

A. gambiae	*Anopheles gambiae*
AMA1	Apical membrane antigen 1
AN	Aggregation number
ATP	Adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
bp	Base pair
BSA	Bovine serum albumin
C1	One carbon units
CCMB	Calcium/magnesium based
CDKs	Cyclin-dependent kinases
CMC	Critical micelle concentration
CSA	Chondroitin sulphate A
CSP	Circumsporozoite protein
CQ	Chloroquine
CQS	Chloroquine sensitive strains
DDT	Dichlorodiphenyl trichloroethane
ddNTP	Dideoxynucleoside triphosphate
DHF	Dihydrofolate
DHFR	Dihydrofolate reductase
DHFS	Dihydrofolate synthase
DHNA	Dihydroneopterin aldolase
DHNP	Dihydroneopterin triphosphate
DHODase	Dihydroorotate dehydrogenase
DHPS	Dihydropteroate synthase
DHT	7-8 Dihydropteroate
DNA	Deoxyribonucleic acid
DNaseI	Deoxyribonuclease I
dNTP	Deoxyribonucleoside triphosphate
dTMP	Deoxythymidine monophosphate
DTT	Dithiotreitol
dUMP	Deoxyuridine monophosphate
EBA	Erythrocyte-binding antigen
E.C.	Enzyme Commission
E. coli	*Escherichia coli*
EDTA	(ethylenedinitrilo) tetraacetic acid
ER	Endoplasmic reticulum
EtBr	Ethidium bromide
EXP	Exported antigen
FAS	Fatty acid biosynthesis
FPGS	Folylpolyglutamate synthase
FV	Food vacuole
GCV	Glycine cleavage complex
GTP	Guanosine triphosphate
GTP-CH	GTP cyclohydrolase

HPPK	Hydroxymethyl pyrophosphokinase
Hsp	Heat shock protein
H. sapiens	*Homo sapiens*
ICAM-1	Intercellular adhesion molecule 1
IDC	Intraerythrocytic Developmental Cycle
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl-β-D-thiogalactopyranoside
ITNs	Insecticide treated bed nets
KAHRP	Knob associated histidine-rich proteins
LB	Luria-Bertani medium
L. casei	*Lactobacillus casei*
LC/MS/MS	Liquid chromatography-tandem mass spectrometry
LDH	Lactate dehydrogenase
LSA	Liver-stage antigen
MCs	Maurer's clefts
MMV	Medicines for Malaria Venture
mRNA	Messenger ribonucleic acid
MSP1	Merozoite surface protein 1
NO	Nitric oxide
NPP	New permeability pathway
OD	Optical density
ORF	Open reading frame
*p*ABA	*p*-amino benzoic acid
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
P. falciparum	*Plasmodium falciparum*
PfATP6	SERCA of *P. falciparum*
*Pf*CRT	*P. falciparum* chloroquine resistance transporter
PfEMP1	*P. falciparum* erythrocyte membrane protein 1
pfmdr 1	*P. falciparum* multidrug resistance 1 protein
PfPFT	*P. falciparum* franesyl transferase
PfRh	*P. falciparum* reticulocyte binding like homolog protein
P. pastoris	*Pichia pastoris*
pRBCs	Parasite infected red blood cells
PV	Parasitophorous vacuole
PVM	Parasitophorous vacuole membrane
PvRII	*P. vivax* Duffy binding protein
PYR	Pyrimethamine
RBCs	Red blood cells
RESA	Ring-infected erythrocyte surface antigen
RNA	Ribonucleic acid
RNase	Ribonuclease
SDS	Sodium dodecyl sulphate
SDX	Sulfadoxine
SERA	Serine repeat antigen
SERCA	Sarcoplasmic-endoplasmic reticulum Ca^{2+} ATPase
S. faecalis	*Streptococcus faecalis*
SHMT	Serine hydroxymethyltransferase

sHsps	Small heat-shock proteins
spp	species
TEMED	N,N,N'N'-tetramethylenediamine
TF	Trigger factor
T. gondii	*Toxoplasma gondii*
THF	Tetrahydrofolate
TIM	Triosephosphate isomerase
TNF	Tumour necrosis factor
TRAP	Thrombospondin-related adhesion protein
Tris	Tris (hydroxymethyl) aminomethane
tRNA	Transfer ribonucleic acid
TS	Thymidylate synthetase
TVN	Tubulovesicular network
U	Enzyme units
X-gal	5-bromo-4-chloro-3-indolyl-β-D galactopyranoside
YE	Yeast extract

Chapter 1
Introduction

…"There is no disputing the fact that African regeneration will remain impaired for as long as the scourge of malaria exists at current levels…we are made poor by malaria…"

His Excellency Chief Olusegun Obasanjo
President of the Federal Republic of Nigeria

The African Summit on Roll Back Malaria, Abuja, 25 April 2000

1. The Burden of Malaria

Malaria, the oldest disease known to mankind, is defined by the Taber's Cyclopaedic Medical Dictionary as an acute and sometimes chronic infectious disease due to protozoa of the genus *Plasmodium* infecting red blood cells (Thomas C.L., 1993). Malaria was derived from the Latin word 'mal-aria' meaning bad air, as it was thought that the noxious swamp gases around Rome caused the outbreaks of disease (Trenholme K.R. and Gardiner D.L., 2004). The oldest known malaria DNA was recently identified from a Roman baby graveyard (Abbot A., 2001). The causative agent for human malaria was first identified in 1880 by Charles-Louis-Alphonse Laveran, a French physician, and by 1897 Sir Ronald Ross demonstrated that malaria is transmitted to humans by the female *Anopheles* mosquito (Anonymous, 2002; Whitfield J., 2002). Human malaria is caused by four species of *Plasmodium*: *vivax*, *malariae*, *ovale* and *falciparum*. These four species differ from one another in geographical distribution, appearance, virulence, clinical features, as well as their resistance to anti-malarial drugs (Bloland P.B., 2001).

Malaria is endemic to tropical and subtropical regions with mostly poor developing countries at risk. It was estimated at the end of 2004 that 107 countries and territories had areas at risk of malaria transmission. This results in ~350-500 million clinical malaria episodes annually and more than 1 million deaths each year. 80% of these deaths occur in sub-Saharan African of which about 18% are children under the age of 5 years (www.rbm.who.int.). Figure 1.1 illustrates the distribution of malaria around the world, as well as the drug resistance to current antimalarials.

Areas where malaria transmission occurs (high/low)
▲ Chloroquine resistance
● Sulfadoxine-pyrimethamine resistance
★ Mefloquine resistance
⬡ Malaria-free areas
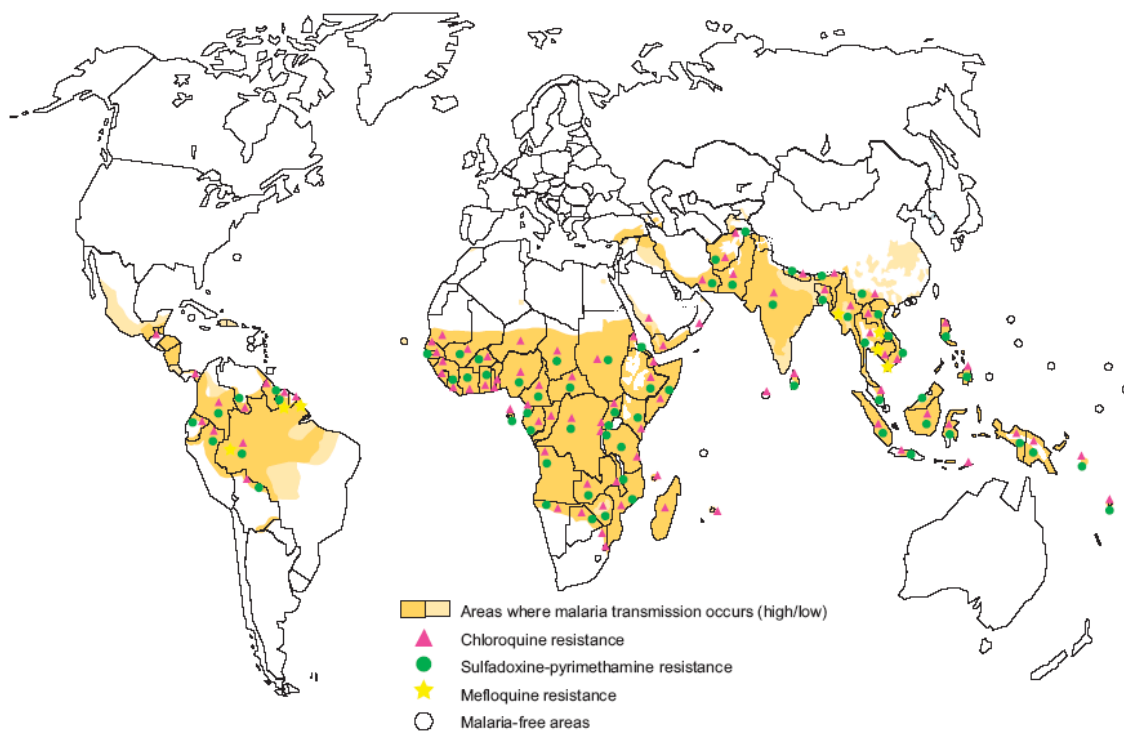
Figure 1-1: World distribution of malaria incidence. Resistance to current antimalarials is indicated in pink, green and yellow (www.rbm.who.int.).

Most of these infections are caused by *P. falciparum* and *P. vivax*. However, the greatest burden of malaria cases is caused by *P. falciparum*. The reason for this overwhelming burden includes the prevalence, virulence and drug resistance of *P. falciparum* (www.rbm.who.int.). The virulence is due to a high rate of asexual reproduction within erythrocytes in the human host (Winstanley P.A, 2000). *P. falciparum* have resistance to nearly all current anti-malarial drugs (Bloland P.B., 2001). Other causative factors for the high malaria mortality include the lack of infrastructure and resources with which to challenge and eradicate this threat in sub-Saharan Africa, contributing to the high prevalence of this parasite in transmission areas (www.rbm.who.int.).

1.1 Life cycle of malaria parasite

The life cycle of malaria involves two hosts, namely humans (blood stages) and mosquitoes (sporogenic cycle). The female mosquito *Anopheles* species (spp.) transmits the parasite during a blood meal. The Intraerythrocytic Developmental Cycle (IDC) starts with the ring stage (immature trophozoites), which mature into trophozoites followed by schizonts, that rupture and release merozoites allowing the infection cycle to repeat. During the IDC, some parasites differentiate into gametocytes, which are ingested by the female *Anopheles* mosquito during a blood meal, multiply (sporogenic cycle) and releases sporozoites into the

salivary glands, to continue the spread of this deadly disease. Each of the two hosts will be discussed in detail below. The parasitic life cycle is outlined in Figure 1.2.
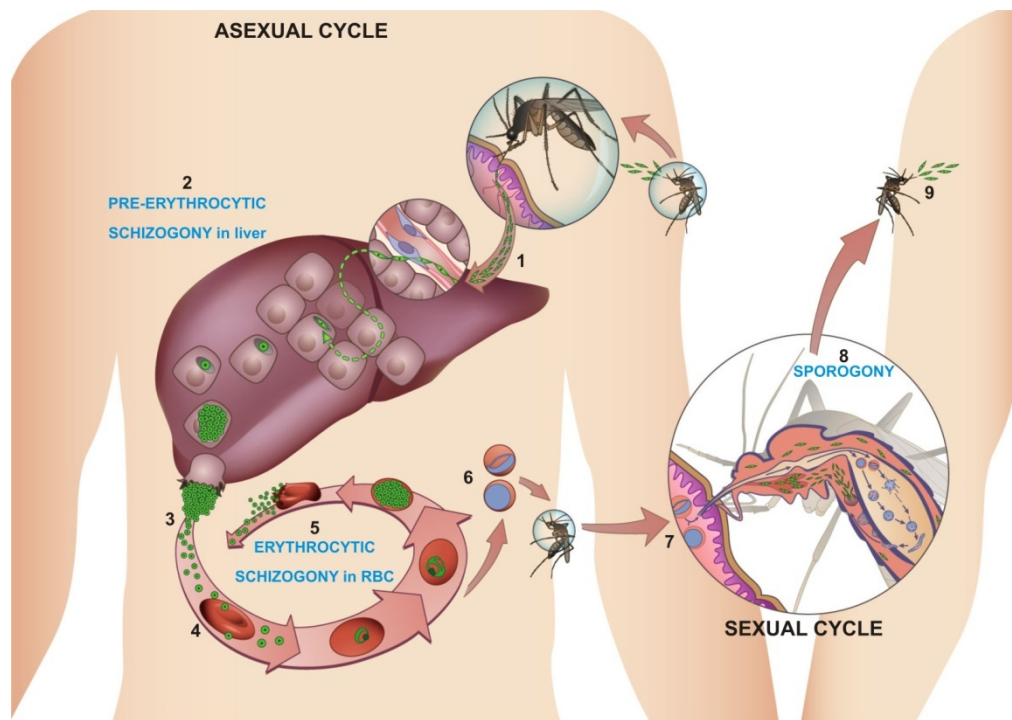
Figure 1-2: The lifecycle of *P. falciparum*. 1) Migration of sporozoites to the liver, 2) Schizogony in the liver, 3) Release of merozoites in the blood stream, 4) Invasion of the red blood cells, 5) IDC, 6) Formation of gametocytes, 7) Ingestion of gametocytes during a blood meal, 8) Sporogony in the mosquito, 9) Re-infection of human host. Adapted from (www.imm.ul.pt/html/uni14i.html).

1.1.1 Mosquito vector stages

The asexual parasite undergoes gametocytogenesis (maturation of gametocytes). The male (microgametocytes) and female (macrogametocytes) gametocytes are ingested during a blood meal by the female mosquito (Baton L.A. and Ranford-Cartwright L.C., 2005; Talman A.M. *et al.*, 2004). The microgametocytes contain mainly proteins for the subsequent gamete formation and possess little synthetic apparatus and cytoplasmic organelles (Baton L.A. and Ranford-Cartwright L.C., 2005). However, the macrogametocytes contain high amounts of ribosomes, the endoplasmic reticulum (ER), mitochondria and the nucleus is small and compact (Baton L.A. and Ranford-Cartwright L.C., 2005). The ingested blood containing the gametocytes as well as host components passes into the midgut lumen of the mosquito. The ingestion of the parasites initiates gametogenesis, which is the rapid development of extra-cellular microgametes and macrogametes (Figure 1.3). Extracellular macrogametes are characterised by their non-motile spherical shape, while the microgametes undergo exflagellation. These gametes rapidly fertilise to form diploid zygotes in the mosquito blood meal. The formation of a mature ookinete from the zygote is the start for oocyst formation and subsequent sporogony (Figure 1.3). The ookinetes migrate through the blood meal to

invade the midgut epithelium. The migration is presumably through a gliding motility. After invasion, amplification of parasite numbers increases markedly in a process known as sporogony. Development underneath the basal lamina of the epithelium is thought to ensure the evasion of the parasite from the immune system of the mosquito (Baton L.A. and Ranford-Cartwright L.C., 2005; Matuschewski K., 2006). The developed sporozoites are released from the oocysts and invade the salivary glands with the same gliding motility as the ookinetes. The sporozoites contains all the structural and functional features for the invasive stages of the parasite in the human host (Baton L.A. and Ranford-Cartwright L.C., 2005; Matuschewski K., 2006). During a blood meal the mosquito, inject the sporozoites into the human host via its secretory ducts. The sporozoites have to leave the site of introduction, enter a blood capillary, adhere to the liver cells and invade a hepatocyte within minutes to ensure its survival (Matuschewski K., 2006).
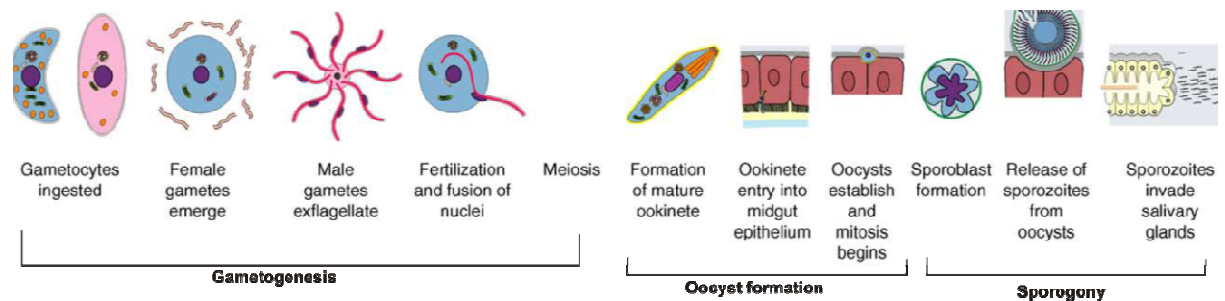
Figure 1-3: Malaria parasite development in the mosquito. Gametogenesis: Development of extracellular male and female gametes for fertilisation and meiosis, **Oocyst formation:** Ookinetes invade midgut to develop into sporozoites, **Sporogony:** Released sporozoites invade salivary glands to re-infect human host. Adapted from (Baton L.A. and Ranford-Cartwright L.C., 2005)

1.1.2 Intraerythrocytic Developmental Cycle

The sporozoites invade liver cells where they undergo asexual development to form merozoites, which are released into the bloodstream of the host. The released merozoites infect red blood cells and asexual development within the erythrocytes commences. The asexual cycle takes about 44-48 h with subsequent release of merozoites from the red blood cells (RBCs) and re-invasion. This cycle of development manifest in the host as reccurring waves of fever when red blood cells cyclically release the merozoites (Bannister L.H. *et al.*, 2000; Marti M. *et al.*, 2005).

Merozoite invasion is highly specific, ordered, and sequential and is of great immunological importance as the merozoite is momentarily exposed to host antibodies between leaving parasite infected red blood cells (pRBCs) and re-infecting new erythrocytes (Bannister L.H. *et al.*, 2000; Pasvol G., 2001). Invasion has various steps, each of which involves different receptor ligand interactions. The various steps include binding of the merozoite to the RBC,

4

re-orientation of its apical part to the membrane and junction formation for entry (O'Donnel R.A. *et al.*, 2006). Re-orientation of the merozoite requires apical membrane antigen 1 (AMA1), a microneme protein that is released before attachment (O'Donnel R.A. *et al.*, 2006). The merozoite is specifically adapted to invade erythrocytes and uses several erythrocyte receptors for attachment via specific ligands (Baum J. *et al.*, 2005). There are two protein families that are thought to act as ligands to help with junction formation (Baum J. *et al.*, 2005; Cowman A.F. and Crabb B.S., 2006; O'Donnel R.A. *et al.*, 2006). The first, Duffy binding like protein family, includes erythrocyte-binding antigen 175 (EBA-175), EBA-181 and EBA-140, (Baum J. *et al.*, 2005; Cowman A.F. and Crabb B.S., 2006; Soldati D. *et al.*, 2004). The second group of proteins are the *P. falciparum* reticulocyte binding-like proteins (PfRh) (Baum J. *et al.*, 2005). It is thought that this protein family is required for sensing and activation of the host erythrocyte for invasion (Soldati D. *et al.*, 2004). After attachment, apical re-orientation of the merozoite to the host cell and the ligand-mediated invasion, additional proteins from the rhoptries, micronemes, and dense granules are discharged. The proteins change the shape and composition of the membrane of the invaded RBC, resulting in the formation of the parasitophorous vacuole (PV). The PV and the parasitophorous vacuole membrane (PVM) act as a semi-permeable barrier between the host and the parasite. The PVM allows transport of nutrients from the host into the parasite and some parasite-derived proteins are exported to the RBC cytoplasm or RBC membrane (Marti M. *et al.*, 2005).

After invasion, the parasite flattens into a thin discoidal ring shape, containing major organelles such as the nucleus, mitochondria, plastid, ribosomes and ER (Bannister L.H. *et al.*, 2000). During the early ring stage, the development of membrane structures in the erythrocyte is initiated. These include a tubulovesicular network (TVN) as well as the Maurer's clefts (MCs), which are sheath-like structures underlying the pRBCs membrane (Bannister L.H. *et al.*, 2000; Marti M. *et al.*, 2005). The parasite obtains nutrients from its host RBC. (Figure 1.4; 0-10h) (Bannister L.H. *et al.*, 2000; Marti M. *et al.*, 2005). The rings eventually change shape to form the trophozoites, which have a rounder or more irregular form. The surface area of the trophozoite is characterised by irregular bulges and tubular invaginations (Bannister L.H. *et al.*, 2000). These bulges can extend to the host cell forming knobs. Knob-associated histidine-rich protein (KAHRP) and parasite-derived surface receptors such as *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) are concentrated in these knobs (Figure 1.4; 10-20h) (Bannister L.H. *et al.*, 2000; Marti M. *et al.*, 2005). The trophozoite subsequently becomes a schizont, which is an intraerythrocytic parasite undergoing repetitive nuclear divisions (Bannister L.H. *et al.*, 2000). The nuclear division results in 16-32 merozoites. Each merozoite contained in the PV is equipped with a set of

rhoptries, micronemes and dense granules and are thus equipped for re-invasion of RBCs, (Figure 1.4; >40h) (Bannister L.H. *et al.*, 2000; Marti M. *et al.*, 2005).
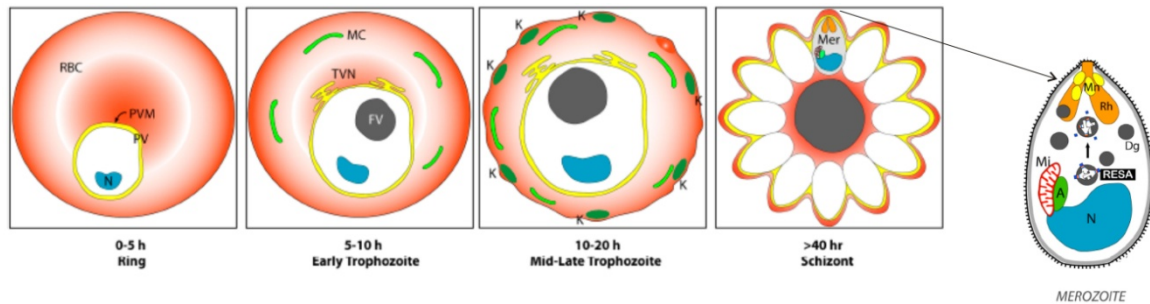
Figure 1-4: Asexual development of malaria parasites in the RBC. 0–5 h: parasites is surrounded by PV and seen as ring stage. 5–10 h: the parasite induces the formation of TVN and MCs. The deposition of haemozoin crystals is visible in FV. 10–20 h: the trophozoite stage induces knob (K) formation on the pRBC surface. >40 h: several rounds of asexual division, results in merozoite formation, which are still surrounded by the PVM until they burst out and invade uninfected RBCs. [Mer] merozoite; [N] nucleus; [Mn] Micronemes; [Rh] Rhoptries; [Dg] Dense granules. Adapted from (Marti M. *et al.*, 2005).

1.2 Pathogenesis of malaria

P. falciparum causes a broad spectrum of clinical features such as parasitic sepsis, the impairment of the nervous system (cerebral malaria), prominent anaemia due to haemolysis, hypoglycaemia, acidosis, circulatory collapse, multi-organ failure and coagulopathies (Heddini A., 2002; Rasti N. *et al.*, 2004; www.cdc.com, 2004).These complications results from the pRBCs adhering to host endothelium (sequestration) and to non-infected erythrocytes (rosetting), eventually leading to the obstruction of microvasculature's in various tissues and organs such as the brain in cerebral malaria as well as tissue hypoxia (Rasti N. *et al.*, 2004). PfEMP1 is the major adhesive ligand mediating these adhesions and plays a key role in the pathogenesis of *P. falciparum* (Rasti N. *et al.*, 2004; Weatherall D.J. *et al.*, 2002). The multiple adhesion domains of PfEMP1 recognise several host receptors to which it bind during sequestration. One such example is the binding of PfEMP1 to the intercellular adhesion molecule 1 (ICAM-1) receptor in cerebral malaria, and chondroitin sulphate A (CSA) in the placenta (Miller L.H. *et al.*, 2002). Other factors contributing to the pathogenesis of malaria include the high levels of tumor necrosis factor (TNF), cytokines and nitric oxide (NO) levels in severe *falciparum* malaria (Pasvol G., 2001). Pro-inflammatory cytokines such as TNF act as homeostatic agents, but can cause pathology in excess levels such as headaches, fevers, rigors, nausea and vomiting. All of these symptoms have a mechanism through inflammatory cytokines (Clark I.A. *et al.*, 2006). It is also suggested that increasing levels of TNF induce the brain cells to express ICAM-1, which could then bind to the pRBCs ensuring sequestration (Miller L.H. *et al.*, 2002). Another important molecule is the haemozoin deposits released from erupting red blood cells. It was found that the haemozoin

inhibits several monocyte functions such as phagocytosis and cytokine production needed for immunity (Skorokhod O.A. *et al.*, 2004; Tekwani B.L. and Walker L.A., 2005).

1.3 Malaria Control

Malaria is a multifaceted problem, with interconnectivity between mosquito and parasite and between mosquito and human. Each of these links is targeted in malaria control strategies, to provide a means of eradicating this devastating disease. The main problems facing malaria control strategies are the insecticide resistance of the mosquito, resistance of the parasite to chemotherapeutic drugs, and social and environmental changes brought about by human migration (Tripathi R.P. *et al.*, 2005). Each of these control strategies (vector control and disease control) will be discussed in more detail in the following sections.

1.3.1 Vector control

Vector control strategies are aimed at the mosquito vector. Vector control can be achieved by reducing the density of the *Anopheles* mosquito, interrupting the life cycle of the parasite in the mosquito, or creating a barrier between the human host and the mosquito vector (Tripathi R.P. *et al.*, 2005). The first insecticide used against adult malaria mosquitoes was bis(4-chlorophenyl)-1,1,1,1-trichlorethane (DDT) also known as dichlorodiphenyl trichloroethane in the 1950s and 1960s (Rogan W.J. and Chen A., 2005). In 1874 DDT was synthesised but it was only in the late 1930's that it's insecticidal properties were described by Paul Müller (Turusov V. *et al.*, 2002). DDT was used worldwide as an insecticide especially against malaria vector mosquitoes due to its repellent, irritant and toxic actions (Roberts D.R. *et al.*, 2000). However DDT was banned in the 1970s based on ecological considerations, such as the persistence of DDT in the environment and possible toxic effects in humans (Roberts D.R. *et al.*, 2000; Rogan W.J. and Chen A., 2005). Due to the ban of DDT other transmission blocking methods were used and investigated such as the use of insecticide treated bed nets (ITNs). The principle insecticide used to impregnate bed nets is pyrethroid. Due to pyrethroid being the only class of insecticide currently used for ITNs, pyrethroid-resistant *Anopheles* is emerging (Guillet P. *et al.*, 2001). A second limitation is the effect of these insecticides on the environment and health concerns for humans if not used correctly. Small doses of pyrethroid could cause a minor risk of acute toxicity, whereas larger doses could affect the nervous system (Hemingway J. *et al.*, 2006; Toure Y.T. *et al.*, 2004). Recently the use of DDT was re-introduced as insecticide since its ban in the 1970s (www.allafrica.com/stories, 2007). The recent availability of the *Anopheles gambiae* (*A. gambiae*) genome provides a platform for improved vector characterisation, understanding of the mechanisms of insecticide resistance and provide new avenues for the development of

novel insecticides (Hoffman S.L *et al.*, 2002; Toure Y.T. *et al.*, 2004). Furthermore, the understanding of the developmental cycle in the midgut and salivary glands of the insect vector could aid in the identification of potential targets for insecticides, or for inhibiting the transmission of parasites to the human host (Kanzok S.M. and Zheng L., 2003). An alternative strategy for vector control is the use of genetically modified mosquitoes. These mosquitoes can be divided in two categories: firstly, the mosquito carries a transgenic gene which would impair pathogen development; secondly, a population carrying a gene, which prevents offspring production (also known as sterile insect technique) (Marrelli M.T. *et al.*, 2006; Moreira L.A. *et al.*, 2002). The use of transgenic mosquitoes have gained a lot of interest and have been used to determine the effect on several vectors of *Plasmodia*, but still requires a lot of in depth-research (Marrelli M.T. *et al.*, 2006).

1.3.2 Vaccines

Control strategies such as anti-malarial drugs and vaccination, are aimed at preventing disease onset in the human host. Vaccination against malaria could present a cost-effective alternative to eradicate this burden (Tripathi R.P. *et al.*, 2005). Several stages of the parasites' life cycle are targeted for vaccine development. These include the pre-erythrocytic stages such as sporozoite invasion of hepatocytes, the hepatic stages and merozoite invasion of red blood cells, the asexual stages as well as the sexual stages of development (Figure 1.5) (Richie T.L. and Saul A., 2002). Pre-erythrocytic vaccines target antigens expressed on the surface of sporozoites such as the circumsporozoite protein (CSP) and thrombospondin-related adhesion protein (TRAP) thus preventing binding interactions with the hepatocytes (Figure 1.5) (Richie T.L. and Saul A., 2002). Several vaccines targeting hepatocyte invasion antigens as mentioned above are in clinical development (Richie T.L. and Saul A., 2002; Tripathi R.P. *et al.*, 2005). The most advanced pre-erythrocytic vaccine is the RTS,S / AS02A, which contains the CSP of *P. falciparum* expressed with hepatitis-B surface antigen together with the adjuvant AS02A (Greenwood B.M. *et al.*, 2005). Antigens targeted in the hepatic stages include the liver-stage antigen 1, 3 (LSA-1, 3) and the exported antigen (EXP-1). Merozoite surface protein 1 (MSP1), AMA1 and serine repeat antigen (SERA) are just some of the antigens present on the merozoite targeted to block erythrocyte invasion (Richie T.L. and Saul A., 2002). Other antigens targeted during the asexual blood stages include the ring-infected erythrocyte surface antigen (RESA) and PfEMP1 on the trophozoites and schizonts, respectively (Greenwood B.M. *et al.*, 2005; Richie T.L. and Saul A., 2002). The antigens targeting the sexual stages of the parasite include Pfs 25,28,48/45, which inhibits exflagellation, fertilisation as well as neutralising ookinete function (Figure 1.5) (Richie T.L. and Saul A., 2002). The post-genomic analyses of

P. falciparum are leading the way to gain better insight into existing vaccine candidates, as well as providing new potential targets for investigation (Waters A., 2006).
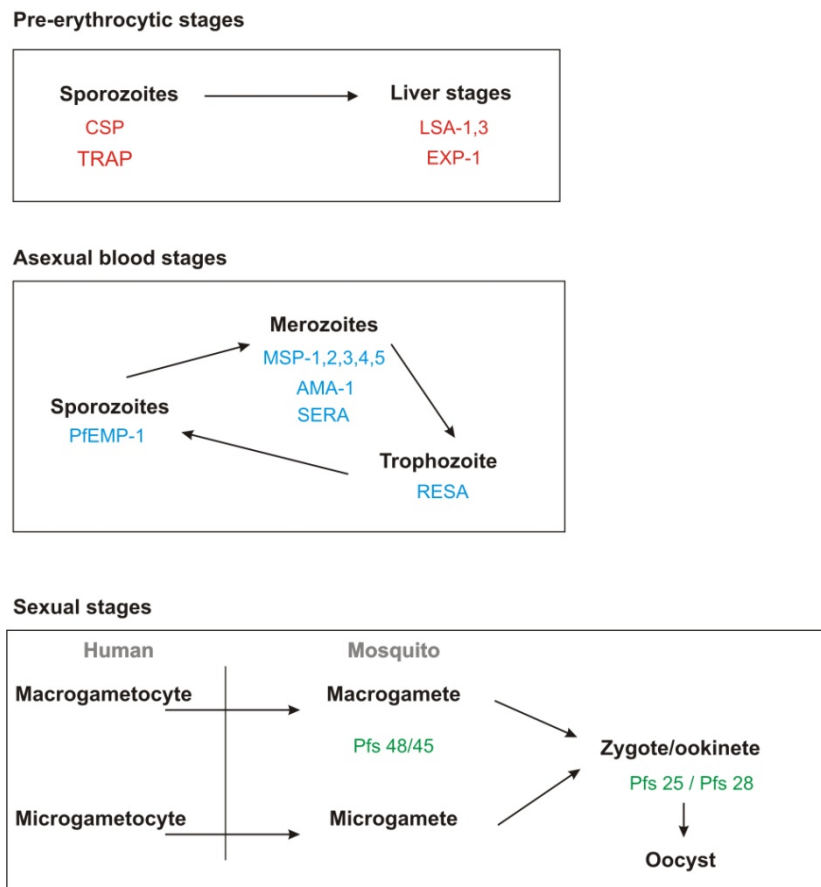
Figure 1-5: Malaria vaccines under development. Antigens targeted are indicated with colours. Adapted from (Greenwood B.M. *et al.*, 2005).

1.3.3 Antimalarial Drugs

Currently the principal component of parasite control strategies is the use of antimalarial drugs, which have been available for over fifty years in contrast to vaccines. However, the foremost problem with the use of drugs is the rapid progression of drug-resistant parasite strains (www.rbm.who.int.). The currently used antimalarials can be divided into three groups: the quinolines (quinine, chloroquine, mefloquine, primaquine), the artemisinin derivatives and the antifolates (sulfadoxine (SDX) and pyrimethamine (PYR)). Table 1.1 illustrates the currently used antimalarials, some of which will be discussed in detail in this section.

Table 1-1: Currently used antimalarial drugs. Adapted from (Ridley R.G., 2002; Vallely A. *et al.*, 2007; Yeh I. and Altman R.B., 2006).

Antimalarial drugs	Functions
4-aminoquinolines	
Chloroquine	Interfere with haemoglobin digestion, causing a build-up of toxic product in the parasite
Amodiaquine	
Halofantrine	
Lumefantrine	
Mefloquine	Lipophyllic drugs, thought to act through disruption of phospholipid metabolism or inhibition of endocytosis
Quinine	
Artemisinin antimalarials	
Artemisinin	Ferrous iron in haem stores of parasite cleaves a peroxide ring, generating free radicals thus increasing oxidative stress. Inhibits P. *falciparum* sarcoplasmic-endoplasmic reticulum calcium ATPase (SERCA)
Dihydroartemisinin	
Artemether	
Arteether	
Artesunate	
Other antimalarials	
Sulfadoxine	Folic acid antagonist Inhibits dihydropteroate synthase (DHPS)
Proguanil	Folic acid antagonist Inhibits dihydrofolate reductase (DHFR)
Pyrimethamine	
Atovaquone	Blocks mitochondrial respiration, by binding to cytochrome c reductase
Combination therapies	
Sulfadoxine-pyrimethamine	Inhibits DHPS and DHFR synergistically
Atovaquone-proguanil (malarone)	Inhibits mitochondrial metabolism - DHFR
Dapsone-proguanil (lap-dap)	Inhibits DHPS and DHFR
Lumefantrine-artemether	Inhibits metabolism of haem-SERCA

The parasite obtains nutrients from its host RBC. Haemoglobin is degraded to its haem derivative (ferriprotoporphyrin IX). This derivative is converted to brown inert haemozoin crystals, which accumulate within the food vacuole (FV) throughout the IDC-cycle, (Bannister L.H. *et al.*, 2000; Marti M. *et al.*, 2005). The current antimalarials such as quinoline and artemisinin are thought to interfere with the breakdown of haem Fe^{2+} to the oxidised form haematin Fe^{3+}, and the subsequent formation of haemozoin. The prevention of haemozoin formation leads to free radicals in the FV, which are toxic to the parasite and may possibly disrupt parasite membranes (Ridley R.G., 2002; Tripathi R.P. *et al.*, 2005).

1.3.3.1 4-aminoquinolines

Quinine was originally isolated from the *Cinchona* tree in the seventeenth century (Ridley R.G., 2002; Tripathi R.P. *et al.*, 2005). The antimalarial activity of quinine was demonstrated in 1820 by Pelletier and Caventau (Spikes J.D., 1998). Due to quinine's undesirable side effects and emergence of quinine resistant strains, development of synthetic antimalarials gained a lot of interest (Spikes J.D., 1998). In the 1940's chloroquine (CQ), a synthetic 4-aminoquinoline was developed that showed less toxicity than quinine (Ridley R.G., 2002; Rieckmann K.H., 2006). CQ is the most widely used 4-aminoquinoline against malaria while mefloquine and halofantrine are used against chloroquine-resistant strains (Ridley R.G., 2002). The antimalarial activity of the quinolines acts by disrupting the detoxification of haem degradation, resulting in the build-up of toxic haem in the digestive FV (Arav-Boger R. and Shapiro T.A., 2005; Biagini G.A. *et al.*, 2003). The main limitations of the quinolines are the emergence of resistance. The mechanism of resistance to these drugs is currently being investigated. Two independent genetic sources have been identified namely, the *P. falciparum* chloroquine-resistance transporter (*Pf*CRT) and the *P. falciparum* multidrug-resistance 1 protein (*pfmdr1*), which both play a role to impart resistance (Arav-Boger R. and Shapiro T.A., 2005). Johnson *et al.* proposed a "charged drug leak" to explain the resistance mechanism of CQ in particular (Johnson D.J. *et al.*, 2004). Evidence suggests that point mutations in the *Pf*CRT not only leads to CQ resistance, but also influences the susceptibility of parasites to unrelated antimalarials (Johnson D.J. *et al.*, 2004). The authors propose the following mechanism prior to mutation: a positively charged lysine residue in the transmembrane *Pf*CRT prevents the movement of CQ through *Pf*CRT. Loss of this positive charge through a point mutation (K76T) results in the efflux of CQ out of the digestive vacuole into the PV. This efflux of CQ reduces the concentration of CQ and confers resistance by less interaction of CQ with haem metabolism (Johnson D.J. *et al.*, 2004). Recently three possible transporter models for CQ resistance were proposed by Sanchez *et al.*: the channel model, co-transporter model and the transporter model (Sanchez C.P. *et al.*, 2007). The authors suggested that *Pf*CRT might function as a channel, which allows the protonated CQ to move against the electrochemical gradient from the FV to the cytoplasm. This channel gets blocked by the positive lysine in the transmembrane domain in CQ sensitive strains (CQS) (Sanchez C.P. *et al.*, 2007). The transporter model propose the use of primary or secondary energy by an energy-dependent *Pf*CRT-linked CQ efflux carrier to expel the protonated CQ from the FV. The proposed co-transporter model involves the binding of CQ and H^+ to the co-transporter in the FV, inter-conversion of these substrates from the FV to the cytoplasm and the subsequent release of these substrates in the cytoplasm. Due to the CQ and H^+ being co-substrates a coupled extrusion of CQ out of the FV can occur due to an outwardly directed proton gradient. In CQS this inter-conversion is blocked resulting in a loss of efflux of CQ out of the FV (Sanchez C.P. *et al.*, 2007). *Pfmdr1* is

a homologue of the major multidrug transporter in mammalian cells (Duraisingh M.T. and Cowman A.F., 2005). It is speculated that the *pfmdr1* transporter regulates pH in the FV, which has an effect on accumulation of weak base drugs in this organelle. Mutations in *pfmdr1* are associated with resistance to halofantrine and mefloquine (Arav-Boger R. and Shapiro T.A., 2005). Point mutations in *pfmdr1* together with point mutations in *Pf*CRT confer CQ resistance (Yeh I. and Altman R.B., 2006). Understanding these mechanisms, and interactions of these two proteins with each other, are crucial to elucidate the resistance of parasites to these antimalarials and overcoming them (Cowman A.F., 2001; Duraisingh M.T. and Cowman A.F., 2005; Kirk K., 2004).

1.3.3.2 Artemisinins

Artemisinin and its derivatives are used in the treatment of multidrug resistant parasites. These compounds are extracted from the plant *Artemisia annua* (Haynes R.K and Krishna S., 2004). The artemisinins affect all stages of the malaria parasite (including young rings). They do not however, affect the hepatic stages but do hinder gametocyte development (Balint G.A., 2001; Haynes R.K and Krishna S., 2004). Importantly, artemisinins pass through the blood-brain barrier, as well as the placenta barriers, making them useful as a treatment for complicated malaria and in pregnancy (Balint G.A., 2001). It is hypothesised that haem is not the target of artemisinins and only function in the activation of artemisinins and its derivatives. Artemisinin and its derivatives are sesquiterpene trioxane lactones that contain a peroxide bridge (Haynes R.K and Krishna S., 2004). This endoperoxide bridge undergoes iron-catalysed activation in the FV to form free radicals. Artemisinin reacts with haem Fe^{2+} to undergo these activations (Messori L. *et al.*, 2006). The free radicals alkylate proteins and damage the organelles and membranes in the parasite by oxidative stress (Arav-Boger R. and Shapiro T.A., 2005; Biagini G.A. *et al.*, 2003). Recent studies proposed that artemisinin inhibit the SERCA of *P. falciparum* (PfATP6) (Krishna S. *et al.*, 2006).

Many commonly used antimalarials (e.g. SDX, PYR, and proguanil) are folic acid antagonists. These will be discussed in more detail in Section 1.4

1.3.4 Validated and new drug targets

The following section will discuss biochemicaly validated drug targets, as well as some possible new targets against which chemotherapeutics could be developed. Many of these targets can be related to functions of specific organelles as can be seen from Figure 1.6 (Ridley R.G., 2002). Current antimalarial drugs target specific subcellular locations such as the lysosomal food vacuole (quinolines), apicoplast (doxycycline), mitochondrion

(atovaquone), or the cytosol (antifolates). For each of these compartments, numerous other potential drug targets exist and are discussed below.
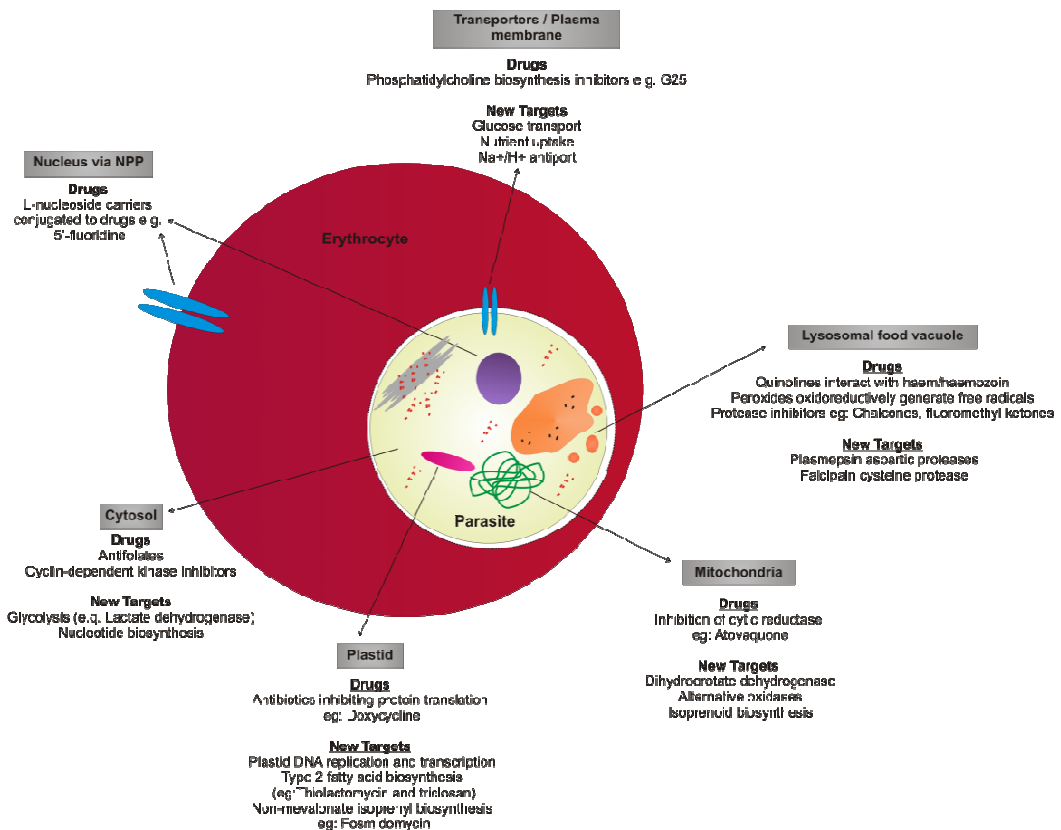
Figure 1-6: Drug targets in the malaria parasite *P. falciparum*. A schematic representation of a trophozoite in a red blood cell. Adapted from (Biagini G.A. et al., 2003; Ridley R.G., 2002).

1.3.4.1 Mitochondrial electron transport

The malaria parasite resides in an oxygen poor environment and depends on the glycolytic pathway as its main source of ATP. Despite the lack of oxidative phosphorylation an active electron transport chain is maintained (Mather M.W. *et al.*, 2007), mainly for the regeneration of ubiquinone (Painter H.J. *et al.*, 2007). Ubiquinone act as an electron acceptor during the coupling of cytochrome c reductase to dihydroorotate dehydrogenase (DHODase), an enzyme involved in nucleoside biosynthesis, (Figure 1.6), (Painter H.J. *et al.*, 2007; Ridley R.G., 2002; Tripathi R.P. *et al.*, 2005). The transmembrane proton gradient, for transport of metabolites across the membrane, is strengthened by the electron transport chain, which is essential for parasite survival (Painter H.J. *et al.*, 2007). The electron transport chain also eliminates oxygen, which could form reactive oxygen species (Mather M.W. *et al.*, 2007). Atovaquone is an inhibitor of cytochrome c reductase and results in the collapse of the mitochondrial membrane potential. Due to emergence of resistance, it is used in combination with proguanil (Biagini G.A. *et al.*, 2003; Mather M.W. *et al.*, 2007). β-methoxy acrylates are other compounds known to inhibit cytochrome c reductase (Tripathi R.P. *et al.*, 2005).

Inhibition of cytochrome c reductase also leads to the inhibition of enzymes such as DHODase, which requires electrons from the mitochondrial electron flow. DHODase is primarily involved in pyrimidine biosynthesis, thus providing a promising new target for control (Biagini G.A. *et al.*, 2003). Another discovery project, by the Medicine for Malaria Venture (MMV), is targeting isoprenoid biosynthesis in the mitochondria. Inhibitors against the lead protein, farnesyl transferase (PfPFT) are under investigation (Biagini G.A. *et al.*, 2003; Medicines for Malaria Venture., 2005).

1.3.4.2 Proteases and haemoglobin degradation in the lysosomal food vacuole

During the parasite IDC, a significant amount of haemoglobin is hydrolysed to haeme in the lysosomal food vacuole to form inert haemozoin. The degradation of haemoglobin also provides a source of free amino acids required for malarial protein synthesis and energy metabolism (Bannister L.H. *et al.*, 2000; Tripathi R.P. *et al.*, 2005). It has been suggested that haemoglobin degradation may have additional functions. These include maintaining osmotic balance, to prevent premature red cell lysis (Francis S.E. *et al.*, 1997). It was recently found in *in vitro* studies that *P. falciparum* requires only isoleucine (only amino acid that is lacking in human haemoglobin) for growth (Liu J. *et al.*, 2006). The authors suggested that haemoglobin degradation supplies most of the amino acid requirements of the parasite, but external amino acid acquisition such as isoleucine is needed (Liu J. *et al.*, 2006). A recent study indicated the influx of isoleucine via the new permeability pathways (NPPs) in exchange for leucine (most abundant amino acid in adult haemoglobin) (Martin R.E. and Kirk K., 2007). The degradation of haemoglobin might thus serve a double role by not only providing amino acids for protein synthesis but also for the involvement with transporters for the influx of isoleucine and possibly other important metabolites. The involvement of proteases such as plasmepsins (aspartic proteases), falcipain (cysteine protease), and the metallopeptidases in haem metabolism, provide key areas of chemotherapeutic target research, as haemoglobin degradation is necessary for parasite survival (Figure 1.6), (Biagini G.A. *et al.*, 2003; Ridley R.G., 2002).

1.3.4.3 Apicoplast metabolism

The apicoplast is an organelle unique to Apicomplexan parasites. Protozoan parasites of the phylum Apicomplexa cause serious disease and death in humans (*Plasmodium* spp., *Toxoplasma* spp., and *Cryptosporidium* spp.) as well as deadly diseases in animals (*Neospora* spp., *Eimeria* spp., *Babesia* spp. and *Theileria* spp.) (Belli S.I. *et al.*, 2005). The apicoplast is similar to a plant's chloroplast and contains its own genome (Fichera M.E. and

Roos D.S., 1997; Gornicki P., 2003). The apicoplast genome contains elements of prokaryotic transcription and translation systems, which are the target of several current antimalarial antibiotics such as tetracycline, doxycycline and clindamycin (Ridley R.G., 2002). Other apicoplast functions include Type-II fatty acid biosynthesis (FAS-II), and the non-mevalonate pathway leading to synthesis of isopentenyl diphosphate subunits (Gornicki P., 2003; Ridley R.G., 2002). These metabolic pathways provide a new possibility for chemotherapeutic development, as shown in Figure 1.6. Several lead molecules have been identified in these pathways. Triclosan inhibits the enoyl-acyl carrier protein reductase, while fosmidomycin acts on 1-deoxy-D-xylo-5-phosphate synthase. Both of these inhibitors lead to arrest of parasite growth (Biagini G.A. *et al.*, 2003; Tripathi R.P. *et al.*, 2005).

1.3.4.4 Cytosolic drug targets

There are several targets in the cytoplasm of the parasite, which are currently controlled by antimalarials. The primary cytosolic target is the folate pathway, but due to the emergence of resistance to the currently used antifolates, potential new drug targets are urgently needed (Ridley R.G., 2002; Tripathi R.P. *et al.*, 2005). Section 1.6 will evaluate folate metabolism of the malaria parasite, its status, targets and future possibilities. Other targets in the cytosol comprise the Cyclin-dependent kinases (CDKs), which are essential for regulation of the cell cycle. Several enzymes of this family have been identified in *P. falciparum* and are thought to play a crucial role in parasite growth and differentiation (Biagini G.A. *et al.*, 2003; Keenan S.M. and Welsh W.J., 2004). The CDKs are attractive targets as there is a high degree of sequence conservation between *P. falciparum* and human CDKs, which suggest that they are activated in a similar manner to the mammalian CDKs, (Figure 1.6), (Biagini G.A. *et al.*, 2003). Another attractive target is the glycolytic pathway, due to the dependence of the parasite on glycolysis for energy production. Two enzymes are currently being investigated for the development of new antimalarials. These enzymes are the cytosolic enzyme lactate dehydrogenase (LDH), which is involved in anaerobic glycolysis and, triosephosphate isomerase (TIM) which is involved in the isomerisation of glyceraldehyde-3-phosphate to dihydroxyacetone phosphate (Tripathi R.P. *et al.*, 2005).

1.3.4.5 Transporters and membrane bound drug targets

A variety of membrane proteins involved in e.g. the transport of glucose, other nutrients and essential ions are also potential points of drug inhibition (Ridley R.G., 2002). During the IDC, the parasite synthesises a considerable number of membranes and increases the permeability of the erythrocyte membrane. A target during membrane biosynthesis is *de novo* phoshatidylcholine biosynthesis, which is inhibited by antimalarial drugs designed to

mimic choline (Biagini G.A. *et al.*, 2003). The NPP induced in the erythrocyte by the parasite is imperative for the flux of essential nutrients required by the parasite, (Figure 1.6), (Kirk K., 2004). The NPP is significantly different from normal RBCs making them attractive drug targets (Biagini G.A. *et al.*, 2003). Several inhibitory compounds of the NPP such as phloridzin, glibenclamide, and a range of acrylaminobenzoates have been shown to inhibit the growth of malaria parasites in culture (Kirk K., 2004). Other potential transporters that can be inhibited include the hexose transporter (Joet T. and Krishna S., 2004), nutrient uptake transporters and Na^+/H^+ antiport (Kirk K., 2004; Ridley R.G., 2002).

1.4 Folate metabolism

1.4.1 Folate metabolic organisation

The activities of the folate enzymes play vital roles in the synthesis and recycling of the folate derivative tetrahydrofolate (THF) (vitamin B9). Folate derivatives (such as the THFs) are important cellular cofactors supplying one carbon (C1) units to major metabolic pathways. These include biosynthesis of methionine, purines and pyrimidines, which are essential for DNA synthesis. Formylmethionine tRNA also requires a C1 unit for the initiation of protein synthesis. Folate derivatives are furthermore involved in metabolism of histidine and glutamic acid and mediate the conversion of serine to glycine (Bollheimer L.C. *et al.*, 2005; Nzila A. *et al.*, 2005b; Stokstad E.L.R. and Koch J., 1967). The availability of these folate derivatives for rapidly dividing cells such as tumours, bacteria and malaria parasites are obviously vital for their rapid growth. Metabolically active folates contain pteridine, *p*-aminobenzoate and glutamate entities, along with the presence of a number of γ-glutamyl residues. These folates are termed folylpolyglutamates and are mainly involved in the acceptance and donation of C1 molecules (Cossins E.A., 2000). Due to the high demand for folate derivatives in rapidly dividing cells the folate pathway has proved to be a good target for antimalarials. Inhibition of specific enzymes in this pathway would ultimately lead to cell death, due to a decrease in DNA and protein synthesis. Several antifolates (methotrexate and antibacterials such as trimethoprim) have been utilised not only as antimalarials, but also as anti-cancer agents (Costi M.P. and Ferrari S., 2001; Nzila A. *et al.*, 2005a; Nzila A. *et al.*, 2005b). The *de novo* folate synthesis is a unique biosynthetic pathway in the human malaria parasite, as shown in Figure 1.7 (Krungkrai J. *et al.*, 1989). This pathway is different from the human host as humans can not synthesise folates *de novo*, these differences thus make this metabolic pathway a excellent antimalarial drug target (Krungkrai J. *et al.*, 1989). As shown in Figure 1.7, folate biosynthesis in the malaria parasite can be divided into two main routes, namely *de novo* synthesis and folate salvage. *De novo* synthesis uses GTP as starting metabolite and ends with the production of 7,8-dihydrofolate (DHF). The DHF is reduced to the

biologically active 5,6,7,8-THF. THF is recycled back to DHF through the thymidylate cycle. The folylpolyglutamates are synthesised by the addition of L-glutamate to THF. Dihydrofolate reductase (DHFR) catalyses two major reactions in the folate biosynthesis pathway. Firstly, the synthesis of endogenous THF *de novo* required for C1 units and, secondly the recycling of DHF for salvage of oxidised forms of folate (Nzila A. *et al.*, 2005a). The recycling of DHF is catalysed by the thymidylate synthetase (TS) activity of dihydrofolate reductase-thymidylate synthetase (DHFR-TS). This reaction involves the reductive methylation of 2'-deoxyuridine-5'-monophosphate (dUMP) to dTMP using THF as C1 methyl donor (Nzila A. *et al.*, 2005a).
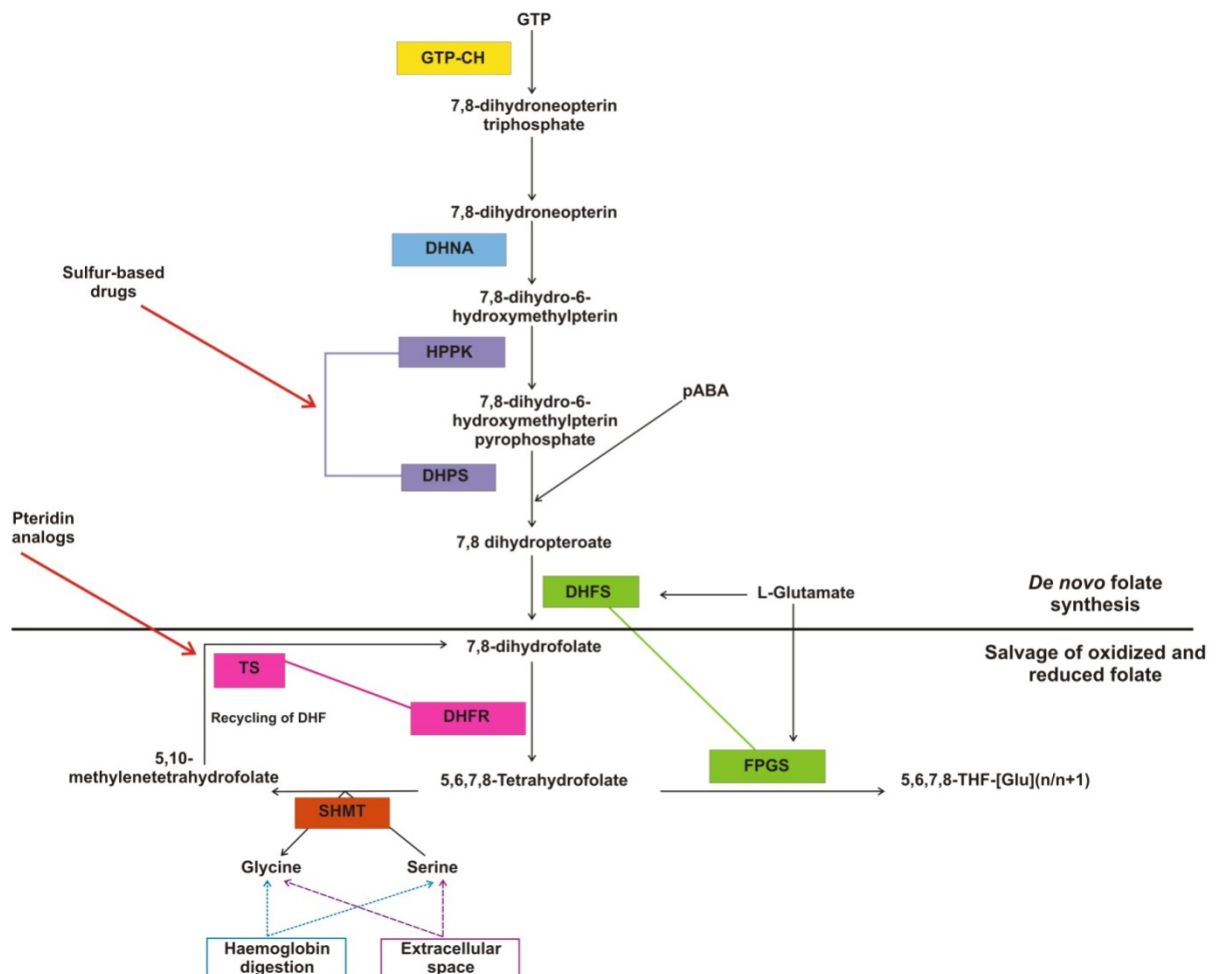
Figure 1-7: Folate metabolism of *P. falciparum*. Enzymes are indicated in coloured boxes. Enzymes with the same colour indicate the bifunctional enzymes. The targets for current antimalarials are indicated with the red arrows. GTPCH, GTP cyclohydrolase; DHNA, dihydroneopterin aldolase; HPPK, hydroxymethyl pyrophosphokinase; DHPS, dihydropteroate synthase; DHFS, dihydrofolate synthase; DHFR, dihydrofolate reductase; FPGS, folylpolyglutamate synthase; SHMT, serine hydroxymethyltransferase; TS, thymidylate synthase. Adapted from (http://sites.huji.ac.il.; Hyde J.E., 2005).

1.4.2 Antifolates

Folate antagonists directed against malarial proteins include the sulpha-based drugs, which inhibit the dihydropteroate synthase activity of bifunctional hydroxymethyldihydropterin pyrophosphokinase-dihydropteroate synthase (HPPK-DHPS). The second class of antagonists are the pteridin analogues, inhibitors of dihydrofolate reductase activity of the DHFR-TS (Nzila A., 2006a). Inhibitors directed against DHFR include PYR, proguanil and chlorproguanil, all of which have a higher binding affinity towards *P. falciparum* DHFR than to the human counterpart (Nzila A., 2006b). HPPK-DHPS is responsible for the formation of DHT in the *de novo* pathway and is inhibited by SDX (Biagini G.A. *et al.*, 2003). The antifolates are also substrate analogues. This unfortunately, leads to the development of mutations due the selective pressure on the active sites in these enzymes, rendering the parasite resistant to the drug (Nzila A., 2006b). SDX resistance is due to 5 mutations in DHPS namely A437Gly, S436F/A, A581G, K540E and A613S/T (De Beer T.A.P. *et al.*, 2006; Gregson A. and Plowe C.V., 2005). If the DHPS of the parasite contains multiple mutations, they act synergistically rather than additively (Gregson A. and Plowe C.V., 2005). PYR resistance arises from 5 mutations in DHFR (I164L, S108N/T, C59R, N51I and A16V). The main mutation identified in all antifolate-resistant strains is S108N. Additional mutations of N51I and C59R show an increase in resistance to PYR. A quadruple mutation of the above with the mutation I164L causes a 1000-fold increase in resistance (Gregson A. and Plowe C.V., 2005; Nzila A., 2006b). Resistance to cycloguanil is associated with point mutations at A16V and S108T (Nzila A., 2006b). These mutations also result in poor binding affinities of enzyme with inhibitors (Gregson A. and Plowe C.V., 2005). Structural analysis indicated that double and quadruple mutants have a high resistance to PYR and cycloguanil, but are sensitive to WR99210, a new class of antifolates (Yuvaniyama J. *et al.*, 2003). Analysis also indicated that the main effects of these mutations on the active site were due to steric constraints (S108N) and main-chain movements, resulting in an opening of the active site, which weakens the binding of inhibitors (N51I) (Yuthavong Y. *et al.*, 2005). Recent analyses of the crystal structure of DHFR-TS provided insight into the structural basis of resistance, which can assist in the development of more successful inhibitors (Yuthavong Y. *et al.*, 2005; Yuvaniyama J. *et al.*, 2003).

1.4.2.1 Drug synergy and effect of exogenous folate utilisation

The treatment of antifolate-resistant strains was modified by the synergistic use of drugs that target different points in the folate pathway. The success of these drugs could be attributed to the fact that they block both *de novo* synthesis and salvage of folates in the folate metabolism (Nzila A., 2006b). Several combination therapies are being used against the folate pathway. These include Fansidar[TM] (pyrimethamine/sulfadoxine), Lapdap[®]

(chlorproguanil/dapsone) and Metakelfin (pyrimethamine/sulfalene). Where drug resistance is emerging, several other combination therapies are also being utilised such as Malarone (proguanil/atovaquone) and several artemisinin-based combinations (Biagini G.A. *et al.*, 2003). Inhibition of the *de novo* folate pathway is hampered by the parasite's ability to salvage exogenous folates (Nzila A., 2006b). It has been demonstrated that the addition of exogenous folate decreases SDX inhibition but that inhibition is revived when PYR is introduced into cultures. This highlights the synergistic role of these two antimalarials (Gregson A. and Plowe C.V., 2005). The exact mechanism of this so called "folate effect" is still unclear and is currently under investigation to enable a better comprehension of resistance and drug synergy in *P. falciparum* (Gregson A. and Plowe C.V., 2005). Initial experiments proposed several mechanisms by which PYR could block exogenous folate utilisation: Firstly, by interfering with the level of folate uptake at the erythrocyte and/or parasite, or secondly, at the level of processing of the folate molecule itself via breakdown intermediates, or lastly a combination of these steps (Wang P. *et al.*, 1999). It was, however, found that the interference is not at the reduction of folic acid to DHF in the folate pathway (Wang P. *et al.*, 1999). Recent investigations into the drug synergy and folate effect of PYR/SDX indicated that DHPS activity is essential for drug synergy, regardless of the exogenous folate pool (Wang P. *et al.*, 2004b). Several reports also support the view of a possible folate reductase activity in DHFR, potentially allowing the parasite to utilise exogenous folic acid (Wang P. *et al.*, 2004a). It was recently suggested by Wang *et al.* that due to the highly polar nature of the folates as well as their dianionic nature, that salvage might utilise a mediated transport process to import these molecules against the inward negative potential (Wang P. *et al.*, 2007). Experiments demonstrated the transport of folates to be a regulated process dependent on cellular energy and the pH gradient across the plasma membrane of the parasite (Wang P. *et al.*, 2007). A proposed model for folate uptake would be mediated through a transport complex or a multi-binding protein, which would involve a proton-symport channel. Three possible candidate proteins for folate transporters were predicted that are related to other putative transporters and are currently being investigated (Wang P. *et al.*, 2007). Thus, the full elucidation of the molecular mechanism underlying drug synergy and the possibility of a folate transporter may allow a unique approach to the inhibition of the exogenous folate pathway (Wang P. *et al.*, 1999; Wang P. *et al.*, 2007).

1.4.3 Putative novel drug targets within the folate pathway

As mentioned, the only two activities currently being targeted in the folate pathway are DHFR and DHPS. Taking into account that the host cannot synthesise folates *de novo*, the other uncharacterised folate metabolising enzymes are promising drug targets. Three novel *P.*

falciparum genes within the *de novo* as well as salvage folate pathway have been identified but their respective enzymes have not been characterised (Lee C.S. *et al.*, 2001). The three genes identified include; GTP cyclohydrolase (GTP-CH, E.C: 3.5.4.16), serine hydroxymethyltransferase (SHMT, E.C: 2.1.2.1), and the bifunctional dihydrofolate synthase-folylpolyglutamate synthase (DHFS-FPGS EC: 6.3.2.12 and EC: 6.3.2.17, respectively) (Lee C.S. *et al.*, 2001). Recently, a glycine-cleavage complex (GCV) was also identified to be part of folate C1 metabolism, opening a new avenue for inhibition studies (Salcedo E. *et al.*, 2005). GTP-CH is the first enzyme in the *de novo* synthesis pathway (Figure 1.7). It catalyses the conversion of GTP to dihydroneopterin triphosphate (DHNP) (Nzila A. *et al.*, 2005b). The second enzyme SHMT involved in the salvage pathway, catalyses the conversion of THF and serine to 5,10-methylene-tetrahydrofolate and glycine, respectively (Figure 1.7) (Nzila A. *et al.*, 2005b). The conversion is essential as a source of C1 units, thus the inhibition of SHMT would affect cell growth dramatically (Nzila A. *et al.*, 2005b). The conversion of serine to glycine is important for the parasite as glycine is used in the mitochondria as a donor of C1 units (Salcedo E. *et al.*, 2005). Lastly, DHFS-FPGS was identified to be a bifunctional protein involved in the addition of L-glutamate moieties to the folates in both the *de novo* and salvage pathways (Lee C.S. *et al.*, 2001). The following section will look at the importance of this enzyme as a drug target.

1.4.4 Bifunctional dihydrofolate synthase-folylpolyglutamate synthase (DHFS-FPGS) as drug target

The DHFS-FPGS enzyme has closely related catalytic functions but distinct roles in C1 metabolism. DHFS is essential in the *de novo* pathway for the formation of the pteroate moiety, while FPGS alters the length of the glutamate tails of the folate molecules in the salvage pathway (Salcedo E. *et al.*, 2001). This is important for the regulation of intracellular folate pools by retaining folates in the cell. Folylpolyglutamates are improved cofactors and substrates compared to the folate moieties due to the increased glutamate tail (Ouelette M. *et al.*, 2002; Stokstad E.L.R. and Koch J., 1967). The DHFS-FPGS of *P. falciparum* is the first bifunctional protein identified outside of the bacterial kingdom to encode both DHFS and FPGS activities in a single gene. The DHFS-FPGS gene resembles the *E. coli* gene, where the activities are dependent on the same set of residues that are distributed throughout the molecule, unlike the other bifunctional proteins of the folate pathway, which have distinct domains for the two activities (Salcedo E. *et al.*, 2001). This enzyme provides a good target for antifolates, as a single inhibitor would have an effect on both the *de novo* and salvage pathway, arresting cell growth and differentiation of the parasite (Salcedo E. *et al.*, 2001).

1.4.4.1 Structural models of DHFS-FPGS

The only crystal structures currently available for homology modelling of the *P. falciparum* DHFS-FPGS, are those of *Lactobacillus casei* FPGS (Sun X. *et al.*, 1998), the FolC (bifunctional DHFS-FPGS) enzyme of *Escherichia coli* (Mathieu M. *et al.*, 2005) and the FPGS from *Thermotoga maritima* (PDB code 1o5z) (Smith C.A. *et al.*, 2006). The *L. casei* crystal structure revealed that the enzyme consist of a two domain architecture linked by a six-residue linker (Figure 1.8) (Sun X. *et al.*, 1998). A mononucleotide binding fold (P-loop), similar to those found in proteins of the adenylate and uridylate kinase family, is found at the N-terminal domain. The C-terminal domain, is very similar to the folate-binding DHFR enzyme (Sun X. *et al.*, 1998). The active site is located between these domains, the classical Mg^{2+} site for ATP binding and hydrolysis are formed by a highly conserved glutamate residue in the P-loop (Smith C.A. *et al.*, 2006; Sun X. *et al.*, 1998). Another important feature is the 10-amino acid Ω-loop, which is essential for interdomain stabilisation through hydrophobic interactions with the C-terminal domain and has an important role in binding of K^+ cations (Sun X. *et al.*, 1998). The proline at the N-terminal of the Ω-loop directs a carbonyl group towards the classical Mg^{2+} binding site adopting a *cis* configuration (Smith C.A. *et al.*, 2006). The Ω-loop is highly conserved in all FPGS sequences (refer to Figure 1.9) and has a consensus sequence of T/S-**SP**-H(Y/F)-xxxx-N/R-**ER** (bold residues indicate complete conservation in all known FPGS sequences) (Smith C.A. *et al.*, 2006). The crystal structure of *E. coli* FolC showed the same structural features (Figure1.8) (Mathieu M. *et al.*, 2005).
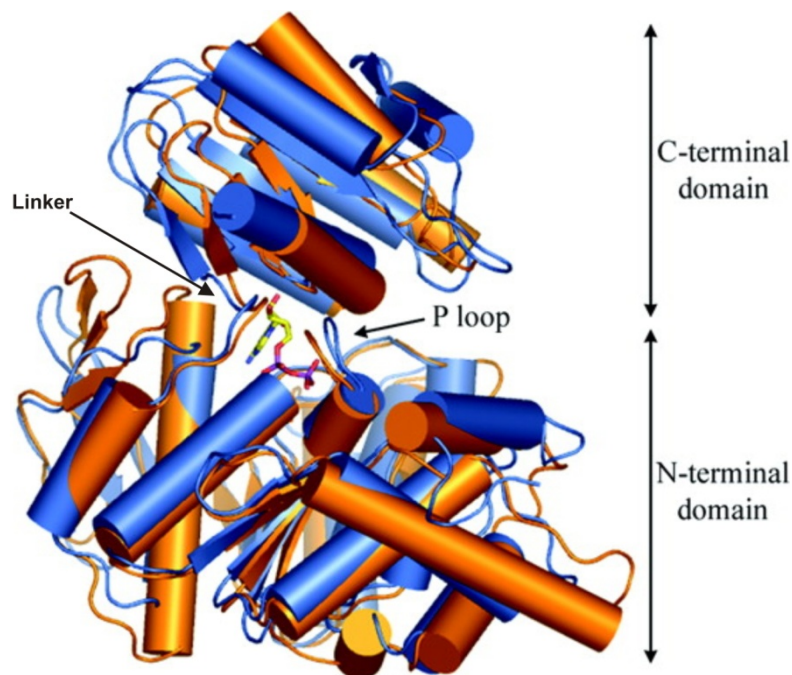
Figure 1-8: Superimposed structure of *L. casei* FPGS (orange) and *E. coli* FolC (blue). The yellow structure in the interdomain cleft represents the ADP molecule (Mathieu M. *et al.*, 2005)

The sequential binding of ATP and folate have been investigated to define FPGS activity (Sun X. *et al.*, 2001). The crystal structures of two substrate-bound FPGS molecules showed that conformational changes between the N-terminal ATPase domain and the C-terminal domain occur only in response to folate binding (Sun X. *et al.*, 2001). It was suggested that the Mg-ATP complex of the enzyme may represent an inactive form, where the initial binding of ATP is non-productive. Folate binding induces a conformational change to allow the productive binding of ATP to occur, specifically when the ATP inserts its adenosine moiety into the deep cleft between the N- and C-terminal domains in a structurally favourable *anti* configuration (Sheng Y. *et al.*, 2003; Sun X. *et al.*, 2001). Based on these domain interactions, a putative L-glutamate binding site was identified, suggesting that the repositioning of the three substrates, Mg^{2+}, ATP and folate might provide a positively charged platform for the binding of L-glutamate (Sun X. *et al.*, 2001). Even though the ATP is bound non-productively, it was shown that the ATP could induce a conformational change in FPGS. Binding of THF is essential for catalysis as it induces a distinct conformational change bringing the domains closer together, which is needed for the formation of the binding pocket for glutamate (Sheng Y. *et al.*, 2003). These binding studies suggest that the THF might be the first substrate to be bound questioning the kinetic Ter-Ter mechanism, which proposes the binding of ATP first followed by THF and glutamate respectively (Sheng Y. *et al.*, 2003). The glutamate binding site further suggests that the species-dependent addition of glutamate moieties may depend on the ability of the enzyme to accommodate domain movements to a different extent in the linker region between the N- and C-terminal domains (Sun X. *et al.*, 1998; Sun X. *et al.*, 2001). The docking studies revealed a new distinct binding site for folate substrates and it was suggested that this pocket is important for the initial addition of the first glutamate residue (Tan X-J. and Carlson H.A., 2005). This binding pocket correlated well with the crystal structure of *E. coli* FolC, which shows a binding pocket for the hydropteroyl-phosphate (Mathieu M. *et al.*, 2005; Tan X-J. and Carlson H.A., 2005). Sequence similarities of DHFS, FPGS and FolC also suggest that both the catalytic site and ATP binding site are conserved at the structural level (Mathieu M. *et al.*, 2005). This smaller FolC pocket also indicates the reason for species specific glutamate addition, as *L. casei* adds up to nine glutamate residues compared to three residues in *E. coli* (Salcedo E. et al., 2001). These conserved residues help accommodate the glutamate tail in the catalytic centre, while the non-conserved residues may be responsible for the binding site's specificity and shape (Tan X-J. and Carlson H.A., 2005).

Thus, due to the high drug resistance of malaria to current antimalarials, targets such as DHFS-FPGS provide new possibilities for rational drug design. Understanding of the enzyme catalytic mechanism would ensure better antifolate development to prevent drug-resistance.

To date, *P. falciparum* DHFS-FPGS has been characterised only in terms of its DNA sequence, the predicted amino acid sequence and functional complementation in *E. coli* (Coetzee L., 2003; Lee C.S. *et al.*, 2001; Salcedo E. *et al.*, 2001). Overall, alignments between the *P. falciparum* DHFS-FPGS and other homologues from *Homo sapiens*, *E. coli* and *L. casei* indicated ~30% similarity, and ~17% identity (Lee C.S. *et al.*, 2001). The presence of the ATP-binding P-loop "GTNGKGX" (Figure 1.9, 130-138), interdomain stabilising Ω loop (Figure 1.9, 163-173), which forms the active site, as well as the FPGS signature sequence (Figure 1.9, 241-256) and selected residues required for catalytic activity were confirmed as present in *P. falciparum*. Mutations of these essential residues causes a loss in function in *L. casei* FPGS (Coetzee L., 2003; Lee C.S. *et al.*, 2001) as well as abolishment of the activity due to conformational changes in the Ω loop (Smith C.A. *et al.*, 2006) (Figure 1.9, stars and black dots, respectively).

```
                      10        20        30        40        50        60        70        80        90       100
Q8L1G2_str   --------------------------------------------------------------------MTYEETLEWIHD--------HLVFG
FOLC_BACSU   -------------------------------------------------------------------MFTAYQDARSWIHG--------RLKFG
FOLC_LACCA   -------------------------------------------------------------------MNYTETVAYIHS--------FPRLA
FOLC_HAEIN   --------------------------------------------------------------MNNMQLKATSPLAEWLSYLEK--------SHFKP
FOLC_ECOLI   --------------------------------------------------------------MIIKRTPQAASPLASWLSYLEN--------LHSKT
FOLC_BUCAI   -------------------------------------------------------------MINKN----YSLSLWLKYLEQ--------LDKKR
FOLC_BUCAP   -------------------------------------------------------------MHKKK----YTFSMWMKYLEK--------FDKKD
FOLC_BUCBP   -------------------------------------------------------------MINKKLARSFSLYEWLYYLDH--------FMLDN
Arabidopsi   ----------------------------------------------------------------------------------------------
FOLC_HUMAN   -----MSRARSHLRAALFLAAASARG------------------ITTQVAARRGLSAWPVPQEPSMEYQDAVRMLNTLQTNAGYLEQVKRQRG---DPQT
FOLC_MOUSE   -----MSWARSRLCSTLSLAAVSARG------------------ATTEGPARRGMSAGPAPQEPGMEYQDAVRTLNTLQTNASYLEQVKRQRS---DPQA
FOLE_CANAL   -----MNQTTETDSMRINLQ-------------------------RTYKDAINALNSLQSNFASIEATKKLGPSVNRNEL
FOLE_YEAST   MHKGKKNYPNLITSFRMNLKKIILNHDRFSHPERWKTNALLRFTFVYIKFLFDLMIIKNPLRMVGKTYRDAVTALNSLQSNYANIMAIRQTGD--RKNTM
FOLE_NEUCR   -----MHHVLRPIAFRLALVSPLRS-------------------LTITHHHLFFTKRTMASSARTYNDAIDALNSLQTPFAVIEARRKAGI--RPDAH
FOLD_YEAST   ----------------------------------------------------------------------------------------------MA
pfalc        ------------------------MEKNQNDKSNKNDIIHMNDKSGNYDKNNINNFIDKNDEHDMSDILHKINNEEKKYEEIKSYSECLELLYKTHA
```

P -loop   Ω -loop

```
                     110       120       130       140       150       160       170       180       190       200
Q8L1G2_str   IKPGLKRMLWVLGQLGNPQKNVK-GVHIV GTNGKGS TVNHLQHIFTTAG------YEVGTFT SPYIMDFKERI SINGRMISEKDLVIAANRIRPLTERLV
FOLC_BACSU   VKPGLGRMKQLMARLGHPEKKIR-AFHVA GTNGKGS TVAFIRSMLQEAG------YTVGTFT SPYIITFNERI SVNGIPISDEEWTALVNQMKPHVEALD
FOLC_LACCA   KTGDHRRILTLLHALGNPQQQGR-YIHVT GTNGKGS AANAIAHVLEASG------LTVGLYT SPFIMRFNERI MIDHEPIPDAALVNAVAFVRAALERLQ
FOLC_HAEIN   IDLGLDRIKSVAEKLDLLHPVPY-VITVG GTNGKGT TCRLLETILLNHG------LRVGVYS SPHLLRYNERI VRIQNQDLPDEAHTASFAFID-------
FOLC_ECOLI   IDLGLERVSLVAARLGVLKPAPF-VFTVA GTNGKGT TCRTLESILMAAG------YKVGVYS SPHLVRYTERI VRVQGQELPESAHTASFAEIE-------
FOLC_BUCAI   IY-NLTELKFLAKKLGLLKSESF-IFTVA GTNGKGT TCAVLERLLLDSG------YQVGLYT SPHLINFVERI VRINGFVLHEEEHIDSFQNVE-------
FOLC_BUCAP   RK-NLFELKLIAKKLGLLNLKSF-FFTVG GTNGKGT TCAMLEKLLLDSG------YQVGLYT SPHLINYSERI KVNGLYLSEKDHIFSFLIID-------
FOLC_BUCBP   IDPTLNRVFYVAKKLGVLKSKAF-VFIVG GTNGKGS TCHVLENLLLNSG------YRVGLYT SPHLMRYTERI VRINGFELEHLYHISAFNDVK-------
Arabidopsi   -------------------------------------------------------------MFT SPHLIDVRERF RIDGLDISEEKFLQYFWECWKLLKEK-
FOLC_HUMAN   QLEAMELYLARSGLQVEDLDRLN-IIHVT GTKGKGS TCAFTECILRSYG------LKTGFFS SPHLVQVRERI RINGQPISPELFTKYFWRLYHRLEET-
FOLC_MOUSE   QLEAMEMYLARSGLQVEDLNRLN-IIHVT GTKGKGS TCAFTERILRNYG------LKTGFFR SPHMVQVRDRI RINGKPISPELFTKHFWCLYNQLEEF-
FOLE_CANAL   SINEVHEFTKRLGYTPTDFNKLN-IIHIT GTKGKGS TCAFTESILKQY-----TISKIGLYT SPHLKSVRERI RINGQPINQEKFAKYFFEVWDKFTTTK
FOLE_YEAST   TLLEMHEWSRRIGYSASDFNKLN-IVHIT GTKGKGS TAAFTSSILGQYK---EQLPRIGLYT SPHLKSVRERI RINGEPISEEKFAKYFFEVWDRLDSTT
FOLE_NEUCR   SVKEMRAYLARIGYSSQDLDRLN-IVHVA GTKGKGG TCAFVDSILTRHQRTHGIPRRIGLFT SPHLIAVRERI RIDSKPISEELFARYFFEVWDRLETSQ
FOLD_YEAST   IELGLSRITKLLEHLGNPQNSLR-VLHIA GTNGKGS VCTYLSSVLQQKS------YQIGKFT TPHLVHVTDSI TINNKPIPLERYQNIRLQLEALN----
pfalc        LKLGLDNPKKLNESFGHPCDKYK-TIHIA GTNGKGS VCYKIYTCLKIKK------FKVGLFS SPHIFSLRERI IVNDEPISEKELIHLVNEVLN------
                                         ● ●                         ●
                                                                     *
```

24

FPGS signature seq

```
                 210       220       230       240       250       260       270       280       290       300
Q8L1G2_str  R------ETDFGEV---TEFEVITLIMFLYFGDMHPVDIAIIEAGLGGLYDSTNVFQ---AMVVVCPSIGLDHQAILGETYADIAAQKAGVLEGG-----
FOLC_BACSU  -------QTEYGQP---TEFEIMTACAFLYFAEFHKVDFVIFETGLGGRFDSTNVVE---PLLTVITSIGHDHMNILGNTIEEIAGEKAGIIKEG-----
FOLC_LACCA  Q------QQADFN---VTEFEFITALGYWYFRQRQ-VDVAVIEVGIGGDTDSTNVIT---PVVSVLTEVALDHQKLLGHTITAIAKHKAGIIKRG-----
FOLC_HAEIN  ---------ENK-TESLTYFEFSTLSALHLFKQAK-LDVVILEVGLGGRLDATNIVD---SHLAVITSIDIDHTDFLGDTREAIAFEKAGIFREN-----
FOLC_ECOLI  ---------SARGDISLTYFEYGTLSALWLFKQAQ-LDVVILEVGLGGRLDATNIVD---ADVAVVTSIALDHTDWLGPDRESIGREKAGIFRSE-----
FOLC_BUCAI  ---------LVRNGVLLTYFEFITLAALILFKRYS-LDCIILKVGLGGRLDATNIID---SDISIITNIGIDHTSILGRDRISIAREKCGVFRKN-----
FOLC_BUCAP  ---------AEKGNVSLTYFEFITLSALFLFSQYS-LDIIILEVGLGGRLDATNIID---SDLSVITNIGIDHTSCLGTDRISIGREKSGIFRKG-----
FOLC_BUCBP  ---------YFQNDVLLTRFEFITLSALILFKSYN-LDIIILEVGLGGRLDATNILS---ADVSVITNIDIDHSKILGVNRSSISVEKSGIFRKN-----
Arabidopsi  --------AVD-GLTMPPLFQFLTVLAFKIFVCEK-VDVAVIEVGLGGKLDSTNVIQK--PVVCGIASLGMDHMDILGNTLADIAFHKAGIFKPQ-----
FOLC_HUMAN  --------KDGSCVSMPPYFRFLTLMAFHVFLQEK-VDLAVVEVGIGGAYDCTNIIRK--PVVCGVSSLGIDHTSLLGDTVEKIAWQKGGIFKQG-----
FOLC_MOUSE  --------KDDSHVSMPSYFRFLTLMAFHVFLQEK-VDLAVVEVGIGGAFDCTNIIRK--PVVCGVSSLGIDHTSLLGDTVEKIAWQKGGIFKPG-----
FOLE_CANAL  SDPQECPTLQPCDQVKPMYFKYLTILSFHVFLQEG-VDTAIYEVGVGGTYDSTNIIDK--PTVTGISALGIDHTFMLGNNIASITENKTGIFKKG-----
FOLE_YEAST  SSLDKFPHMIPGS--KPGYFKFLTLLSFHTFIQED-CKSCVYEVGVGGELDSTNIIEK--PIVCGVTLLGIDHTFMLGDTIEEIAWNKGGIFKSG-----
FOLE_NEUCR  LAKD---EVELGS--KPIYARYLTLMSYHVYLSEG-VDVAIYETGIGGEYDATNVVDR--PVVSGISTLGIDHVFVLGDTVDKIAWHKAGIMKTG-----
FOLD_YEAST  ----------KSHSLKCTEFELLTCTAFKYFYDVQ-CQWCVIEVGLGGRLDATNVIPGANKACCGITKISLDHESFLGNTLSEISKEKAGIITEG-----
pfalc       --------KAKKLYINPSFFEIITLVAFLHFLNKK-VDYAIIETGIGGRLDATNILTK--PEVIVITSIGYDHLNILGDNLPIICNEKIGIFKKD-----
                                              *                                                   *
```

```
                 310       320       330       340       350       360       370       380       390       400
Q8L1G2_str  -ETLVFAVENP--SAREVFLTKAEQVGASIWEWQEQFQMAENASG-----YRFTSPLGVISDIHIAMPGHHQVSNAALAIMTCLTL--KDRYPRLTSD--
FOLC_BACSU  -IPIVTAVTQP--EALQVIRHEAERHAAPFQSLHDACVIFNEEALPAGEQFSFKTEEKCYEDIRTSLIGTHQRQNAALSILAAEWLN-KENIAHISDE--
FOLC_LACCA  -IPVVTGNLVP--DAAAVVAAKVATTGSQWLRFDRDFSVPKAKLHGWGQRFTYEDQDGRISDLEVPLVGDYQQRNMAIAIQTAKVYAKQT-EWPLTPQ--
FOLC_HAEIN  -CPVVIGEPNV--PQTMLDQAEKLHCQVARRDVDWLFEQ---------NAENWQWQNKKVRLENLPFCQ-IPLANAATVLAAVQYLP-----FDISEQ--
FOLC_ECOLI  -KPAIVGEPEM--PSTIADVAQEKGALLQRRGVEWNYSV---------TDHDWAFSDAHGTLENLPLPL-VPQPNAATALAALRASG-----LEVSEN--
FOLC_BUCAI  -KISVIGETDI--PCSMYQIAKEKKTILKKIDIDWSWEK---------KRNYWNFFHSTIQLYNLPETQ-VPLSSAATALSTLYYSR-----FKIKEK--
FOLC_BUCAP  -KIAVIGEKNI--PISVDEIAKEKKTILKKIDVDWFWTRT--------KIDSWNFIHSNIELYNLPVSR-IPLSNTAIALASLFYSG-----LEVNLK--
FOLC_BUCBP  -KIAIVADNNF--PKVAQYLAKKKKVRLRIVNIDWIYKK---------IEFEWSFCSSKITWLHLPLPRNVSLDSVATALSAVSESG-----IKINQK--
Arabidopsi  -IPAFTVPQLS--EAMDVLQKTANNLEV--TIVAPLEPKKLD-----GVTLGLSGDHQLVNAGLAVSLSRCWLQRTGNWKKIFPNESKE---TEIPVA--
FOLC_HUMAN  -VPAFTVLQPE--GPLAVLRDRAQQISCPLYLCPMLEALEEGGP---PLTLGLEGEHQRSNAALALQLAHCWLQRQDRHGAGEPKASRPGLLWQLPLAPV
FOLC_MOUSE  -VPAFTVVQPE--GPLAVLRDRAQQIGCPLYLCPPLEALEEVGL---PLSLGLEGAHQRSNAALALQLAHCWLERQDHQDIQELKVSRPSIRWQLPLAPV
FOLE_CANAL  -VPAFVSRQLEYPETHELIEKRAKQLGVSSLEFVDTEDLP-------NVKLGLSGEFQKQNAALAIRIANSHLKTIGITQDLPEFNNNDGKIKKLSNK--
FOLE_YEAST  -APAFTVEKQP-PQGLTILKERAEERKTTLTEVPPFKQLE-------NVKLGIAGEFQKSNASLAVMLASEILHTSNILEEKIKCSSN----ASIPEK--
FOLE_NEUCR  -SPAFTIEQVP--SATQVLKDRAVEKGVDLKIPDVDPRLN-------GVKIRPDAVFQKKNATLAIALAETALKKL-----DPSFKPG---TDSLSPE--
FOLD_YEAST  -VPFTVIDGTNEASVINVVKERCKALGSELSVTDSQLNGN----------MIDTNSWGCFDLAKLPLNGEYQIFNLRVAMGMLDYLQMN-ELIDITKN--
pfalc       --ANVVIGPSVAIYKNVFDKAKELNCTIHTVVPEPRGERYNEENSRIALRTLEILNISIDYFLKSIIPIKPPLRIQYLATEQIQHIKKKFSPDNLEHN--
```

25

linker

```
                    410       420       430       440       450       460       470       480       490       500
Q8L1G2_str  -----HIREGLANSLWLGRTELLAPN------------------LMIDGAHNNE----SVAALVAVLKNNYND----KKLHILFGAI--DTKPIADMLV
FOLC_BACSU  -----ALRSGLVKAAWRGRLELVQEHP------------------PVYLDGAHNEE----GVEKLAETMKQRFAN----SRISVVFSAL--KDKPYQNMIK
FOLC_LACCA  -----NIRQGLAASHWPARLEKIS-DT----------------PLIVIDGAHNPD----GINGLITALKQLFS-----QPITVIAGIL--ADKDYAAMAD
FOLC_HAEIN  -----TLRKSLQEVELVGRFQAIKTDKREKLADYLGVPVE-TLPTIIVDVGHNPH----AAKYLSEKLTALKRSI--EGKMIAVCGML--KDKDANGVFE
FOLC_ECOLI  -----AIRDGIASAILPGRFQIVSES------------------PRVIFDVAHNPH----AAEYLTGRMKALPK----NGRVLAVIGML--HDKDIAGTLA
FOLC_BUCAI  -----IIRKSISNVQLPGRFQVISTF------------------PYIIVDVAHNPN----AAFYLSQKIDEINI----TGKIYAVVGIL--KDKDILGIID
FOLC_BUCAP  -----KLKSSISKVQLSGRFQTVFNS------------------PRIILDVAHNVH----AALYLSEKIDEIDT----EGQIYAVFGIL--KDKDVAGVVQ
FOLC_BUCBP  -----VFRSCISEITLCGRFETISYN------------------PIIILDVAHNPH----SARYLFKKMSSFKK----NGNIFAVVGIL--KEKNIKDIVS
Arabidopsi  ------FCRGLATARLHGRAQVVHDVVSDPQDSSDSMETPCGDLIFYLDGAHSPESMEACGRWFSSAVRGDKS-----LSTAILLFNCM--EVRDPQVLLP
FOLC_HUMAN  FQPTSHMRLGLRNTEWPGRTQVLRR-----------------GPLTWYLDGAHTASSAQACVRWFRQALQGRERPSGGPEVRVLLFNAT--GDRDPAALLK
FOLC_MOUSE  FRPTPHMRRGLRDTVWPGRTQILQR-----------------GPLTWYLDGAHTTSSVQACVHWYRQSLERSKRTDGGSEVHILLFNST--GDRDSAALLK
FOLE_CANAL  ------FIKGLENVDWPGRCQIINNNPTG--------------ITWYIDGAHTIESINASSTWFKQEQIKLEKP----KRRALLFNQQ--GRENYAELLE
FOLE_YEAST  ------FIIGLQNTKWEGRCQVLEKGKN---------------VWYIDGAHTKDSMVAASTWFR-DMVRLSK-----RKKILLFNQQ--SRD-----AN
FOLE_NEUCR  ------FVQGLEQVVWRGRCEVKEEDQA---------------VWHLDGAHTVDSLKVAGRWFVEECVKKAKG----GPKVLIFNQQ--GRSEAVDFLD
FOLD_YEAST  -----EVSTRLAKVDWPGRLYRMDYRFDKVSN---------RTVPILMDGAHNGS----AAVELVKYLRKEYGN----QPLTFVMAVT--HGKNLEPLLQ
pfalc       ------VQYPLAVILDVGHNETAIDRLCT-----------DINYFHKGQNIRICISITKPRNLSVFHPFIAQFGDTLKDIFYLPSLNERTYDFEEIVE
                                                               *  *
```

```
                    510       520       530       540       550       560       570       580       590       600
Q8L1G2_str  ALEQ-------------------IG-DLQVTSFH-YPN---AYPLEKCPERFGR--VADFKD-------------------------------------
FOLC_BACSU  RLET------------------IAHAIHFASFD-FPRASLAKDLYDASEISNKSWSEDPDD-------------------------------------
FOLC_LACCA  RLTA------------------AFSTVYLVPVPGTPRALPEAGYEALHEGRLKDSWQEALA-------------------------------------
FOLC_HAEIN  HLTP-------------------IIDEWHCVTLGGYRGQSGDELVEKLKSHFPNIQATSDNS-------------------------------------
FOLC_ECOLI  WLKS-------------------VVDDWYCAPLEGPRGATAEQLLEHLG------NGKSFDS-------------------------------------
FOLC_BUCAI  PLAN-------------------KIHHWFTAPLKTIRTATKHELKKFFPIH----NTSILKS-------------------------------------
FOLC_BUCAP  VLQK-------------------KINYWYVVNLKTNRSASINYLKKKLSLN----NALFFNN-------------------------------------
FOLC_BUCBP  PLIP-------------------IVDYWYCITLLTHRSATSSEIIKYLPNH----NSQISKN-------------------------------------
Arabidopsi  KLVTTCASSGTHFSRALFVPSMSTYNKVISGASAIPSDTRRKDLTWQFRLQRLWEKSIQGTDAG-LDHTLKPDGITALPPHDFLCGDAPQCGGPAGTPVT
FOLC_HUMAN  LLQP------CQFDYAVFCPNLTEVSSTGNADQQNFTVTLDQVLLRCLEHQQHWNHLDEEQASPDLWSAPSPEPGGS--------ASLLLAPHPPHTCSA
FOLC_MOUSE  LLQP------CQFDYAVFCPNVTEVSSIGNADQQNFTVTLDQVLLRCLQHQQHWNGLAEKQASSNLWSSCGPDPAGP--------GSLLLAPHPPQPTRT
FOLE_CANAL  KLFN-------VTYGTGSEPQIKFDDVIFTTNTTWSSGQFNSELISKNTSEDAVKKLEVQNNLSEVWRKLDGGV----------------------S
FOLE_YEAST  ALVN-------NLYSSVS-PEITFDDVIFTTNVTWKSGSYSADLVSMNTSQEDVEKLKVQESLVKNWNKIDDNR----------------------A
FOLE_NEUCR  GLCN-------TVKSADP-EGTGFSHVIFCTNVTYATTGYKKDFVNHQYNPKDIENMTQQRVFAERWSTLDPS-----------------------A
FOLD_YEAST  PLLR--------------PIDQVILTRFNNVEGMPWIHATDPEEIKDFILTQGYTKEIVIEN-------------------------------------
pfalc       MLNN--------------EEEIKNEIKELILSSSKKVGKWLAHEKQGNINEEDALKLYKRG-------------------------------------
```

```
                          610        620        630        640        650        660        670        680
Q8L1G2_str   ------FLALRKHAK------------------------------ADDFFVITGSLYFISEIRRYWKKHIEKSVLLTH------
FOLC_BACSU   ------VIKFIESKKG-----------------------------SNEIVLITGSLYFISDIRKRLK----------------
FOLC_LACCA   ----------ASLNDV-----------------------------PDQPIVITGSLYLASAVRQTLLGGKS-------------
FOLC_HAEIN   ---VMDGVCTALKSAV-----------------------------KNDVVLVFGSFHTVAEFWAVVE-----------------
FOLC_ECOLI   ---VAQAWDAAMADAK-----------------------------AEDTVLVCGSFHTVAHVMEVIDARRSGGK-----------
FOLC_BUCAI   ---IEIAYKKALILVK-----------------------------KEDAIIIFGSFLTVSE---FLSLKI---------------
FOLC_BUCAP   ---INESWQAIKKVIT-----------------------------KKDIILVFGSFFTVSE---FMSIKDRRLTLY---------
FOLC_BUCBP   ---MTVALEKIFDKVT-----------------------------NNDIVLIFGSFITVCEANKFLANKVKNFKLL---------
Arabidopsi   SSAVMPSLPLTINWL---RDCVRRNPS------------------LKLEVLVTGSLHLVGDVLRLLKR----------------
FOLC_HUMAN   SSLVFSCISHALQWISQGRDPIFQPPSPPKGLLTHPVAHSGASILREAAAIHVLVTGSLHLVGGVLKLLEPALSQ----------
FOLC_MOUSE   SSLVFSCISHALLWISQGRDPIFQPQSLPRNLLNHPTANSGASILREAAAIHVLVTGSLHLVGGVLKLLDPSMSQ----------
FOLE_CANAL   KRHVFADIETAVNYLKDLGD---------------------------KDLQVFVCGSLHLVGGFLVVLDNERD-------------
FOLE_YEAST   KTHVTASIEEANELIETLYD-------------------------EPADIFVTGSLHLVGGLLVVFDRIDVK------------
FOLE_NEUCR   NVMLIPTIEEAINKARSLVDTTE--------------------GEQKVQALITGSLHLVGGALGILEKADAL------------
FOLD_YEAST   --DLHQVLPSLAHVSDE----------------------------QRRPIVVCGSLYLCGELLRIHNSHLRN------------
pfalc        --CIPLIIKNAFLECCK----------------------------DNSILLVCGTFFVFDEVLNVFDIHSDMQDTIFMNEPSLV
                                                                      **
```

Figure 1-9: Alignment of *P. falciparum* DHFS-FPGS with homologous proteins. Q8L1G2_str: *Streptococcus* DHFS, FOLC_BACSU: *B. subtillus* DHFS-FPGS, FOLC_LACCA: *L. casei* FPGS, FOLC_HAEIN: *H. influenzeae* FPGS, FOLC_ ECOLI: *E. coli* DHFS-FPGS, FOLC_BUCAP: *B. aphidicola* DHFS-FPGS, FOLC_BUCBP: *B. biphidicola* DHFS-FPGS, Arabidopsi: *A. thaliana* FPGS, FOLC_HUMAN: human FPGS mitochondrial precursor, FOLC_MOUSE: mouse FPGS mitochondrial precursor, FOLE_CANAL: *C. albicans* FPGS, FOLE_YEAST: yeast FPGS, FOLE_NEUCR: *N. crassa* FPGS, FOLD_YEAST: yeast DHFS. p.falc: *P. falciparum* DHFS-FPGS. Thick black lines indicate areas of low complexity in the *P. falciparum* sequence. Stars indicate essential residues in the *L. casei* enzyme. Black dots indicate essential residues in *L. casei* for the conformation of the Ω loop. Boxed areas indicate the P-loop, Ω-loop, FPGS signature sequence and *L. casei* linker features. The grey highlighted residues indicate conserved hydrophobic residues flanking the linker region of *L. casei.* Neutral amino acids are coloured green, negatively charged amino acids are coloured different shades of red according to pKa values, positively charged amino acids are coloured shades of blue according to pKa values, proline and glycine normally found in turns are coloured yellow and purple respectively (Coetzee L., 2003).

A superimposed homology model of *P. falciparum* DHFS-FPGS (red ribbons) on the *L. casei* FPGS crystal structure (green ribbons) is indicated in Figure 1.10. The yellow loop indicates the *P. falciparum* linker-area in the alignment (Coetzee L., 2003).
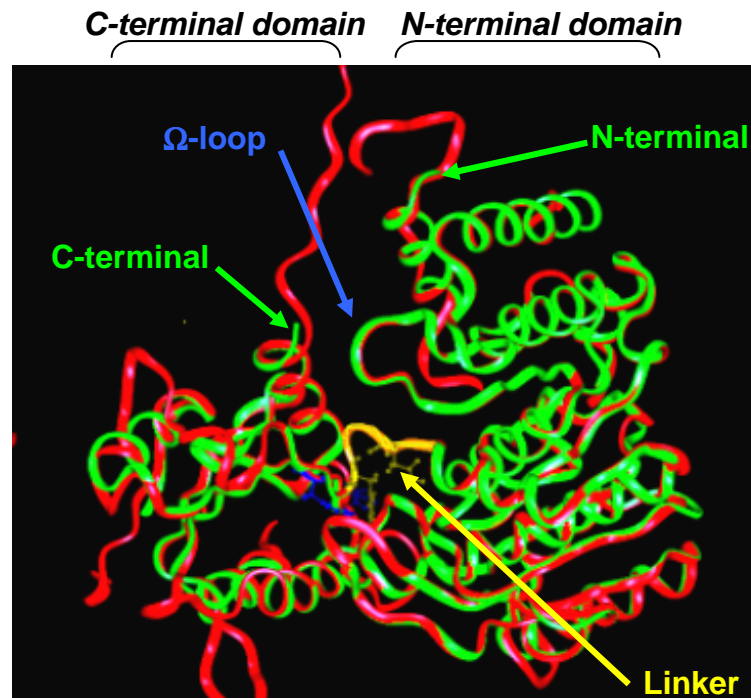
Figure 1-10 Superimposed ribbon backbones of the preliminary homology model of *P. falciparum* DHFS-FPGS (red) on the *L. casei* FPGS (green) crystal structure. The aligned Ω-loop (blue arrow and linker area (yellow loop) are indicated. Catalytic residues (D449 and H552) that do not align are indicated in blue ball and stick structures (*L. casei*) and yellow ball and stick structures (*P. falciparum)* (Coetzee L., 2003).

1.5 Research aims

The primary aim of this study is to functionally express sufficient quantities of soluble *P. falciparum* DHFS-FPGS for kinetic assays.

- **Chapter 2** focuses on the heterologous recombinant expression of *P. falciparum* DHFS-FPGS from the synthetic *dhfs-fpgs* gene and the investigation into several solubilisation protocols for improved soluble expression.
- **Chapter 3** focuses on the verification of enzymatic activity by means of a radioactivity-based assay as well as the determination of the kinetic properties of P*. falciparum* DHFS-FPGS.
- **In Chapter 4** the relationship between the above chapters, and future perspectives are discussed.

Chapter 2
Heterologous expression of recombinant native and synthetic *P. falciparum* DHFS-FPGS

2. Introduction

E. coli remains a central tool for the heterologous expression of recombinant proteins. This is also true within the malaria research field where *E. coli* continues to be the workhorse for the expression of recombinant malaria proteins (Flick K. *et al.*, 2004). This method however, has disadvantages including expression of recombinant proteins as truncated forms, or as insoluble protein in inclusion bodies (Flick K. *et al.*, 2004). Inclusion bodies can be defined as refractile aggregates of protease resistant miss-folded protein (Carrio M.M. and Villaverde A., 2002). A major reason for this inclusion body formation is the very high A+T versus G+C content of the *Plasmodium* genome (Gardner M.J *et al.*, 2002b), which results in *P. falciparum* having different codon preferences to those of *E. coli*. *P. falciparum* also has several rare codons that are very low in frequency in *E. coli* (Cinquin O. *et al.*, 2001). An example of these are the arginine codons AGA and AGG, which are the rarest in *E. coli* genes but occur at a high frequency in *Plasmodium* (Sayers J.R. *et al.*, 1995). This excessive presence of rare codons contributes to early termination of the mRNA translation process during high level protein expression, as the tRNAs will rapidly be exhausted (Flick K. *et al.*, 2004). This exhaustion of tRNAs causes the ribosome to pause and bring about the early termination of expression (Zhou Z. *et al.*, 2004).

2.1 Improving recombinant protein expression

There are various strategies to increase soluble expression of recombinant proteins in *E. coli,* as listed in Table 2.1. Several examples are given where strategies improved expression of *P. falciparum* proteins, some of which will be examined in more detail in the following sections. These strategies include the optimisation of codon usage, high cell-density culture systems, co-expression of chaperone plasmids and refolding protocols for

the resolubilisation of recombinant proteins (Jana S. and Deb J.K., 2005; Sorensen H.P. and Mortensen K.K., 2005).

2.1.1 Expression systems

Several essential elements can be changed to obtain increased recombinant protein expression in *E. coli*. These include using different host expression cell lines, a strong inducible promoter to control high-level gene expression and monitoring of mRNA stability (Jana S. and Deb J.K., 2005; Sorensen H.P. and Mortensen K.K., 2005). Important requirements of the expression strain comprise the absence of harmful natural proteases, expression plasmid stability and conferring the genetic elements (example: DE3) relevant to the expression plasmid (Sorensen H.P. and Mortensen K.K., 2005). Several strains are available for protein expression, of which the BL21 strains are the most commonly used. Several derivatives of this cell line can be utilised to improve expression. These include the *recA* negative strains (Novagen BLR strain) for the stabilisation of plasmids containing repetitive sequences, *lacY* mutants (Tuner series), which enable adjustable levels of protein expression and the *trzb*/*gor* negative mutants (Novagen Origami) for improved cytoplasmic disulfide bond formation (Sorensen H.P. and Mortensen K.K., 2005). Factors such as nutrient composition, temperature and pH can additionally affect protein production levels, as well as soluble protein expression. Manipulation of any of these factors could result in an increase in soluble protein expression (Jana S. and Deb J.K., 2005). Gene expression is also controlled by the decay rate of mRNA. It is suggested that host strains that are deficient in specific RNases may enhance expression by protecting the mRNA from degradation (Sorensen H.P. and Mortensen K.K., 2005). Several factors thus need to be considered to obtain the most effective soluble expression.

Table 2-1: **Strategies to improve recombinant protein expression in an *E. coli* host**

Techniques	Comments	Examples	Advantage/Disadvantage	References
Expression vector system				
Host strain	Choice of host strain impacts expression levels, changing tRNA pool of host to accommodate rare codons.	Several cell lines available from Stratagene and Novagen.	Changes in levels of tRNA might have metabolic effects on host.	Stratagene and Novagen
Plasmid	Copy number determines gene dosage. Moderate to high copy number plasmids are favoured.			(Jana S. and Deb J.K., 2005)
Co-expression plasmids	Co-transformation with plasmid able to overproduce appropriate tRNAs may increase soluble expression.	Over-expression of *Pf.* DHPS by co-transformation of RIG-plasmid.	RIG plasmid does not have effect on problems such as transcription of A+T rich genes, mRNA stability and secondary structures.	(Baca A.M. and Hol W.G.J., 2000; Sayers J.R. *et al.*, 1995)
Promoter	Several desirable features: Strong promoter, low basal expression levels (tightly regulated), easily transferable to other *E. coli* strains.	Commonly used promoters are: *lac* (*lacUV5*), *trp* and T7	Most of these promoters do have leaky expression.	(Jana S. and Deb J.K., 2005)
Codon usage				
Codon-optimisation	Utilising the optimal *E. coli* codons frequently improves protein yield.	Synthetic F2 domain of EBA-175 showed 4-fold increase in *E. coli.* Codon optimisation of FALVAC-1 gene increased 3-fold in *E. coli.* 42kDa fragment of MSP1 showed a 2-fold increased expression.	Costly and time consuming. May result in increased protein levels but not in soluble protein.	(Darko C.A. *et al.*, 2005; Yadava A. and Ockenhouse C.F., 2003; Zhou Z. *et al.*, 2004)
Translational attenuation	Effectiveness and spacing of transcription terminators are needed to prevent ribosome skipping.	These characteristics, together with codon optimisation, can be seen as codon harmonisation where the codon bias, secondary structures and ribosomal pause sites of gene of interest are taken into account.		(Kincaid R.L. *et al.*, 2002)
mRNA stability	The secondary structure of mRNA and stability impacts the yield of protein expression.			
Translational signalling	Efficiency of gene expression is influenced by secondary structure at translation initiation region of mRNA.			

	Growth conditions			
Temperature	Temperature has a pronounced effect on protein folding and stability.	Lowering of the temperature to 15-20℃ for protein expression.		(Jana S. and Deb J.K., 2005)
Medium	Nutrient composition, pH and oxygen levels could affect protein production.	Cultures in complex media produce increased amounts of protein in the absence of an inducer.	Results in increased soluble protein levels.	(Grossman T.H. *et al.*, 1998; Studier F.W., 2005)
Induction time	Growth conditions/growth rates affect protein yield.	Induction of protein expression at post-log phase instead of mid-log phase significantly increased soluble expression of PfEMP1 protein.		(Flick K. *et al.*, 2004)
	Chaperone proteins			
Small heat-shock proteins and Hsp70/90/60	Interact and stabilise partially folded target proteins to prevent aggregation. Aid correct folding and refolding of proteins during expression.	Human pro-urokinase co-expressed with plasmid containing chaperone genes showed less aggregation and more soluble expression.		(Treweek T.M. *et al.*, 2003) (Nishihara K. *et al.*, 1998)

2.1.2 Codon modification

Two different strategies have traditionally been used to avoid codon bias. Firstly, altering the codons of the target genes to the preferred codons of the expression host (codon-optimisation) (Prapunwattana P. *et al.*, 1996; Zhou Z. *et al.*, 2004). Secondly, the intracellular tRNA pool can be expanded by including plasmids, which encode for tRNA's that are rare in the expression organisms such as *E. coli* (Baca A.M. and Hol W.G.J., 2000). Recently, codon-harmonisation or translational attenuation, has been proposed to improve the classical codon-optimisation strategies (Kincaid R.L. *et al.*, 2002). Codon-harmonisation involves substitution of codons to replicate the codon frequency preference of the target gene, rather than *E. coli* frequency preferences (Figure 2.1). These changes ensure that the positional codon frequency of low/intermediate and high codons stays the same in the non-natural host (*E. coli)*, which allows the translational processes to match that of the natural host. The translational machinery would thus move and pause at the correct sites, allowing the folding of the secondary and tertiary structures to occur as it would have in the natural host (Hillier C. *et al.*, 2005; Kincaid R.L. *et al.*, 2002; Thanaraj T.A. and Argos P., 1996).

As shown in Figure 2.1, harmonisation of a *P. falciparum* gene will ensure that both the codon usage and frequency stay the same between the original *P. falciparum* and the non-natural host (*E. coli*), compared to the situation with codon-optimisation alone.
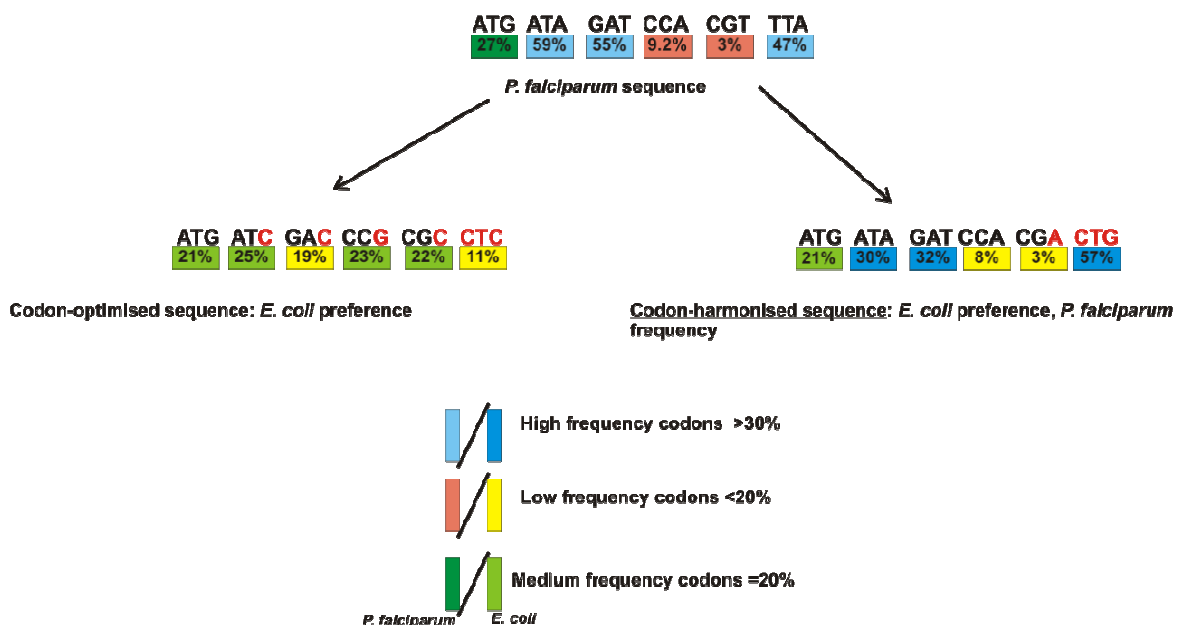
Figure 2-1: Diagram of codon-harmonisation

Narum *et al.* demonstrated that optimising codon usage of antigenic fragments from organisms whose codon usage differs from that of mammals, could improve protein expression and subsequently the immunogenicity of gene fragments (Narum D.L. *et al.*, 2001). Two leading malaria vaccine candidate antigens, EBA-175 and MSP-1 were chosen. The optimised sequence for both antigens contained a G+C content of about 56% compared to the previous content of 26%. Both optimised *P. falciparum* antigens genes led to increased protein expression levels in mammalian cells. An elevated antibody titre was also observed, which is an indication of enhanced immunogenicity (Narum D.L. *et al.*, 2001). The effect of codon-optimisation was also compared for the F2 domain of EBA-175 of *P. falciparum* by adapting the synthetic gene to *E. coli,* as well as *Pichia pastoris* codon usage (Yadava A. and Ockenhouse C.F., 2003). The synthetic genes produced improved protein expression levels but recombinant protein expression in *E. coli* was found to be insoluble. This was thought to be due to the high speed of the translational machinery resulting in incorrectly folded protein. In contrast, the optimised *P. pastoris* genes showed very little aggregation with enhanced expression of correctly folded protein. This is due to its ability to produce complex molecules, as well as the formation of disulfide bonds. Codon-optimisation was recently used to improve the yield and purity of a recombinant blood stage vaccine for *P. vivax* (Yazdani S.S. *et al.*, 2006). Due to the receptor-binding domain of *P. vivax* Duffy binding protein (PvRII) eliciting high antibody titres it was used in a recombinant expression system. A codon-optimised synthetic gene was designed to increase the low levels of native PvRII expression and reduce the levels of truncated fragments being observed (Yazdani S.S. *et al.*, 2006). Results indicated a 2-fold increase of protein expression with the synthetic construct. The synthetic construct yielded only the full length PvRII. The authors suggested that increased expression was due to two reasons: the abundance of tRNA for the optimised codons, as well as the destabilisation of the secondary structures at the translation initiation site.

Codon-harmonisation instead of codon-optimisation was used on a *P. falciparum* liver stage antigen to obtain a protein-based vaccine against this pre-erythrocyte protein (Hillier C. *et al.*, 2005). This approach resulted in significantly enhanced efficiency of protein expression in *E. coli* compared to previous codon-optimisation experiments, as well as to experiments of protein co-expressed with a rare frequency tRNA plasmid (Hillier C. *et al.*, 2005). The field of codon modification involving both codon-harmonisation as well as codon-optimisation is immense, but provides a good alternative, compared to codon-optimisation alone, for increasing the expression of problematic proteins.

2.1.3 Auto-induction

As mentioned previously, *E. coli* is the preferred organism for recombinant expression of proteins, but this system has a few drawbacks, one of which is the leaky (uninduced) expression of protein when the lactose operon is used. The *lac* operon consists of three structural genes which encode for an enzyme responsible for the hydrolysis of lactose, as well as a permease for the transport of lactose across the cell wall. The operon is regulated by binding of the *lac* repressor to the operator site in the absence of an inducer or lactose. This operon is widely used in systems such as the pET vector system where isopropyl-β-D-thiogalactopyranoside (IPTG) is used as a synthetic inducer. This stimulates the transcription of the *lac* operon and subsequent protein expression by binding to the *lac* repressor (Novagen, EMD Biosciences Germany). Several expression systems are known to produce some levels of leaky expression. It was shown that leaky expression increases markedly when cells are grown in complex auto-inducing media, and could be a means to improve expression levels without inducers present (Grossman T.H. *et al.*, 1998; Studier F.W., 2005). It was concluded that the leaky expression was to some extent regulated by nutrient availability as the levels of leaky expression were influenced by medium composition and cell growth stage. The authors also indicated that cyclic AMP, acetate and a low pH are necessary to effect high-level expression in the absence of the IPTG inducer (Grossman T.H. *et al.*, 1998). The same analyses were performed by Studier *et al.* whose goal was to develop a reliable method for protein expression in T7 expression strains with or without induction (Studier F.W., 2005). It was found that with a sufficient amount of nutrients, the main factors influencing the expression are the continuation of a neutral pH and the availability of oxygen (Studier F.W., 2005). From these complex media studies several requirements for expression were thus found to be of importance: 1) low phosphate levels, which ensures buffering at neutral pH as higher phosphate concentrations slow down cell growth; 2) the presence of trace metals; and 3) glycerol as an improved carbon source to support cell growth since glucose catabolism causes pH to drop to acidic levels, which subsequently arrests the growth of the cells (Studier F.W., 2005). Even though several requirements are needed for the high level expression of proteins in complex inducing media, it provides a new way of increasing expression levels. These media also provides a way to increase the soluble protein expression of proteins.

2.1.4 Protein refolding

The immense drawback during high level protein expression in *E. coli* is the production of insoluble proteins in inclusion bodies. Inclusion bodies are refractile protein aggregates of

recombinant proteins that are expressed at high levels in a recombinant expression system (Carrio M.M. and Villaverde A., 2001; Carrio M.M. and Villaverde A., 2002). Due to the reversibility of protein aggregation in bacteria, several strategies are employed to isolate and refold inclusion bodies since they are a good source of reasonably pure polypeptides, which could provide protein in the native conformation after refolding (Villaverde A. and Carrio M.M., 2003). Inclusion bodies almost exclusively contain the over-expressed recombinant protein and very little host protein, ribosomal components or DNA/RNA fragments (Singh S.M. and Panda A.K., 2005). Protein refolding is a complicated procedure requiring optimisation steps for several of the factors involved. The first is the isolation of pure inclusion bodies without membranes or other contaminants, which co-precipitates during inclusion body recovery (Vallejo L.F. and Rinas U., 2004). The isolated inclusion bodies are solubilised by using high concentrations of chaotropic reagents such as urea and guanidine hydrochloride, as well as the addition of the anionic detergent sodium dodecyl sulphate (SDS). Protein is reduced by dithiotreithol (DTT) to dissociate the aggregates and prevent non-native disulphide bond formation and the metal ions are chelated by (ethylenedinitrilo)-tetraacetic acid (EDTA) to prevent the metal-catalysed oxidation of cysteines. Solubilisation of inclusion bodies result in soluble protein in its non-native conformation and needs to be refolded (Vallejo L.F. and Rinas U., 2004). Refolding of the solubilised protein is accomplished by the removal of the excess reagents by means of dilution, dialysis or filtration (Carrio M.M. and Villaverde A., 2001; Carrio M.M. and Villaverde A., 2002; Singh S.M. and Panda A.K., 2005; Villaverde A. and Carrio M.M., 2003). Sirawaraporn *et al.* showed that disaggregated and refolded *P. falciparum* DHFR from inclusion bodies yielded active protein with an activity of 20-30 units/mg. The refolded protein was present in sufficient amounts for enzyme function determinations (Sirawaraporn W. *et al.*, 1993). This indicates that refolding can be a good alternative to obtain soluble protein and allow further enzyme characterisation.

2.1.5 Chaperone proteins

Molecular chaperones (also called heat shock proteins (Hsp)) are a varied group of proteins that interact with incompletely/incorrectly folded proteins to stabilise the non-native conformation (Treweek T.M. *et al.*, 2003). Protein misfolding is exacerbated during recombinant protein expression due to the use of strong promoters and elevated inducer concentrations to increase the protein yield (Baneyx F. and Mujacic M., 2004). Chaperones help with protein folding by shielding interactive surfaces from one another, as well as from the solvent. They also accelerate isomerisation of peptidyl-prolyl bonds from uncharacteristic *cis* to *trans* conformation and allow the re-shuffling of disulfide bonds

(Baneyx F. and Mujacic M., 2004). The best characterised molecular chaperones are the DnaK-DnaJ-GrpE and GroEL-GroES systems (Thomas J.G. *et al.*, 1997). The DnaK-DnaJ-GrpE chaperones belong to the Hsp70 family and require ATP for the refolding of proteins (Treweek T.M. *et al.*, 2003). DnaK is targeted to a protein by DnaJ, which acts as a co-chaperone, and substrate binding occurs by hydrolysis of the DnaK-bound ATP. The hydrolysis of ATP closes the binding site and locks the substrate, thus forming a stable protein/DnaK complex. The release of correctly folded protein is controlled by GrpE, which catalyses an ADP/ATP exchange (Baneyx F. and Mujacic M., 2004; Treweek T.M. *et al.*, 2003; Walter S. and Buchner J., 2002). GroEL-GroES is part of the Hsp60 family of chaperones and recognise partially and incorrectly folded polypeptides by virtue of exposed hydrophobic surfaces (Treweek T.M. *et al.*, 2003; Walter S. and Buchner J., 2002). These two chaperones are stacked as homoheptameric rings, which refold protein in the central chamber by reversible GroES capping and conformational changes orchestrated by ATP binding and hydrolysis. The GroE assisted protein folding can thus be divided into three steps: 1) capture by the GroEL ring; 2) folding mediated by the conformational change in GroEL due to the binding of the GroES and ATP; 3) release of peptide after ATP hydrolysis (Baneyx F. and Mujacic M., 2004; Thomas J.G. *et al.*, 1997; Treweek T.M. *et al.*, 2003; Walter S. and Buchner J., 2002). Another very important system is the trigger factor (TF) chaperone. TF binds to ribosomes to interact with short nascent chains and exhibits a petidyl-prolyl *cis/trans* isomerase activity (Baneyx F. and Mujacic M., 2004). Other proteins that aid the folding pathway is the small heat-shock proteins (sHsps) characterised by monomeric sizes of 16-43 kDa (example: Ibp A/B) (Baneyx F. and Mujacic M., 2004). These sHsps complex with stressed proteins to prevent aggregation and precipitation and make the subsequent refolding of the protein with other chaperone systems possible (Treweek T.M. *et al.*, 2003; Walter S. and Buchner J., 2002). Therefore, chaperone proteins can be utilised together with any of the other factors mentioned in section 2.1.1 – 2.1.4 to improve soluble protein expression levels. Several of the chaperones can also be used alone, or in conjunction with each other to improve protein yields.

2.1.6 Previous attempts at the high level expression of soluble, functional *P. falciparum* DHFS-FPGS

In a previous study (Coetzee L., 2003), a full length *P. falciparum dhfs-fpgs* gene (~1573bp) was successfully constructed by re-synthesis of the gene to contain *E. coli* codon preferences (Appendix A). This codon-optimised synthetic *dhfs-fpgs* gene (referred

to in text as the synthetic gene) had a 17% reduction in A+T content and was predicted to increase the recombinant expression of the protein in *E. coli*. Table 2.2 indicates some characteristics of the synthetic gene. The ability of the synthetic gene construct to complement an *E. coli* K-12 mutant strain, SF4 (F⁻ *folC strA recA tn10: srlC*) was performed to determine the efficiency of the DHFS and FPGS activities (Bognar A.L. *et al*., 1986; Coetzee L., 2003; Salcedo E. *et al*., 2001). This specific strain has a single point mutation, (A309T) that reduces endogenous *E. coli* DHFS and FPGS activities to less than 3% of the wild type activities (Bognar A.L. *et al*., 1986). Cells therefore produce basal folate levels to overcome thymine deficiency but are auxotrophic for methionine. Glycine is added to stimulate growth, as the SF4 growth on minimal media is very poor in the absence of glycine. Complementation studies indicated that the synthetic gene construct complemented cell growth better than the native gene (Coetzee L., 2003) (Refer to Table 2.2).

Table 2-2: **Differences between the synthetic and native *dhfs-fpgs* gene constructs used for heterologous recombinant expression.** Arrows indicate levels of complementation in *E. coli*.

Synthetic *dhfs-fpgs*	Native *dhfs-fpgs*
Codon-optimised *dhfs-fpgs* gene	***dhfs-fpgs* gene**
E. coli codons (Coetzee L., 2003)	*P. falciparum* codons
Similar distribution of four nucleotides	**A+T rich**
42% G+C and **58% A+T**	27% G+C and **73% A+T**
Complementation studies	**Complementation studies**
↑ ↑ ↑	↑

In addition, the synthetic gene was subsequently used to determine recombinant expression levels of DHFS-FPGS in *E. coli*. Preliminary results of this study indicated that the maximum levels of expression occurred in BL21 DE3 Star *E. coli* cells, but that most of the protein was present in inclusion bodies (Coetzee L., 2003).

This chapter therefore describes various strategies for the optimisation of heterologous soluble expression of DHFS-FPGS in *E. coli* to determine the effectiveness of the optimised *dhfs-fpgs* gene. Comparative expression studies of DHFS-FPGS from the synthetic gene and native gene are also described*.*

2.2 Materials and methods

2.2.1 Primer design

The native *P. falciparum* *dhfs-fpgs* gene cloned into the *Pf*1 expression plasmid was obtained as a kind gift from Prof J Hyde, University of Manchester, UK (Salcedo E. *et al.*, 2001). The gene sequence was obtained from Genbank (Accession number AF161264). Primers were designed for subcloning of the insert into a pET22b vector. The Oligo 4 Primer analysis software program (Molecular Biology insights, MBI) was used to analyse the internal stability, specificity, and mispriming of all designed primers. The primer 'pfdffwd' was used for the forward priming reaction and the two reverse primers were 'pfdrev' (without the stop codon for read-through and expression of a C-terminal His_6-tag) and 'pfdffrev 2' (with the stop codon), respectively. These forward and reverse primers were designed with *Nde*I (CAT↓ATG) and *BamH*I (G↓GATCC) restriction enzyme sites, for in-frame cloning into the pET22b vector. The ATG in the *Nde*I restriction site (CAT↓<u>ATG</u>) was used as the start codon. Refer to section 2.2.9 for an illustration of the constructs. Primers were obtained from Inqaba Biotec (SA) and were dissolved in 10 mM Tris-HCl, pH 8.6 at 37℃ to a final stock concentration of 100 µM. Aliquots of primers were stored at -20℃ in silanised tubes.

2.2.2 Nucleic acid concentration determinations

Absorbance of each primer was determined spectrophotometrically at 260nm using the GeneQuant Pro UV/Vis Spectrophotometer (Amersham Biosciences, Uppsala, Sweden). The molar concentrations (pmol/µl) of primers were calculated using the manufacturer's value for oligonucleotide OD as follows:

Molar Concentration = <u>A_{260nm} x OD (ng/µl) x dilution factor</u>
Molecular mass (ng/nmole)

The DNA concentration of PCR products and plasmids were determined similarly, with one OD unit equivalent to 50 ng/µl for double stranded DNA.

2.2.3 PCR gene amplification

The native *dhfs-fpgs* gene was amplified from the *Pf*1 plasmid (Salcedo E. *et al.*, 2001) using the primers as described in section 2.2.1. The PCRs were performed in a Gene-AMP PCR System 9700 (Perkin Elmer, Coneticut, USA) in 0.2 ml thin-walled PCR tubes. The reactions contained 10 pmol of pfdfdw and pfdrev or pfdrev2, 1x Takara *ExTaq* DNA polymerase

buffer (2 mM Mg^{2+}, Proprietary mix), 2.5 U of Takara *ExTaq* DNA polymerase (TAKARA Bio Inc, Japan), and 250 µM of each dNTP in a final volume of 50 µl. *ExTaq* DNA polymerase is a polymerase consisting of *Taq* polymerase (thermostable) and *Pfu* polymerase (3' to 5' exonuclease activity) for the amplification of longer targets. The PCR profile consisted of a 5 min denaturing step at 94°C and 30 cycles of denaturation at 94°C for 30 sec, annealing at 58°C for 30 sec and extension at 72°C for 2 min. PCR products were stored at 4°C. The A' overhangs produced in a template independent fashion by *Taq* polymerase on the 3' termini of PCR products allowed subsequent direct A/T cloning (section 2.2.5.1).

2.2.4 Agarose gel electrophoresis and purification of PCR products

PCR products and digested plasmids (sections 2.2.3 and 2.2.6) diluted 1:6 with loading dye (0.025% (w/v) bromophenol blue, 30% (v/v) glycerol) were separated on a 1.5% (w/v) agarose gel (Promega, Wisconsin, USA) containing 0.5 ng/ml ethidium bromide (EtBr), in 1X TAE buffer (0.04 M Tris-acetate, 1 mM EDTA, pH 8). Samples were electrophoresed at 4-10 V/cm in a HE33 Mini Horizontal Agarose submarine unit (Hoefer Scientific Systems, USA). Bands were visualised at 312nm on a Spectroline TC-312A UV transilluminator (Spectronics Corporation, USA), as the intercalated EtBr emits fluorescence at a wavelength of 300nm (Karcher S.J., 1995). Bands were excised from the agarose gel and purified with the High Pure PCR Product purification kit (Roche, Basel, Switzerland) according to manufacturer's protocols: In short, 300 µl of binding buffer (3 M guanidine-thiocyanate, 10 mM Tris-HCl, 5% (v/v) ethanol, pH 6.6) was added for every 100 mg of agarose. The gel slices were completely dissolved at 56°C where-after 150 µl of isopropanol was added for every 100 mg of agarose and loaded onto a High Pure filter column. DNA binds to silica in the presence of the chaotropic salt guanidine thiocyanate, which is highly positively charged. This electrostatic interaction is strong enough to allow for several wash steps with a high salt buffer to wash out the contaminating primers, nucleotides and salts. The bound DNA is eluted using a low salt solution, which ensures the breaking of the interaction, by pH competition. The filter was washed twice with wash buffer (20 mM NaCl, 2 mM Tris-HCl, pH 7.5), by means of centrifugation at 16 000xg for 1 min at 22°C (Microlitre Centrifuge, Hermle, Labortechnik Germany). The DNA was eluted in a clean microcentrifuge tube with 50 µl elution buffer (10 mM Tris-HCl, pH 8.5) and collected by centrifugation as above.

2.2.5 Cloning

2.2.5.1 The pGEM® T-Easy vector system

The A overhangs of the PCR products generated by *Taq* polymerase enabled A/T cloning into the pGEM® T-Easy vector according to the manufacturers protocol (Promega, USA) (Appendix B). The A-tailed PCR product and pGEM® T-Easy vector were mixed in a 3:1 molar ratio. The quantity of insert required was determined by:

ng of insert = $\underline{\text{ng of vector x kb size of insert}}$ x molar ratio of insert / vector

kb size of vector

3 Weiss units of T4 DNA ligase and 1x T4 DNA ligase buffer (30 mM Tris-HCl pH 7.8, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP and 5% (v/v) polyethyleneglycol) were added to vector (50 ng) and the A-tailed PCR product (75 ng) in a final volume of 10 µl. This reaction was incubated on ice overnight. The ligation mixture was cloned into calcium/magnesium based- and/or calcium-competent DH5α [genotype F⁻Φ80 d*lacz* ΔM15 Δ(*lac*ZYA –argF) U169 *deo*R *rec*A1 *end*A1 *hsd*R17 (r_K^+, m_K^+) *pho*A *sup*E44 λ *thi*-1 *gyr*A96 *rel*A1] *E. coli* cells (see section 2.2.5.2 and 2.2.5.3).

2.2.5.2 Preparation of competent cells

2.2.5.2.1 Calcium/magnesium based (CCMB) method

This method employs the incubation of *E. coli* at low temperatures in the presence of calcium-chloride and magnesium-chloride to induce competence for DNA uptake (Hanahan D. *et al.*, 1991). Hanahan *et al.* suggested three roles that the divalent cations may play in transformation of bacteria: 1) they protect the negative charges on the outside of the cell as well as on the DNA thus a closer association is probable; 2) some regions of the membrane may crystalise in the presence of cations at low temperatures, which can bring about channels for DNA uptake; 3) the cations may also rearrange the lipopolysaccaharides away from the channels. This procedure was employed on DH5α and BL21 *E. coli* cell lines, first for routine cloning procedures and secondly for expression. 1 ml of an *E. coli* cell culture was grown overnight at 30℃ with shaking at ~150-200 rp m and diluted 50-fold into SOB (2% (w/v) Tryptone, 0.5% (w/v) Yeast extract, 0.05% (w/v) NaCl, 0.01 M $MgCl_2$ and 10 mM $MgSO_4$). Cells were grown at 30℃ until OD_{600nm}=0.3 (6 x 10^8 cells/ml), cooled down on ice for 10 min and collected by centrifugation at 1000xg for 15 min at 4℃ (Beckman J-6 centrifuge). The cell pellet was dissolved in 16.7 ml cold CCMB80 (80 mM $CaCl_2.6H_2O$, 20 mM $MnCl_2.4H_2O$, 10 mM $MgCl_2.6H_2O$, 10 mM potassium acetate, pH 7 and 10% (v/v) glycerol, pH 6.4) and incubated on ice for 20 min. Cells were centrifuged as before. The cell

pellet was dissolved in 4.2 ml of cold CCMB80. Competent cells were stored at -70℃ as 200 µl aliquots.

2.2.5.2.2 Calcium chloride competent cells

5 ml of an overnight *E. coli* cell culture was diluted 10-fold and grown at 37℃ until OD_{600nm} = 0.4 (8 x 10^8 cells/ml). The cells were cooled down on ice for 10-20 min and then pelleted at 5000xg in a Beckman Avanti J-25 centrifuge with fixed angle rotor for 5 min at 4℃. The cell pellet was washed by resuspending in 25 ml ice cold 0.1 M $CaCl_2$. This was then centrifuged as before. The supernatant was removed and the pellet resuspended in 2.5 ml ice cold 0.1 M $CaCl_2$ and 15% (v/v) cold glycerol. The competent cells were incubated on ice for 1 hour followed by snap freezing of 200 µl of aliquots in liquid nitrogen (J. Hyde University of Manchester, UK, personal communication).

2.2.5.3 Transformation

Transformation of competent cells was performed as follows: The competent $CaCl_2$ or CCMB80 cells were thawed on ice, to which 50 ng of plasmid or 5 µl of ligation reaction was added followed by incubation on ice for 30 min. A 42℃ heat shock was applied for 90 sec in a water bath. Cells were quickly cooled down on ice for 2 min. 900 µl of warm (~ 37 -40℃) Luria-Bertani (LB)-glucose (2 mM) was added to the cells. The cells were then grown with shaking (~150-200 rpm) for 1 hour at 37℃. Cells we re plated out on LB (1% Tryptone, 0.5% Yeast Extract, 1% NaCl (w/v) pH 7) agar plates (2% w/v) containing 100 µg/ml ampicillin (Roche, Basel, Switzerland) for primary selection of recombinant clones in pET22b vector. 40 µg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactoside (Xgal) (Promega, Wisconsin, USA) and 0.2 mM of IPTG were added for blue-white screening to identify the recombinant clones. Blue-white screening involves the insertional inactivation of the *lacZ* gene and the production of a white colony as an indicator for this incorporation by the gene of interest. The *lacZ* gene encodes for β-galactosidase and is induced by IPTG. The expressed enzyme utilises X-gal as synthetic substrate to yield a blue product. Recombinant plasmids prevent the formation of the blue color due to the insertional inactivation of the *lacZ* gene.

2.2.6 Identification of recombinant clones

2.2.6.1 Plasmid isolation

2.2.6.1.1 STET-prep isolation

An adapted STET (Sucrose, TritonX-100, EDTA, Tris)-prep protocol was used as a quick and convenient method for plasmid isolation and screening of inserts. The method employs boiling to lyse bacteria and to denature all the DNA where-after the insoluble genomic DNA, cell walls and debris are removed by centrifugation. Plasmid DNA is then recovered by isopropanol precipitation (Holmes D.S. and Quigley M., 1981). Clones were picked from solid media plates and grown overnight in 10 ml LB liquid media containing 50 µg/ml ampicillin at 37°C with shaking (~150-200 rpm). 2 m l of this overnight culture was pelleted at 3500xg at 22°C for 10 min in a Microlitre Centrifuge (Hermle, Labortechnik Germany). The supernatant was removed and the pellet resuspended. To this, 300 µl of STET buffer (8% (w/v) sucrose, 5% (v/v) Triton X-100, 50 mM EDTA, 50 mM Tris, pH 8) was added. Lysozyme (Roche, Basel, Switzerland), was added to a final concentration of 0.1 mg/ml to digest the cell walls. The samples were placed in boiling water for 1 min and then centrifuged for 15 min at 16000xg at 22°C in a Microlitre Centrifuge (Hermle, Labortechnik Germany). The genomic and plasmid DNA is denatured by the high pH of the buffer, the plasmid DNA reanneals fully due to its topological constraints and stays in the aqueous phase. The genomic DNA forms insoluble globules with the salts and EDTA which are present in the pellet. Pellets were discarded and 300 µl isopropanol added to the supernatant for precipitation of the plasmid DNA at 22°C. The samples were centrifuged for 15 min at 16000xg at 22°C and the supernatant removed. The pellets containing plasmid DNA were air-dried and resuspended in 25 µl TE buffer (10 mM Tris, 1 mM EDTA, pH 8) containing 0.5 mg/ml RNase (Roche, Basel, Switzerland).

2.2.6.1.2 High Pure Plasmid isolation Kit

Plasmid DNA is purified on the same principle as described in section 2.2.4. The DNA is bound to silica in the presence of the chaotropic salt, guanidine-HCl. Chaotropic salt solutions and ethanol-containing buffers are used in several wash steps to remove contaminating bacterial components. The DNA is subsequently eluted using a low salt buffer. The High Pure Plasmid Isolation kit (Roche, Basel, Switzerland) was used to obtain pure plasmid samples following the manufacturer's protocol. Overnight cell culture (5-10 ml) was pelleted for 10 min at 5000xg (Heraeus Sepatech Medifuge). Cell pellets were resuspended in 250 µl suspension buffer (50 mM Tris-HCl, 10 mM EDTA, pH 8.0, 10% (w/v)

RNAse). To this, 250 µl of Lysis buffer (0.2 M NaOH, 1% SDS) was added and incubated for 5 min at room temperature. After lysis, 350 µl chilled (4°C) binding buffer (4 M guanidine-HCl, 0.5 M Potassium acetate, pH 4.2) was added. The sample was mixed by inversion and incubated on ice for 5 min. Cell walls and large contaminants such as genomic DNA and proteins were pelleted by centrifugation at 16000xg for 10 min at 22°C. The supernatant, containing plasmid DNA, was passed through a High Pure Filter column for binding of the plasmid DNA to the silica membrane and washed with wash buffer I (5 M guanidine-HCl, 20 mM Tris-HCl, pH 6.6) and II (20 mM NaCl, 2 mM Tris-HCl, pH 7.5), respectively. Plasmid was eluted in 100 µl elution buffer (10 mM Tris-HCl, pH 8.5).

2.2.6.2 Screening of recombinant clones

Restriction enzyme digestions were utilised for screening of clones. An *Eco*RI restriction enzyme site in the multiple cloning site of the pGEM® T-Easy vector (Appendix B) was used to screen for the presence of recombinant clones. This enzyme digests the plasmid (pGEM® T-Easy) at both sides of the multiple cloning site and additionally cuts the *P. falciparum dhfs-fpgs* insert internally at 948nt. Digestion reactions were each set up in a total volume of 10 µl and consisted of ~600 ng plasmid DNA, 10 U *Eco*RI (Promega, Madison, USA) and 1x Buffer H (pH 7.5, 90mM Tris-HCl, 10mM MgCl$_2$, and 50mM NaCl) (Promega, Madison, USA). The reactions were incubated at 37°C for a minimum of 3 hours and stopped by heat inactivation at 65°C for 10 min or by loading onto an agarose gel.

2.2.7 Subcloning into the pET22b vector system

For protein expression, the IPTG inducible *T7lacUV5* promoter of the pET22b vector system was employed (Novagen, EMD Biosciences Germany) (Appendix C). In this system, protein expression from the plasmid is tightly regulated to prevent the overproduction of protein, which might be toxic to cell growth. This system can produce C-terminally His$_6$-tagged fusion proteins. The presence of these tags allows for affinity purification of recombinant proteins in *E. coli*. Double digestions were performed on recombinant pGEM® T-Easy with *Nde*I (Promega, Madison, USA) and *Bam*HI (Promega, Madison, USA), which cuts on either side of the *dhfs-fpgs* insert. These enzymes were also used for the digestion of the pET22b vector for subcloning of the insert into this vector. Digestion reactions were each set up in a total volume of 10 µl and consisted of ~2.5 µg plasmid DNA, 10 U *Nde*I and *Bam*HI and 1x Buffer H (pH 7.9, 6 mM Tris-HCl, 6 mM MgCl$_2$, and 150 mM NaCl) (Promega, Madison, USA). The reactions were incubated at 37°C for a minimum of 3 hours and stopped by

loading onto an agarose gel. Digested plasmid and insert were separated on a 1.5% (w/v) agarose gel (Promega, Madison, USA) as in section 2.2.4. Bands were excised from the agarose gel and purified with the High Pure PCR product Purification Kit (Roche, Basel, Switzerland) as described in section 2.2.4. The purified gene fragment was used for subcloning into the pET22b vector. A 3:1 molar ratio of vector to insert was used in the ligation reaction containing 3 Weiss units of T4 DNA ligase and 1x T4 DNA ligase buffer (30 mM Tris-HCl pH 7.8, 10mM $MgCl_2$, 10 mM DTT, 1 mM ATP) (Promega, Madison, USA), 100 ng vector DNA and 80 ng digested insert to a final volume of 20 µl. This reaction was incubated overnight at 4°C. Half of the ligation mixture was cloned into heat-shock competent *E. coli* cells (see section 2.2.5.2).

2.2.8 Nucleotide Sequencing

Sequencing reactions were performed on recombinant plasmids isolated with the High Pure Plasmid Kit (section 2.2.6.1.2) using primers directed against the T7 promoter and T7 terminator sites on the pET22b vector (Promega, Wisconsin, USA) (Appendix C). Sequencing reactions were performed by utilising the incorporation of fluorescent dideoxynucleotides (ddNTPs) to terminate the template, as the ddNTPs act as chain terminators (Turner P.C. *et al.*, 2000). This sequencing is based on the method of Sanger *et al.*, which make use of the ddNTPs in addition to the normal nucleotides (Sanger F. and Coulson A.R., 1975). The incorporation of the ddNTPs leads to chain termination due the presence of the hydrogen group on the 3' carbon instead of a hydroxyl group. The use of the fluorescent ddNTPs allows the sequencing to occur in a single tube. The fluorescence of the four dyes are read at different wavelengths to determine a sequence chromatogram (Sanger F. and Coulson A.R., 1975; Turner P.C. *et al.*, 2000). Single-primer reactions were set up containing 200-500 ng pure plasmid DNA as template, 2 µl of 5x sequencing buffer (proprietary mix) (Perkin Elmer, Wellesley, USA), 2 µl Terminator Big-Dye Ready reaction mix (proprietary mix) version 3.1 (Perkin Elmer, Wellesley, USA) and double distilled deionised H_2O to a final volume of 10 µl. The sequencing cycling profile consisted of 25 cycles of denaturation at 96°C for 10 sec, annealing at 50°C for 30 sec and extension at 60°C for 4 min. The sequencing reactions were performed on a GeneAmp 9700 thermocycler (Perkin Elmer, Wellesley, USA).
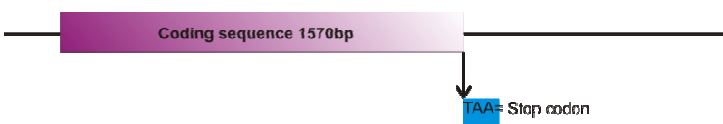
The products obtained from sequencing were cleaned up by means of DNA precipitation to remove all contaminants such as dNTPs, primer, and protein. 2.5 volumes of ice-cold 95% (v/v) ethanol and 2 µl of 3 M NaOAC, pH 4.6 were added to all sequencing reactions followed by centrifugation at 16000xg for 25-30 min at 4°C. The supernatant was removed

and the pellet washed with 12.5 volumes of 70% (v/v) ethanol followed by centrifugation as above. After centrifugation, the DNA pellet was dried *in vacuo* and stored at -20℃ in the dark to protect the light sensitive dyes. Sequencing was performed on an ABI PRISM 3100 capillary sequencer (Perkin Elmer, Wellesley, USA). The sequences were analysed with BioEdit version 5.0.9 software (Hall 2001, North Carolina State University, Department of Microbiology, USA). Sequences were additionally subjected to the NCBI Basic Local Alignment Search Tool (BLAST) to perform a Nucleotide Nucleotide BLAST to confirm the identity of the products (www.ncbi.nlm.gov).

2.2.9 Constructs, vectors and cell lines

Different constructs and cell lines were used for the comparative protein expression studies. The constructs obtained from both the previous study and this study is indicated in Table 2.3. The names used for these constructs are also indicated.

Table 2-3: Representation of the gene constructs indicating the annotations used. The changed stop codons are indicated in yellow and blue after the coding sequence (pink). The presence of the 6x Histag is indicated in green after the silenced stop codon.

Constructs	Vector	Insert origin	Annotation in Text		Synthesised
Tagless			**Gene Name**	**Protein name**	
Stop codon present	pET22b	Synthetic *Pf. dhfs-fpgs*	*pPfs-TagI*	*Pfs*-DF TagI	(Coetzee L., 2003)
No Histidine Tag		Native *Pf. dhfs-fpgs*	*pPfn-TagI*	*Pfn*-DF TagI	This study
Histag					
NO Stop codon present	pET22b	Synthetic *Pf. dhfs-fpgs*	*pPfs-HisT*	*Pfs*-DF HisT	(Coetzee L., 2003)
6x Histidine Tag		Native *Pf. dhfs-fpgs*	*pPfn-HisT*	*Pfn*-DF HisT	This study

Three different *E. coli* cell lines (BL21 (DE3) pLysS, BL21 Star (DE3) and BL21 Gold (DE3) pLysS) were used as in the previous study to determine which system gave optimal protein expression of the synthetic gene (Table 2.4) (Coetzee L., 2003). The same cell lines were also used in this study for the comparative protein expression studies of the native and synthetic genes. BL21 cell lines are *E. coli* B protein expression strains that lack both the *lon* protease and the *ompT* outer membrane protease, both of which can degrade proteins during purification. Thus, the lack of these two proteases reduces the degradation of heterologous proteins expressed in these strains. These cell lines also contain a λDE3 lysogen that carries the gene for T7 RNA polymerase under control of the *lacUV*5 promoter. BL21 cells containing a pLysS plasmid have increased expression control of toxic genes expressed by the system. The pLysS plasmid encodes a small amount of T7 lysozyme, which binds to T7 RNA polymerase and thereby reduces basal expression levels of the gene of interest. The presence of the pLysS also confers resistance to chloramphenicol (Cam^R). The BL21-Gold strain features the Hte phenotype, which increases the transformation efficiency. Additionally, the gene that encodes for endonuclease I (*endA*) is inactivated thereby ensuring no degradation of plasmid DNA during isolation allowing for direct use in cloning strategies. In summary, BL21 Star (DE3) cells were chosen for their high level of gene expression and easy induction, BL21 (DE3) pLysS cells for providing tighter control of toxic proteins and the BL21 Gold (DE3) pLysS cells for ease of transformation (Invitrogen Technical Data, La Jolla, USA).

Table 2-4: *E. coli* strains used as hosts for protein expression

Expression host	Genotype	Antibiotic resistance	Annotation used in text
BL21 (DE3) pLysS (Stratagene, La Jolla, USA)	B strain F^- *dcm ompT hsdS$_B$* ($r_B^-m_B^-$) *gal* λ(DE3) [pLysS Cam^r]	Chloramphenicol	BL21
BL21 Star (DE3) (Invitrogen, La Jolla, USA)	B strain F^- *ompT hsdS* ($r_B^-m_B^-$) *gal dcm rne131λ* (DE3)	None	Star
BL21 Gold (DE3) pLysS (Invitrogen La Jolla, USA)	B strain F^- *dcm ompT hsdS* ($r_B^-m_B^-$) dcm$^+$ Tet^r *gal* λ(DE3) *endA* Hte [pLysS Cam^r]	Chloramphenicol	Gold

2.2.10 Protein expression

$CaCl_2$ competent expression hosts (section 2.2.5.2) were freshly transformed with plasmid containing the different gene constructs (section 2.2.9) and plated onto LB-agar plates supplemented with ampicillin (100 µg/ml) for recombinant plasmid selection, as well as chloramphenicol (100 µg/ml) for selection of the pLysS plasmid in the BL21 and Gold cells. Positive clones were inoculated into 10 ml LB media containing ampicillin (50 µg/ml) and grown overnight at 37°C with shaking (~150-200 rpm). The overnight cultures were subsequently diluted 20-fold in LB-Amp and grown at 30°C with shaking (~150-200 rpm) until an OD_{600nm} of 0.4 was reached. Uninduced samples (10 ml each) were taken and a final concentration of 0.1 mM IPTG added to the remaining culture for induction of protein expression by the pET vector *T7lacUV5* promoter. Expression was carried out at 30°C for 3.5 hours, after which cells were pelleted by centrifugation at 5000xg for 15 min at 22°C. The pellets were dissolved in Bugbuster Protein Extraction Reagent (Novagen, EMD Biosciences, Germany), which is a proprietary zwitterionic detergent mixture for gentle lysis of cells. A volume of 5 ml "Bugbuster" per gram wet cell paste was used and 1/1000 DNaseI (10mg/ml) (Roche, Basel Switzerland) was added to prevent excess sample viscosity. For protein extraction, the samples were left in Bugbuster/DNaseI at 4°C overnight. Protease inhibitors were not added, as the BL21 cell lines are deficient in the *lon* protease and the *ompT* outer membrane protease. The induced samples were split into two fractions; one was used for total protein analysis and the other was centrifuged at 16000xg for 20 min. The resulting pellet was analysed as the insoluble fraction and the supernatant as the soluble fraction. Total fractions refer to the soluble and insoluble proteins together before centrifugation.

2.2.11 Protein concentration determination

The Bradford method was used for protein concentration determination (Bradford M.M., 1976). The determination involves the binding of Coomassie Brilliant Blue G-250 to protein, shifting the absorbance of 465nm to 595nm, which is monitored spectrophotometrically. The dye binds primarily the aromatic and basic amino acids on the surface of the protein. Bovine serum albumin (BSA) (Promega, Madison, USA) was diluted to obtain a standard concentration range of 200, 100, 50, 25, 12.5, and 6.25 µg/ml protein. To 100 µl of each BSA standard or protein sample, 100 µl Quick Start Bradford Dye (Bio-Rad, Hercules, USA) was added and the absorbency measured at 595nm in 96 well microtitre plates (Bibby Sterilin Ltd., Stone Staffordshire, UK) on a Multiskan Ascent spectrophotometer (Thermo Labsystems).

2.2.12 SDS-PAGE analysis

Polyacrylamide gel electrophoresis (PAGE) was performed with a 4% acrylamide-methylene-(bis) acrylamide stacking gel containing 0.1% (v/v) SDS, 0.125 M Tris-HCl (pH 6.8), 0.5% (v/v) N,N,N'N'-tetramethylenediamine (TEMED), 0.05% (w/v) ammonium persulfate, and a 12.5% acrylamide-methylene-(bis)acrylamide, separating gel containing 0.1% (v/v) SDS, 0.375 M Tris-HCl (pH 8.8), 0.5% TEMED and 0.5% (w/v) ammonium persulfate. Electrophoresis was performed in a 0.1% (v/v) SDS, 0.25 M Tris, 192 mM Glycine buffer (pH 8.3). Samples were diluted 1:1 in reducing SDS sample buffer (0.0625 M Tris-HCl pH 6.8, 2% (w/v) SDS, 10% (v/v) glycerol, 0.05% (v/v) β-mercaptoethanol and 0.0025% (w/v) bromophenol blue) and denatured by boiling for 5 min. 10 x 8 cm x 1mm gels were run at 60 Volts and 10 mA for the stacking gel and at 110 V and 20 mA for separation of protein bands in the Minigel G-41 system (Biometra, Germany). For larger 16 x 18cm x 1mm gels the running conditions were: a beginning voltage, current and power of 111 V, 50 mA and 6 W respectively for the stacking gel. After 3 hours a voltage, current and power of 280 V, 35.9 mA and 10 W, respectively, were applied for the separating gel (Hoefer SE 600 Ruby system).

Protein bands were visualised with Coomassie staining (0.1% (w/v) Coomassie brilliant blue G250 in 40% (v/v) methanol, 10% (v/v) acetic acid) and destained with 40% methanol, 10% acetic acid. Alternatively, Colloidal Coomassie was used. The gels were stained for 24 hours in 4 parts Coomassie Blue (10% (w/v) ammonium sulphate in 2% phosphoric acid with 0.1% of Coomassie Brilliant Blue G) and 1 part Methanol. Gels were quickly fixed with 25% methanol and 10% acetic acid and subsequently destained with 25% methanol to reduce background (Neuhoff V. *et al.*, 1988). The diffusion of molecularly dispersed dye molecules is crucial during colloidal staining. The high concentration of ammonium sulphate in the staining solution ensures that the dye shifts into its colloidal form thus minimising the background. This is most favourable for quantitative analysis. The inclusion of an acidic alcoholic media (20% methanol) shifts the equilibrium from the colloidal form to dispersed dye, which is needed to facilitate diffusion into the gel matrix. This diffusion results in a more rapid and complete staining of the gels' cross-section (Neuhoff V. *et al.*, 1988).

Gels were scanned electronically on the Versadoc 4000 image scanner (Biorad, Hercules, USA) and analysed using the Quantity One™ Software Package from Biorad (Hercules,USA), Microsoft™ Excel™ format. The following parameters were used in the analysis: Peak Density, Average Density, Trace Quantity, and Relative Quantity.

The protocols described next were used to increase the protein expression levels for the synthetic *Pf. dhfs-fpgs* gene. The synthetic gene was previously constructed as explained in section 2.1.6

2.2.13 Unfolding/refolding protocol

After protein extraction and separation of the soluble and insoluble fractions by centrifugation (refer to section 2.2.10), the inclusion bodies from the Star cells expressing the *Pfs*-DF HisT protein were further purified and washed from the insoluble fraction using the manufacturer's protocol for Bugbuster Protein Extraction Reagent (Novagen, EMD Biosciences, Germany). Cell pellets resulting from liquid culture were resuspended in the equivalent (1:1) amount of "Bugbuster" (section 2.2.10) and mixed well by vortexing. Lysozyme was added (200 µg/ml) and the mixture vortexed for 5 min at 22°C. To this, 6 volumes of a 10-fold dilution of "Bugbuster" was added, the mixture vortexed for 1 min and centrifuged at 16000xg for 15 min at 4°C. The supernatant was removed and the pellet resuspended in half the original sample volume of 10-fold diluted "Bugbuster". The sample was mixed well and centrifuged at 16000xg for 15 min. This process was repeated twice. The purified inclusion bodies were solubilised according to the method of Sirawaraporn *et al* (Sirawaraporn W. *et al.*, 1993). The purified inclusion body pellet was resuspended in 10 ml of Buffer A (20 mM potassium phosphate buffer pH 7, 0.1 mM EDTA, 10 mM DTT), supplemented with 0.2 M KCl and 6 M guanidine HCl for a 300 ml culture, and incubated at 4°C for 1 hour to solubilise the protein. The protein mixture was subsequently added in a drop-wise manner to achieve a 20-fold dilution in Buffer A (supplemented with 0.2 M KCl and 20% (v/v) glycerol) at 4°C for the removal of excess denaturant and reducing agent to ensure satisfactory refolding of the protein. This mixture was incubated overnight at 4°C and then centrifuged at 16000xg for 20 min to pellet all the remaining insoluble proteins. The refolded-soluble fraction was then used for affinity purification (section 2.2.17), and analysed with Liquid Chromatography-Tandem Mass Spectrometry LC/MS/MS (section 2.2.18).

2.2.14 Auto-induction protocols

The following media (Table 2.5) and protocols for auto-induction of cultures were adapted from (Grossman T.H. *et al.*, 1998; Studier F.W., 2005).

Table 2-5: Media used for auto-induction of cultures

Composition of inducing media (Final conc.)	Culturing conditions	Annotation in text	Reference
25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM Na_2SO_4, 2 mM $MgSO_4$, 0.5% Glucose	Overnight at 37℃	Non-inducing media	(Studier F.W., 2005)
1% Tryptone, 0.5% Yeast extract, 25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM Na_2SO_4, 2 mM $MgSO_4$, 0.5% Glycerol, 0.05% Glucose and 0.2% Lactose	Overnight at 37℃	Auto-inducing media	(Studier F.W., 2005)
3.2% (w/v) Bactro-tryptone, 3% (w/v) Bacto-yeast extract, 0.5% (w/v) NaCl	37℃ for 24hours / 18℃ for 48h	4x YT media	(Grossman T.H. *et al.*, 1998)
Added Yeast extract to a final concentration of 6% to LB-media grown to saturation	37℃ for 24hours / 18℃ for 48h	6% YE	(Grossman T.H. *et al.*, 1998)

pPfs-HisT in Star cells was grown overnight at 37℃ and diluted 20-fold into 4xYT media. The 4xYT cultures were grown at 37℃ for 24 hours or at 18℃ for 48 hours. Alternatively, *pPfs-HisT* in Star cells was inoculated into LB-media and grown to an OD_{600nm}=2.0 after which yeast extract (YE) was added to a final concentration of 6% (v/v). The 6% YE media cultures were subsequently grown under the same conditions as the 4xYT media cultures. For both media, growth curves were determined by measuring the OD_{600nm} at specific time intervals (Grossman T.H. *et al.*, 1998).

The second method of auto-induction makes use of cultures grown in non-inducing media to retain the plasmid of interest. These non-inducing cultures are then inoculated into complex auto-inducing media (no inducer) and grown to saturation. *pPfs-HisT* in Star cells was inoculated into non-inducing media and grown overnight at 37℃. The overnight non-inducing cultures were aliquoted as 1 ml stocks in 10% (v/v) glycerol concentration and stored at -70℃. For the auto-induction, cultures grown in non-inducing media were diluted 1000-fold into auto-inducing media and incubated overnight with shaking at 37℃. (Studier F.W., 2005). SDS-PAGE analysis (section 2.2.12), purification (section 2.2.17), and activity assays (section 3.1.3) of the protein were performed for the culturing methods.

2.2.15 Chaperone proteins

The chaperone containing plasmid pG-KJE8 was chosen for co-expression with the target protein (Appendix D). This plasmid contains the *E. coli* chaperones dnaK-dnaJ-grpE-groES-groEL. The dnaK-dnaJ-grpE is part of the Hsp70 family, which associates immediately after protein synthesis with partially folded proteins thereby stabilising the extended chains. It additionally preserves folding capability. The groES-groEL is part of the Hsp60 family, which associate with already aggregated proteins to refold them (Treweek T.M. *et al.*, 2003). The chaperone genes are situated downstream of either the araB (L-Arabinose) or Pzt-1 (Tetracycline) promoters and can thus be induced individually with the inducers indicated in brackets.

Star cells were transformed with the chaperone plasmid pG-KJE8 (Takara Bio Inc, Japan) (Appendix D). Transformation reactions were plated out onto LB-agar plates containing 20 µg/ml of chloramphenicol for selection of the plasmid (Boehringer Manheim, Germany). Positive clones were subsequently used for preparing competent cells and transformed with *pPfs-TagI* or *pPfs-HisT* (section 2.2.5.3). Cells were plated out onto a double selection LB-Agar plate containing 20 µg/ml chloramphenicol and 50 µg/ml ampicillin. Positive clones were grown overnight at 37°C in 5 ml non-inducing media (section 2.2.14). The overnight cultures were diluted 1000-fold into auto-inducing media containing 50 µg/ml ampicillin, 20 µg/ml chloramphenicol. 10 ng/ml tetracycline and 4 mg/ml L-Arabinose were subsequently added as inducers of the chaperone protein expression. The auto-inducing media cultures were grown overnight at 37°C with shaking (~150-200 rpm). Expressed proteins were isolated (section 2.2.10) and characterised (section 2.2.12).

2.2.16 Detergent extraction

The insoluble pellets as obtained in section 2.2.14 were resuspended with the same amount of Tris-EDTA-KCl buffer (10 mM Tris, 1 mM EDTA, 0.2 M KCl, pH 8) as initial Bugbuster volume. n-Octyl-β-glucopyranoside, a non-ionic detergent, (Fluka, Sigma-Aldrich USA) was added to a final concentration of 1% (v/v). Cell lysates were rotated on a Rotater HAG (Fine PCR, Seoul, Korea) at 4°C for 30 min and centrifuged at 16 000xg for 30 min. The pellet was resuspended in 1/5 of the initial Bugbuster volume of TE-KCl buffer and n-Lauroylsarcosine-sodium salt, an ionic detergent, (Sigma-Aldrich, St.Louis, USA) was added to a final concentration of 0.3-1% (v/v). The lysates were rotated and centrifuged as above. The solublised fraction (supernatant) was used for protein purification. The sarcosine

concentration was adjusted to 0.3% and pH 8 with equilibration buffer (section 2.2.17) (Angov E., Walter Reed Army Institute of Research, USA, personal communication).

2.2.17 Protein purification

2.2.17.1 Affinity chromatography

Purification of histidine-tagged protein was performed with immobilised metal ion affinity chromatography (IMAC) using HIS-select HC Nickel affinity gel (Sigma-Aldrich, St.Louis, USA). This method of affinity purification relies on the electrostatic interaction between the negatively charged hexahistidine peptide tag of the target protein (pH 8) and the positively charged Ni^{2+} immobilised on the column resin. Unbound proteins are washed off and the target protein with the hexahistidine tag eluted by the addition of a high imidazole concentration. The imidazole binds competitively to the Ni^{2+} column resin. A large-scale purification protocol was used for purification of His-tag fusion proteins. The Ni^{2+} affinity gel suspension with a binding capacity of 40-45 mg/ml, was added to a chromatography column for a bed volume of 1 cm^3. The gel beads were washed with 2 column bed volumes of dddH$_2$O and 3 column bed volumes of equilibration buffer (50 mM Na$_2$HPO$_4$/NaH$_2$PO$_4$, pH 8, 0.3 M NaCl). The supernatants (adjusted to pH 8 with equilibration buffer) from sections 2.2.13 and 2.2.16, were added to the affinity gel. The affinity column was washed twice with 10 column bed volumes of equilibration buffer. The target protein was eluted with 3-10 column bed volumes of elution buffer (50 mM sodium phosphate, pH 8, 0.3 M NaCl, 250 mM imidazole). Fractions (1 ml each) were collected for every wash and elution step and analysed with SDS-PAGE (section 2.2.12).

2.2.18 Mass spectrometry analysis

Mass spectrometry was carried out by Dr. P. Sims, Manchester Interdisciplinary Biocentre, University of Manchester, UK. Bands were excised from gels and digested with trypsin before analysis by LC-MS/MS (Nirmalan N *et al.*, 2004). Peptide/protein identities were established using a locally mounted version of Mascot (Matrix Science, (Perkins D.N. *et al.*, 1999)).

2.3 Results

A comparative study of the expression of *P. falciparum* DHFS-FPGS in *E. coli* from both the previously constructed synthetic *dhfs-fpgs* gene and the isolated native *dhfs-fpgs* gene was performed. This was done to determine the efficiency of expression of the codon-optimised gene as well as to obtain a sufficient amount of protein for activity determination.

2.3.1 Primers

The primers designed for the isolation of the *pPfn-TagI* and *pPfn-HisT* genes from the *Pf*1 plasmid (Prof J Hyde, (Salcedo E. *et al.*, 2001) are indicated in Table 2.6. The forward primer contains the *Nde*I restriction site "CAT<u>ATG</u>" and the reverse primers the *Bam*HI site "GGATCCT" for directional cloning into the pET22b vector.

Table 2-6: Forward (fwd) and reverse primers (rev) used for the amplification of native *dhfs-fpgs* gene.

Primer name	Sequence (5' to 3')	Length (nt)	Tm (℃) *	Construct
pfdffwd	CGCGGA**CAT**↓**ATG**GAAAAAAATC	22	59	*pPfs-TagI / pPfs-HisT*
pfdfrev▪	CTC**G**↓**GATCCT**G<mark>TTT</mark>AACAAGAGATG	25	62	*pPfs-HisT*
pfdfrev2●	CTC**G**↓**GATCCT**GT<mark>TTA</mark>AACAAGAGATC	26	63	*pPfs-TagI*

*Tms were calculated using the formula: Tm=69.3 +0.41(%GC) – 650/primer length (Rychlik W., 1993)

▪The rev primer was used to silence the stop codon (indicated in yellow) at the 3' end of the gene to ensure read-through and His-tag expression.

●The rev2 primer containing the stop codon (indicated in purple) was used to obtain a Tagless expression product.

2.3.2 Optimisation of the native *Pf dhfs-fpgs* gene amplification

Amplification of the *pPfn-TagI* and *pPfn-HisT* gene constructs indicated that 30 cycles at 58℃ provided the most specific PCR product (Figure 2.2). Both reverse primers pfdfrev2 and pfdfrev, respectively were used for the generation of the native constructs with and without the stop codons, *pPfn-TagI/pPfn-HisT*, (Figure 2.2 lanes 1 and 2). Correct sizes of ~1573bp were obtained and PCR products were subsequently purified as described in section 2.2.4. From Figure 2.2 it is evident that the amplification of *pPfn-HisT* results in less product formation possibly due to inefficient amplification as reflected by the presence of the primer dimers (bottom of lane 2).
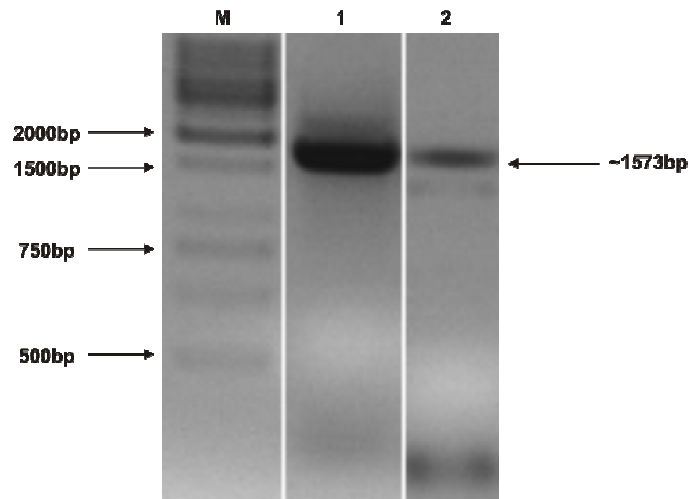
Figure 2-2: The full-length *pPfn-TagI/pPfn-HisT* after large scale PCR amplification. M is the molecular marker (Generuler 1kb DNA ladder, Fermentas); Lane 1: Full length *pPfn-TagI*, Lane 2: Full length *pPfn-HisT*. The arrow on the right indicates the expected fragment size. PCR products were run on a 1.5% agarose gel.

The concentrations of the purified products obtained were 26 ng/µl and 31 ng/µl for the *pPfn-TagI* and *pPfn-HisT* products, respectively.

2.3.3 Screening of recombinant clones

PCR products obtained were ligated into pGEM® T-Easy vector and transformed into DH5α cells. Blue-white colony screening was used to determine the presence of positive clones. Potential recombinant plasmids were isolated using the STET-prep method and subsequently used in restriction digestions to confirm presence of insert (section 2.2.6.). The plasmid vector map is indicated in Figure 2.3, illustrating the restriction sites of the enzymes used (*Eco*RI, *Bam*HI*,* and *Nde*I).
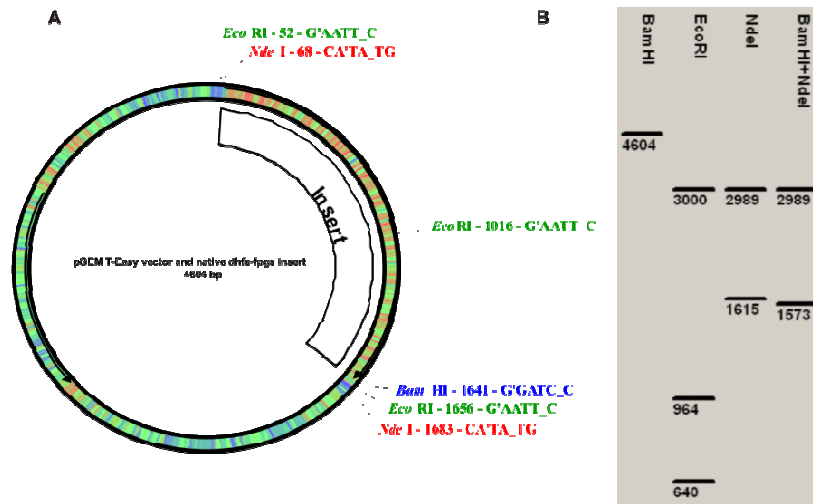
Figure 2-3: The *Eco*RI, *Bam*HI and *Nde*I restriction map of the pGEM® T-Easy vector containing the native *Pf dhfs-fpgs* insert. A) The *Eco*RI sites are indicated in green and the *Bam*HI and *Nde*I indicated in blue and red, respectively. The nucleotide sequence positions are also indicated. **B)** The expected band sizes in bp after digestions as would be visualised after electrophoresis on a 1.5% agarose gel detection.

Results obtained from the restriction enzyme digestions showed that, of the 4 colonies screened for the *pPfn-TagI* construct and 10 screened for the *pPfn-HisT* construct, 3 recombinant clones each were obtained. Figure 2.4 indicates an example of the results obtained for the digestion of one of the positive *pPfn-HisT* plasmids with both *Nde*I*/Bam*HI and *Eco*RI on a 1.5% agarose gel. Similar results were obtained for the *pPfn-TagI* plasmid (Results not shown).
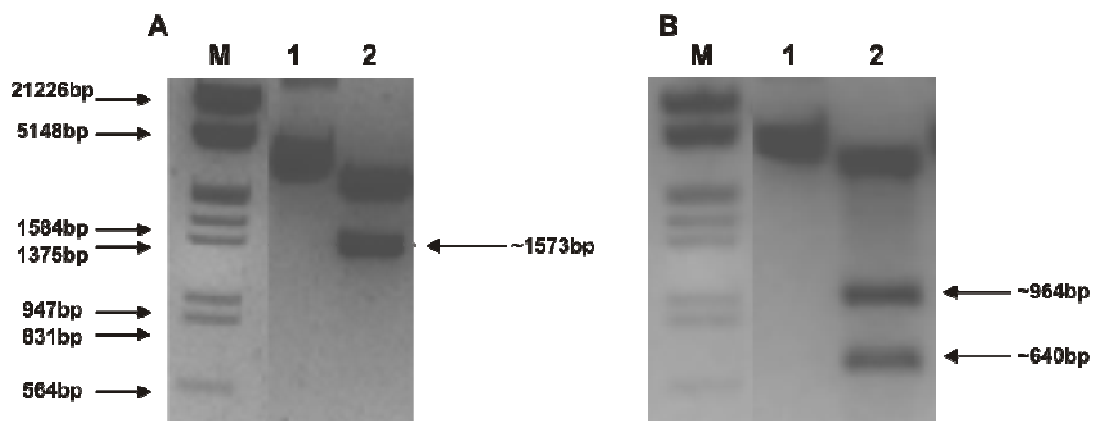
Figure 2-4: *Nde*I*/Bam*HI (A) and *Eco*RI (B) digestions of recombinant *pPfn-HisT* plasmid. M: Molecular mass marker (*Eco*RI-*Hind*III digested λ-phage DNA). Lane 1 and 2 undigested and digested *pPfn-HisT* respectively. Expected band sizes are indicated by the arrows on the right. Products were electrophoresed on a 1.5% agarose gel.

These positive recombinant plasmids were purified and subsequently digested on a large scale for use in sub-cloning of the insert into the pET22b vector. The vector map of pET22b containing the inserts is given in Figure 2.5.
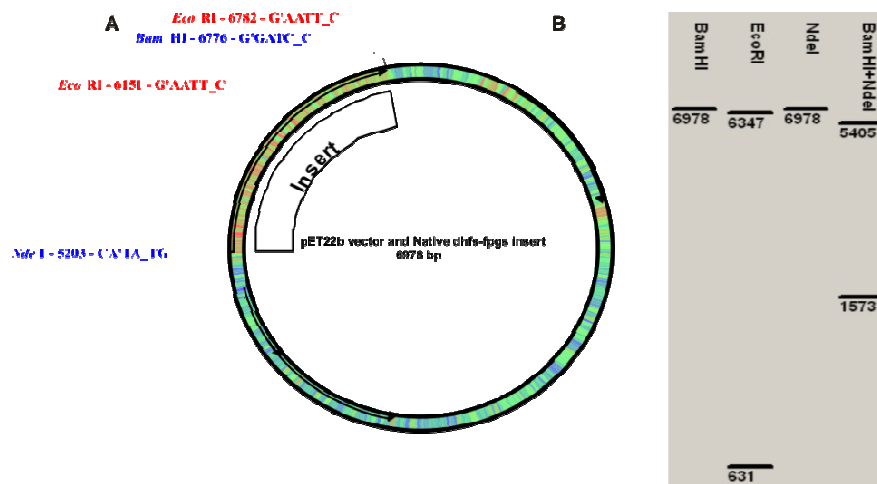
Figure 2-5: The *Eco*RI, *Bam*HI and *Nde*I restriction map of the pET22b vector containing the native *Pf dhfs-fpgs* insert. A.) The *Eco*RI sites are indicated in red and the *Bam*HI and *Nde*I indicated in blue respectively. The nucleotide sequence positions are also indicated. **B)** The expected band sizes in bp after digestion as would be visualised after electrophoresis on a 1.5% agarose gel detection.

Clones obtained after sub-cloning of *pPfn-TagI* and *pPfn-Hist* into pET22b were isolated and screened using *Nde*I/*Bam*HI and *Eco*RI digestion (refer to Figure 2.6). Figure 2.6 indicates an example of the results obtained for the digestion of one of the positive *pPfn-Hist* plasmids with both *Nde*I/*Bam*HI and *Eco*RI on a 1.5% agarose gel. Similar results were obtained for the *pPfn-TagI* plasmid (Results not shown).
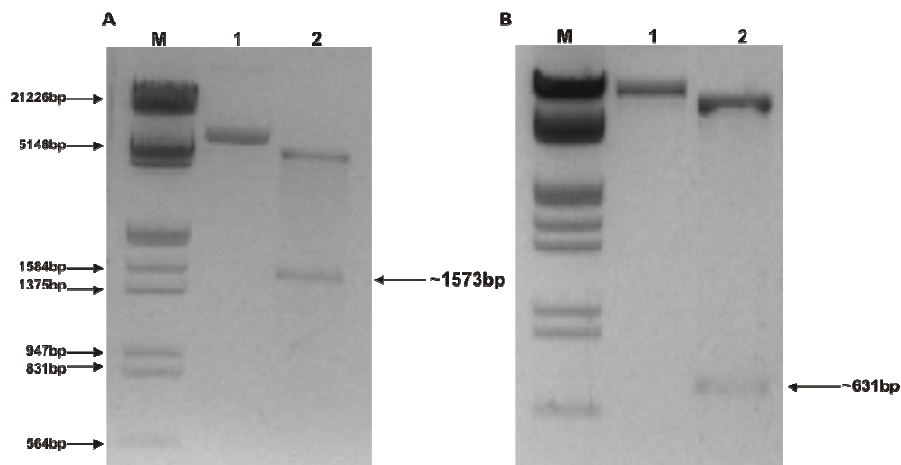
Figure 2-6: *Nde*I/*Bam*HI (A) and *Eco*RI (B) digestion of recombinant pET22b plasmid. M: Molecular mass marker (*Eco*RI-*Hind*III digested λ-phage DNA. Lane 1 and 2 Undigested and digested *pPfn-HisT*, respectively. Expected band sizes are indicated by the arrows on the right. Products were electrophoresed on a 1.5% agarose gel.

2.3.4 Comparative protein expression of DHFS-FPGS constructs

Both the native DHFS-FPGS construct (*pPfn-HisT*/*pPfn-TagI*) as well as the construct derived from the synthetic, codon-optimised gene (*pPfs-HisT*/*pPfs-TagI*) cloned into pET22b were used in subsequent expression studies (Refer to Table 2.3). All of the constructs were transformed into the three different *E. coli* cell lines; BL21, Star, and Gold (refer to section 2.2.9). Total protein fractions (soluble + insoluble proteins) were isolated for each construct, in each cell line. The total protein fractions were subsequently further fractionated into soluble and insoluble fractions. 750 ng of each protein fraction was electrophoresed on 12% SDS-PAGE gels. Gels were stained with Coomassie Brilliant blue and analysed using the Quantity One Software package. Results of *Pfs*-DF HisT and *Pfn*-DF HisT are compared in Figures 2.7, 2.8 and 2.9, and analyses of *Pfs*-DF TagI and *Pfn*-DF TagI are indicated in Figures 2.10, 2.11 and 2.12.

Figure 2-7: Comparison of *Pfn*-DF HisT and *Pfs*-DF HisT expression in (A) BL21 (B) Gold and (C) Star cells. Lane 1: M: Protein Molecular Weight marker, Fermentas. The arrows indicate the position of the expected 60kDa band.

From the SDS-PAGE results (Figure 2.7 C) it is clear that the ~60kDa *Pfs*-DF HisT protein expresses optimally in Star cells even though ~50% of the proteins are insoluble (Figure 2.9). The expression levels in the other two cell lines appear to be the same between *Pfn*-DF HisT and *Pfs*-DF HisT. To determine the fold difference in total protein expression of each construct (Figure 2.8), as well as the expression levels of *Pfn*-DF HisT and *Pfs*-DF HisT in the soluble and insoluble fractions (Figure 2.9), the gels were analysed using the relative quantity measure obtained from the Quantity One program. The relative quantity measured

is the quantity of a particular band as measured by its intensity expressed as a percentage of the total intensity of all the bands in the lane. The total induced fraction was normalised against the total uninduced fraction. The uninduced fraction was set at a relative quantity of 1 (Figure 2.8). The fold differences were calculated between the total induced fractions of the *Pfs*-DF HisT compared to the *Pfn*-DF HisT, or visa versa.

	Star	BL21	Gold
■ Pfn-DF HisT	2.00	1.70	1.56
■ Pfs-DF HisT	4.25	0.54	0.93
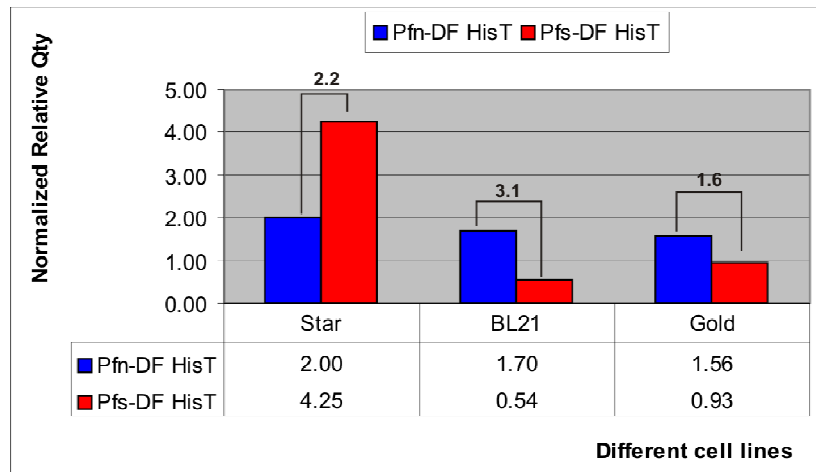
Figure 2-8: Relative quantity band analysis of total *Pfn*-DF HisT and *Pfs*-DF HisT protein expression in three different cell lines. (Uninduced = 1). Fold differences are indicated above the bars.

From the relative quantity band analysis on the total protein fraction (Figure 2.8), the *Pfs*-DF HisT showed a 2.2-fold higher total protein expression than the *Pfn*-DF HisT in Star cells. The *Pfn*-DF HisT showed a 3.1-fold and 1.6-fold higher total protein expression level than the *Pfs*-DF HisT in BL21 and Gold cells, respectively. Even though the *Pfn*-DF HisT construct demonstrates higher total protein expression levels in the other two cell lines, it is still markedly lower (2.5-fold) than the total protein expression of *Pfs*-DF HisT in the Star cells.

Relative quantity band analysis was also performed on the soluble and insoluble fractions obtained to determine the expression levels of *Pfn*-DF HisT and *Pfs*-DF HisT in these fractions. The relative quantity obtained for the soluble and insoluble fractions were not normalised against uninduced sample as no uninduced soluble/insoluble fractions were analysed on the SDS-PAGE gels. The relative quantities obtained for *Pfn*-DF HisT and *Pfs*-DF HisT in the insoluble and soluble fractions are indicated in Figure 2.9.

Figure 2-9: Relative quantity band analysis of the soluble and insoluble fractions of _Pfn_-DF HisT and _Pfs_-DF HisT protein expression in three different cell lines. Fold differences are indicated above the bars

From the relative quantity band analysis of the soluble and insoluble fractions (Figure 2.9), it is observed that both the _Pfn_-DF HisT and the _Pfs_-DF HisT have similar distribution of ~50% soluble and ~50% insoluble expression in all three cell lines. The only exception is _Pfs_-DF HisT in the Gold cells, where a 2.2-fold difference between soluble and insoluble was observed. The _Pfn_-DF HisT indicated the highest soluble expression in the BL21 cells while the _Pfs_-DF HisT showed the highest soluble expression in the Gold cells.

A comparable profile of expression levels for the _Pfn_-DF TagI and _Pfs_-DF TagI protein expression is indicated in Figure 2.10. 750 ng of each protein fraction (total, soluble and insoluble) was electrophoresed on 12% SDS-PAGE gels. Gels were stained with Coomassie Brilliant blue and analysed using the Quantity One Software package.
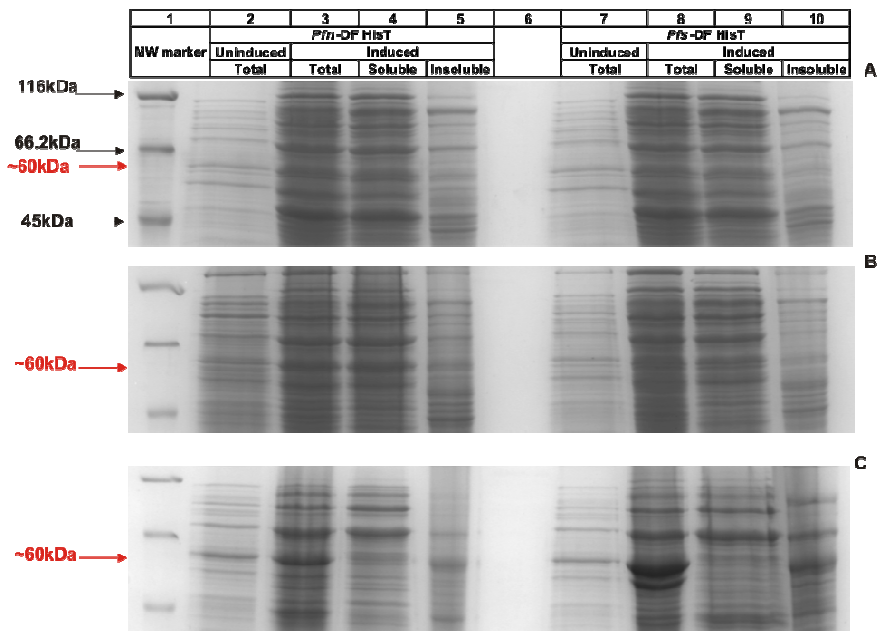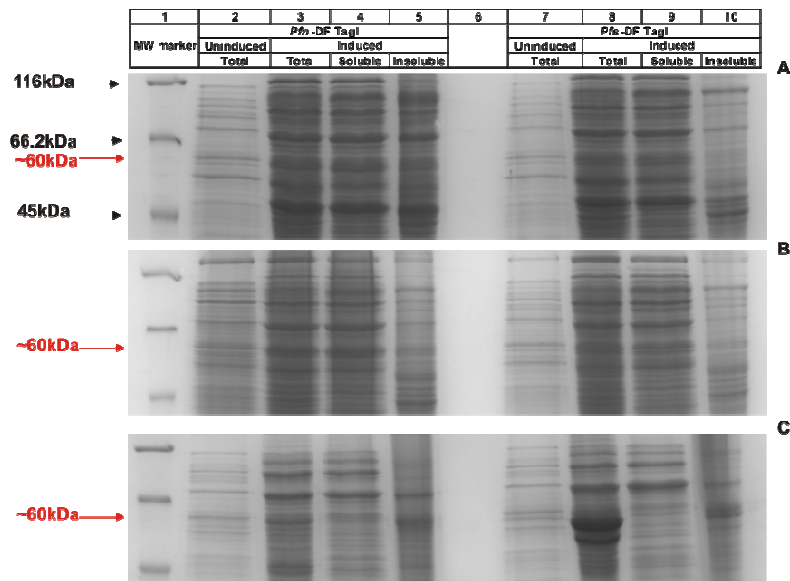
Figure 2-10: Comparison of *Pfn*-DF TagI and *Pfs*-DF TagI) expression in (A) BL21 (B) Gold and (C) Star cells. Lane 1: M: Protein Molecular Weight marker, Fermentas. The arrows indicate the position of the expected 60kDa band.

From the SDS-PAGE results (Figure 2.10 C), the ~60kDa *Pfs*-DF TagI protein express the best in the BL21 Star cells but mostly as insoluble protein (Figure 2.12), while the *Pfn*-DF TagI expresses at the same levels as the *Pfs*-DF TagI in the other two cell lines. The gel was analysed as before using the relative quantity measure as obtained from the Quantity One program. The relative quantity was used to determine the fold difference in total protein expression (Figure 2.11) as well as expression levels of *Pfn*-DF TagI and *Pfs*-DF TagI in the soluble and insoluble fractions (Figure 2.12). The fold differences were calculated between the total induced fractions of the *Pfs*-DF TagI compared to the *Pfn*-DF TagI, or visa versa. The total induced fraction was normalised as mentioned before (Figure 2.11).

	Star	BL21	Gold
■ Pfn-DF TagI	2.07	0.56	2.35
■ Pfs-DF TagI	7.11	0.92	0.48

Different cell lines
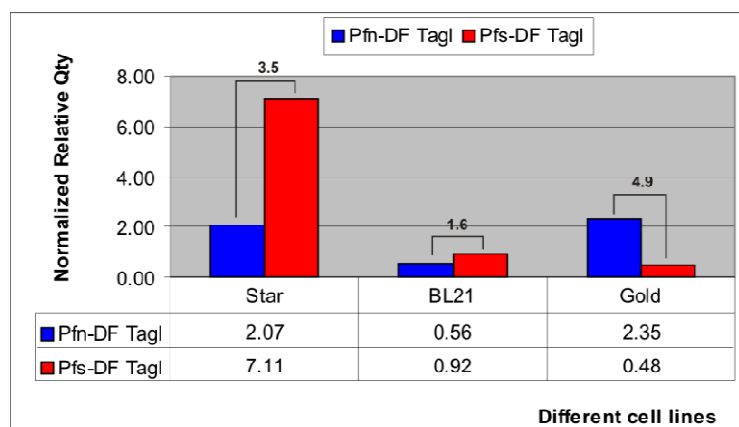
Figure 2-11: Relative quantity band analysis of total *Pfn*-DF TagI and *Pfs*-DF TagI protein expression in three different cell lines. (Uninduced =1). Fold differences are indicated above the bars.

The relative quantity band analysis (Figure 2.11) of *Pfs*-DF TagI showed a 3.5-fold higher total protein expression than *Pfn*-DF TagI in Star cells, as well as a 1.6-fold higher total protein expression in the BL21 cells when compared with *Pfn*-DF TagI. In contrast, *Pfn*-DF TagI showed a 4.9-fold higher total protein expression level in the Gold cells. Even though the *Pfn*-DF TagI construct exhibits the highest fold expression in the Gold cell line, it is still markedly lower (3-fold) than the total protein expression of the *Pfs*-TagI protein in Star cells.

The relative quantities obtained for *Pfn*-DF TagI and *Pfs*-DF TagI in the insoluble and soluble fractions are indicated in Figure 2.12. Relative quantity band analysis was performed as mentioned earlier in this section.
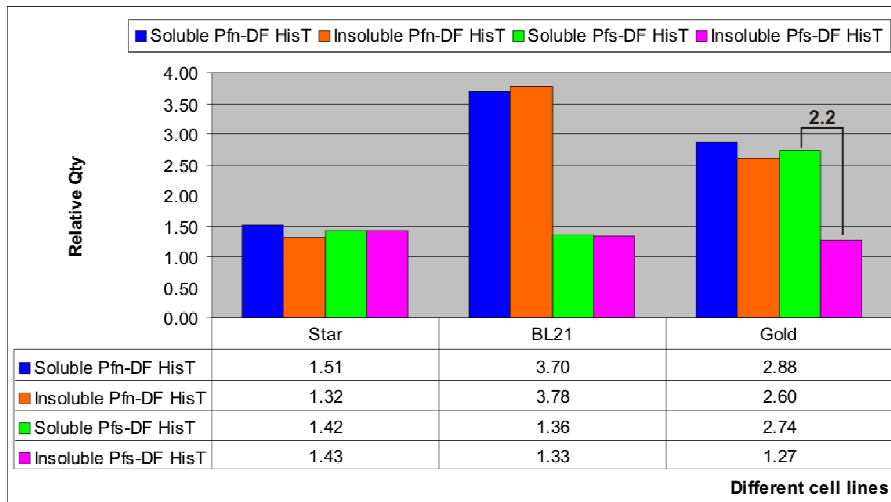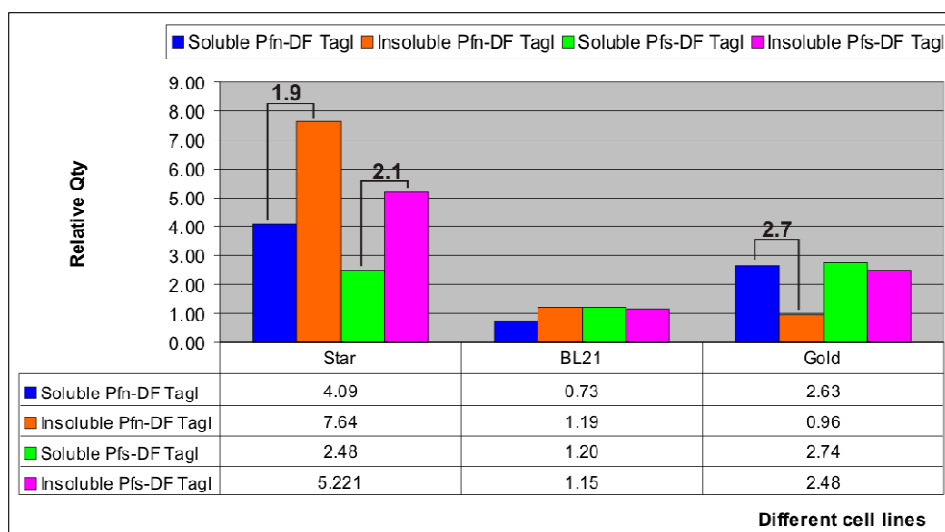

Figure 2-12: Relative quantity band analysis of the soluble and insoluble fractions of *Pfn*-DF TagI and *Pfs*-DF TagI protein expression in three different cell lines. Fold differences are indicated above the bars.

From the relative quantity band analysis of the soluble and insoluble fractions (Figure 2.12), it is indicated that the *Pfn*-DF TagI and the *Pfs*-DF TagI have the highest insoluble protein expression in the Star cells, where a 1.9 and 2.1-fold differences between soluble and insoluble were observed for the two constructs, respectively. Comparison of the soluble protein levels indicated that *Pfn*-DF TagI also has the highest soluble protein expression in the Star cells compared to the other two cell lines as well as to *Pfs*-DF TagI. The *Pfn*-DF TagI and the *Pfs*-DF TagI have similar soluble and insoluble protein levels of ~50% in the BL21 cells. Both the *Pfs*-DF TagI and the *Pfn*-DF TagI showed higher soluble expression than insoluble expression in the Gold cells. The *Pfn*-DF TagI expression indicated a 2.7 fold difference between the soluble and insoluble fraction.

Results taken as a whole are indicated in Table 2.7. The table indicates the relative quantities obtained after band analysis (off all fractions) for both the native constructs (*pPfn-HisT/pPfn-TagI)* as well as the construct derived from the synthetic, codon-optimised gene (*pPfs-HisT/pPfs-TagI*) expressed in the three different cell lines. Thus overall, the *Pfs*-DF HisT and *Pfs*-DF TagI total protein expression levels are ~3-fold higher than the *Pfn*-DF HisT and *Pfn*-DF TagI expression levels in Star cells by comparing the relative quantities obtained (Refer to Table 2.7). The highest soluble levels are found in the BL21 cells and the Star cells with *Pfn*-DF HisT and *Pfn*-DF TagI respectively.

Table 2-7: Relative quantities obtained after band analysis of all constructs in three different cell lines. The highest expression levels of total protein and soluble protein are indicated with red and green respectively. The synthetic constructs were compared to the native constructs.

Cell lines	Histag					
	Native			Synthetic		
	Total	Soluble	Insoluble	Total	Soluble	Insoluble
BL21 Star	2.00	1.51	1.32	4.25	1.42	1.43
BL21	1.70	3.70	3.78	0.54	1.36	1.33
BL21 Gold	1.56	2.88	2.60	0.93	2.74	1.27
Cell lines	**Tagless**					
	Native			Synthetic		
	Total	Soluble	Insoluble	Total	Soluble	Insoluble
BL21 Star	2.07	4.09	7.64	7.11	2.48	5.25
BL21	0.56	0.73	1.19	0.92	1.20	1.15
BL21 Gold	2.35	2.63	0.96	0.48	2.74	2.48

All of the ~60kDa bands were analysed using LC-MS/MS (refer to section 2.2.18) to establish whether the expression of the *P. falciparum* enzyme is detectable against the background expression level of the native *E. coli* enzyme. Analysis indicated that most of the soluble fractions contained *E. coli* proteins, which was expected based on the SDS PAGE analysis. *Pfs*–DF HisT/TagI proteins were detectable in the protein band analysed but the yield is unknown. Based on these results it was decided that the best construct to use in further experiments was *pPfs*-HisT to express *Pfs*–DF HisT in Star cells as it provides the highest amount of protein compared to the other constructs and cell lines and secondly, an added advantage of using *Pfs*–DF HisT is its ease of purification compared to *Pfs*-DF TagI. This decision was also supported by the previous finding where the synthetic construct showed higher complementation of cell growth compared to the native gene in functional complementation studies (refer to section 2.1.6) (Coetzee L., 2003) Thus the main aim was to ensure high levels of *Pfs*–DF HisT were obtained without the presence of the *E. coli* background proteins as they could possibly interfere with activity assays. Even though the SF4 cell line for complementation reduces the endogenous *E. coli* DHFS and FPGS activities to about 3% this cell line was not used for recombinant expression. This was based

on previous experiments, which indicated very low levels of *Pfs*–DF HisT/TagI are obtained and after about 8 hours cell growth decreases indicating possible toxicity of the *Pfs*–DF HisT/TagI to the cells (Coetzee L., 2003). Therefore several methods were investigated to improve soluble recombinant expression of *P. falciparum* DHFS-FPGS.

2.3.5 Unfolding/refolding of protein in inclusion bodies

In a first effort to isolate soluble *P. falciparum* DHFS-FPGS, protein was expressed from *p*Pfs-*HisT* in Star cells. More than 50% of the protein was present in the insoluble fraction as inclusion bodies. Isolated inclusion bodies were subsequently used for the unfolding and refolding of the protein using the method of Sirawaraporn *et al.* (section 2.2.13) (Sirawaraporn W. *et al.*, 1993). Refolded-soluble protein was purified with Ni^{2+} affinity chromatography and purified protein was analysed on SDS-PAGE gels as indicated in Figure 2.13. The refolded purified protein gave two bands at the expected protein size. The sizes were calculated from Rf values and the molecular mass standard curve to be ~60kDa (Figure 2.13 band 1) and ~57kDa (Figure 2.13 band 2). The equation used for this prediction was derived from the standard curve (Figure 2.14). The concentration of pure refolded protein obtained after resolublisation was 7.04 µg/ml (63.39 µg eluted protein in 9ml). Purified refolded protein was subsequently used in the determination of enzyme activity as described in section 3.1.3.
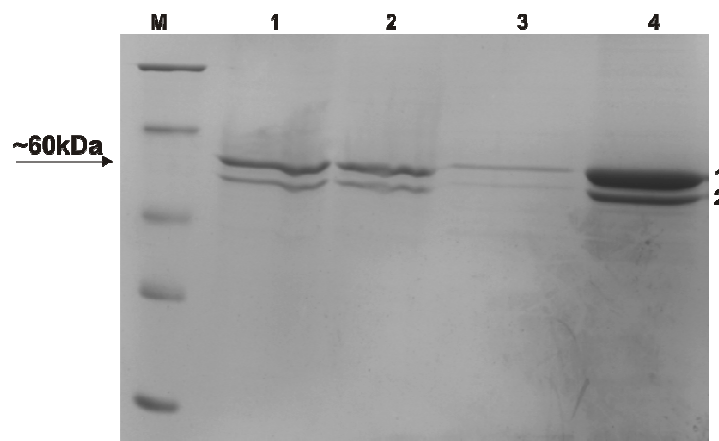
Figure 2-13: Affinity purification of resolubilised *Pfs*-DF HisT obtained from Star cells after Ni^{2+} chromatography. M: Protein Molecular Weight marker, Lane 1: Resolubilised protein (1.5mg loaded onto column), Lane 2: Flow through of binding of protein to Ni^{2+} matrix, Lane 3: Wash steps one and two, Lane 4: Eluted refolded protein. Band 1 (~60kDa) and Band 2 (~57kDa).
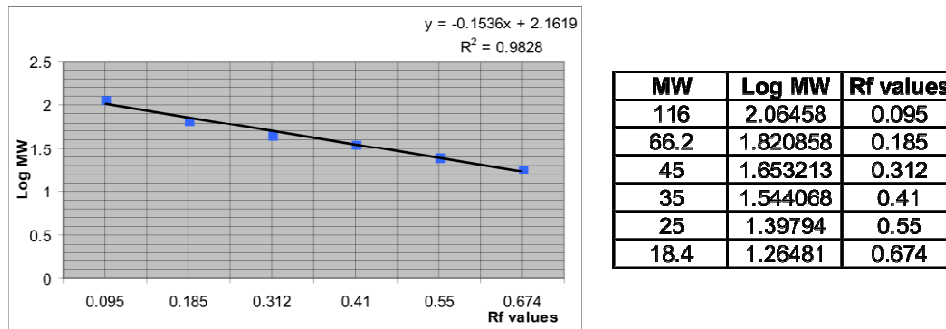
Figure 2-14: A logarithmic plot of molecular mass vs. Rf value. Data points correspond with 116kDa, 66.2kDa, 45kDa, 35kDa, 25kDa, 18.4kDa, respectively.

As can be seen from Figure 2.13 (Lane1) relative clean refolded-soluble protein is obtained from the inclusion bodies, suggesting that *Pfs*–DF HisT is the main protein aggregates in the inclusion bodies as no other proteins are visible. The protein was further purified to ensure removal of contaminants possibly present in small amounts and not visible as well as to concentrate the sample (Figure 2.13, Lane 4). In order to determine the identity of the second ~57kDa protein (band 2), the band was excised and analysed using LC-MS/MS (refer to section 2.2.18). MS analysis established the identity of this protein as an N-terminally truncated version of full length *Pf*. DHFS-FPGS (Figure 2.15). The identified peptides are indicated in red. In Band 1 (~60kDa) four peptides were located before position 150, whereas the first peptide identified in band 2 (~57kDa) starts at position 150 (truncated by 109 amino acids).

```
Band 1
    1 MEKNQNDKSN KNDIIHMNDK SGNYDKNNIN NFIDKNDEHD MSDILHKINN
   51 EEKKYEEIKS YSECLELLYK THALKLGLDN PKKLNESFGH PCDKYKTIHI
  101 AGTNGKGSVC YKIYTCLKIK KFKVGLFSSP HIFSLRERII VNDEPISEKE
  151 LIHLVNEVLN KAKKLYINPS FFEIITLVAF LHFLNKKVDY AIIETGIGGR
  201 LDATNILTKP EVIVITSIGY DHLNILGDNL PIICNEKIGI FKKDANVVIG
  251 PSVAIYKNVF DKAKELNCTI HTVVPEPRGE RYNEENSRIA LRTLEILNIS
  301 IDYFLKSIIP IKPPLRIQYL ATEQIQHIKK KFSPDNLEHN VQYPLAVILD
  351 VGHNETAIDR LCTDINYFHK GQNIRICISI TKPRNLSVFH PFIAQFGDTL
  401 KDIFYLPSLN ERTYDFEEIV EMLNNEEEIK NEIKELILSS SKKVGKWLAH
  451 EKQGNINEED ALKLYKRGCI PLIIKNAFLE CCKDNSILLV CGTFFVFDEV
  501 LNVFDIHSDM QDTIFMNEPS LV

Band 2 (Truncated N-terminal protein)
    1 MEKNQNDKSN KNDIIHMNDK SGNYDKNNIN NFIDKNDEHD MSDILHKINN
   51 EEKKYEEIKS YSECLELLYK THALKLGLDN PKKLNESFGH PCDKYKTIHI
  101 AGTNGKGSVC YKIYTCLKIK KFKVGLFSSP HIFSLRERII VNDEPISEKE
  151 LIHLVNEVLN KAKKLYINPS FFEIITLVAF LHFLNKKVDY AIIETGIGGR
  201 LDATNILTKP EVIVITSIGY DHLNILGDNL PIICNEKIGI FKKDANVVIG
  251 PSVAIYKNVF DKAKELNCTI HTVVPEPRGE RYNEENSRIA LRTLEILNIS
  301 IDYFLKSIIP IKPPLRIQYL ATEQIQHIKK KFSPDNLEHN VQYPLAVILD
  351 VGHNETAIDR LCTDINYFHK GQNIRICISI TKPRNLSVFH PFIAQFGDTL
  401 KDIFYLPSLN ERTYDFEEIV EMLNNEEEIK NEIKELILSS SKKVGKWLAH
  451 EKQGNINEED ALKLYKRGCI PLIIKNAFLE CCKDNSILLV CGTFFVFDEV
  501 LNVFDIHSDM QDTIFMNEPS LV
```

Figure 2-15: Mass Spectrometry analysis of DHFS-FPGS refolded protein. Band 1 (60kDa) and Band 2 (~57kDa) indicated in Figure 2-13. Identified peptides from Mascot are indicated in red.

The activity assays indicated the refolded purified protein to have a very low specific activity (section 3.2.2.2). Based on these results several other methods were investigated to increase soluble protein expression levels.

2.3.6 Auto-inducing studies

Based on work by Grossman *et al.* and Studier (Grossman T.H. *et al.*, 1998; Studier F.W., 2005), a series of auto-inducing experiments were performed to improve the solubility of *Pfs*-DF HisT. *pPfs*-HisT in Star cells was grown in the different auto-inducing media (Table 2.5) as described in section 2.2.14. Growth curves were determined to examine the efficiency of growth of *Pfs*-DF HisT in the different auto-inducing media. The growth curves were also used to compare the expression levels between the BL21 and Star cell lines. Figure 2.16 shows the *Pfs*-DF HisT growth curves found with 4xYT media as well as with 6% YE.

A

B
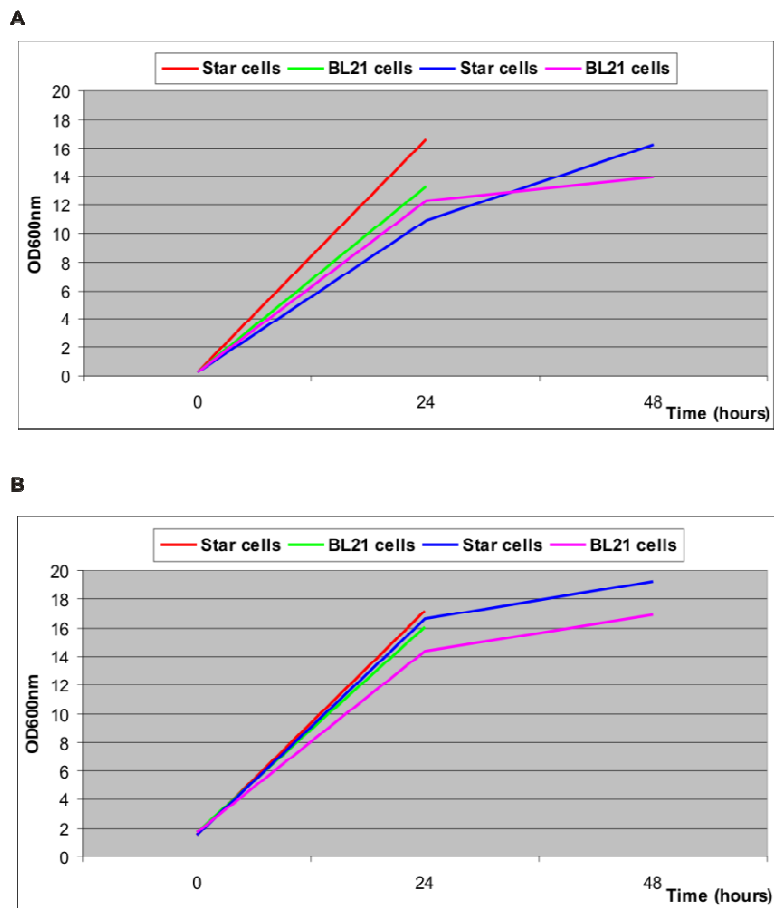
Figure 2-16: Growth curves obtained for *Pfs*-DF HisT after monitoring both BL21 and Star cell line growth in A) 4xYT media and B) 6% YE. In each media, the cell lines were monitored at 37°C for 24 hours (red and green lines) and at 18°C for 48 hours (blue and pink lines).

In both media, the Star cells showed better growth than the BL21 cells. In the 6% YE media, the cell growth is higher than in the 4xYT media. A comparative study was performed using the different inducing media as described in section 2.2.14 (Table 2.5) as compared with normal IPTG induction (section 2.2.10). *pPfs-HisT* was grown for 24 hours at 37°C in both 4xYT media and 6% YE respectively, as well as in the auto-inducing media for 16 hours at 37°C. The total and soluble fractions of protein expressed for each of these media were analysed on a 12% SDS-PAGE gel and compared to the normal IPTG induction (Figure 2.17 A). Samples were loaded qualitatively and stained with Colloidal Coomassie. Band analysis of the total protein was performed using the Quantity One software program. The relative quantity obtained form the Quantity One program was used in the comparative band analysis (refer to Figure 2.17 B).
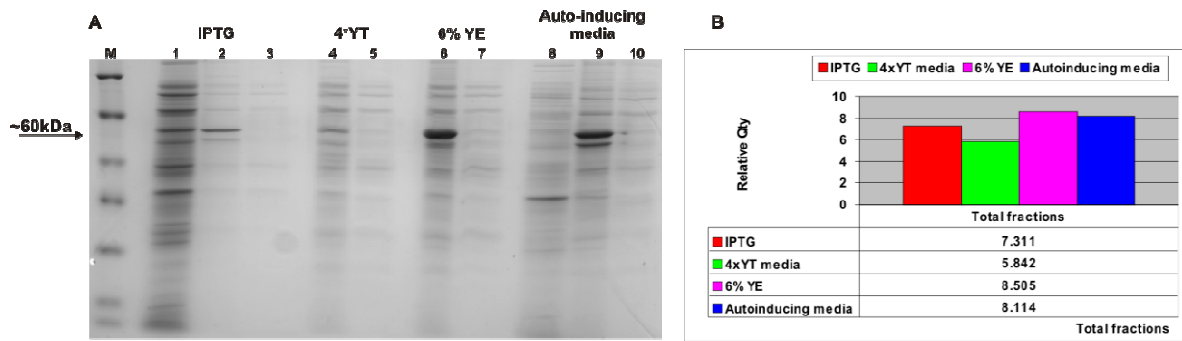
Figure 2-17: A) Comparison of protein expression levels in different media with *Pfs*-DF HisT Star cells. M: Protein Molecular Weight marker, Lane 1, 8: Uninduced samples, Lane 2, 4, 6, 9: Total fraction, Lane 3, 5, 7, 10: Soluble fraction. The arrow indicates the position of the expected 60kDa band. **B) Comparative band analysis of total protein expression levels in different media.** Analysis explained in text.

From the SDS- PAGE gel (Figure 2.17 A) it appears that the 6% YE media produced the highest amount of total protein. It is important to note that there is no apparent shift to production of soluble protein. The band analysis indicates that the 6% YE and the auto-inducing media produced the same amount of total protein (refer to Figure 2.17 B), while the 4xYT media produces much less *Pfs*-DF HisT. Due to the non-reproducibility of the 6% YE and the 4xYT media yielding the same amount of proteins with independent experiments (results not shown), it was subsequently decided to use the auto-inducing media for further experiments incorporating chaperone proteins (section 2.2.15) and detergent extraction (section 2.2.16).

Activity determination of *Pfs*-DF HisT using this auto-inducing media strategy showed that the DHFS-FPGS protein obtained was active (section 3.2.2.3).

2.3.7 Chaperone proteins

The *pPfs-HisT* and *pPfs-TagI* were co-expressed in Star cells with the pG-KJE8 chaperone plasmid. This plasmid was chosen as it contains the major chaperone proteins DnaK-DnaJ-GrpE and GroES-GroEL. The BL21 and Gold cell lines could not be used due to the presence of the pACYC origin of replication. These cell lines confer chloramphenicol resistance and use of Cm[r] gene as selection for the chaperone plasmid would therefore not be possible (refer to section 2.2.15). Co-transformed Star cells were grown in auto-inducing media supplemented with 10 ng/ml tetracycline and 4 mg/ml L-arabinose (refer to section 2.2.15). The protein expression obtained for *Pfs*-DF HisT is indicated in Figure 2.18. Results for *Pfs*-DF TagI are only indicated in analysis (Figure 2.19).
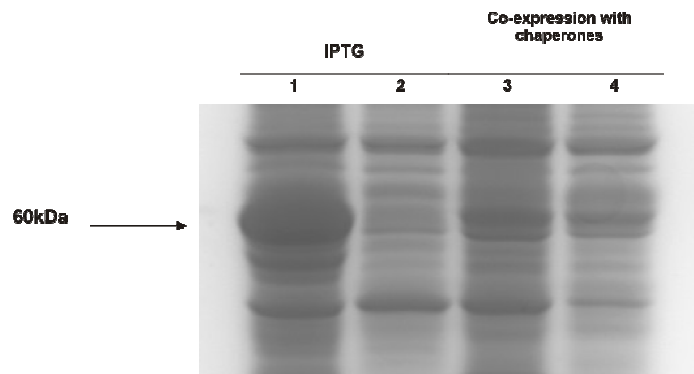
Figure 2-18: Co-expression of *pPfs*-DF *HisT* construct with pG-KJE8 chaperone plasmid in Star cells. *Pfs*-DF HisT expression on a 12% SDS-PAGE gel. Lane 1, 3: Total protein fraction. Lane 2, 4: Soluble protein fraction.

The fold differences obtained by co-expression of the chaperones on the total protein expression, as well as the distribution of the protein in the soluble fractions were determined by analysing the SDS-PAGE gels using the relative quantity obtained from the Quantity One program (Figure 2.19).

Fraction	Total	Soluble
Histag	12.5	3.8
Tagless	10.6	3.6
Histag with chaperones	5.6	5.4
Tagless with chaperones	5.3	5.8
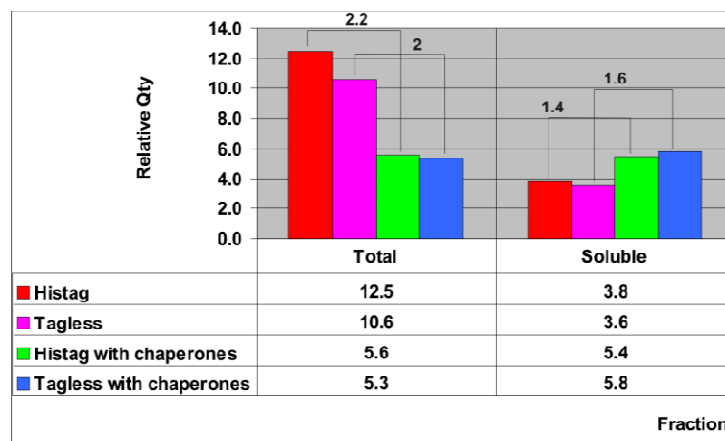
Figure 2-19: Comparison of total and soluble protein expression levels of *Pfs*-DF HisT and *Pfs*-DF TagI in the presence and absence of the chaperone proteins. Fold differences are indicated above bars.

In the presence of the chaperone proteins the SDS-PAGE gel analysis indicated a shift of about 1.4 fold and 1.6 fold to the soluble form of the *Pfs*-DF HisT and *Pfs*-DF TagI proteins, respectively (Figure 2.19). However, the overall total protein expression with co-expression of chaperones with both constructs is ~2-fold lower. Thus the use of chaperones in the expression system yields more soluble protein but the overall protein yield is much lower. Previous experiments indicated that purification of soluble fractions yield very little DHFS-FPGS protein and several contaminating proteins (Coetzee L., 2003). Thus purification of soluble protein would require a two-step purification procedure, requiring high yields of

soluble protein. Based on results indicating lower yield in protein an alternative method for solubilising the DHFS-FPGS protein was investigated.

2.3.8 Detergent extraction

DHFS-FPGS was expressed from *pPfs-HisT* in BL21 Star cells using the auto-inducing media. The insoluble fraction containing the inclusion bodies were subsequently used for detergent extraction of the protein using a combination of ionic and non-ionic detergents (refer to section 2.2.16). The initial sarcosine concentrations used were 0.3, 1, 3, 6, and 9%, which resulted in a shift of protein to the solubilised fraction (Figure 2.20). Lane 1 and 2 were loaded qualitatively, while the rest of the samples were loaded quantitatively.
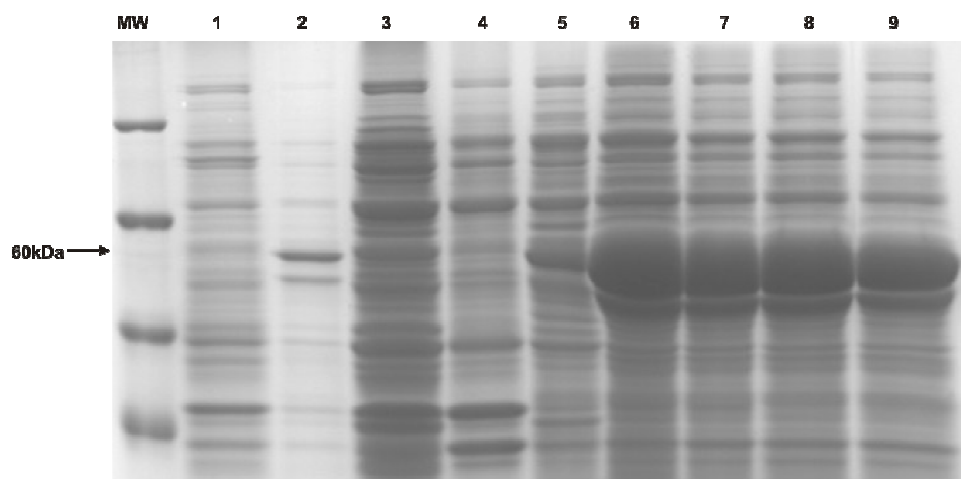
Figure 2-20: Comparison of *Pfs*-DF HisT expression levels obtained with detergent extraction. MW: Protein molecular weight marker, Lane 1: Uninduced sample, Lane 2: Total protein fraction, Lane 3: Soluble protein fraction, Lane 4: glucopyranoside fraction, Lane 5: 0.3% Sarcosine fraction, Lane 6: 1% Sarcosine fraction, Lane 7: 3% Sarcosine fraction, Lane 8: 6% Sarcosine fraction, Lane 9: 9%: Sarcosine fraction. The arrow indicates the position of the expected 60kDa protein.

Analysis of the SDS PAGE (Figure 2.20, Lane 5) indicates that the solubilisation occurs with as little as 0.3% sarcosine. The glucopyranoside fraction shows very little solubilisation of *Pfs*-DF HisT (Figure 2.20, Lane 4 and Figure 2.21, Lane 3). The presence of other contaminant proteins in the solubilised fractions is due to using the insoluble fraction directly without washing of inclusion bodies as in section 2.2.13. This can be included in future experiments to ensure more optimal resolubilisation and removal of contaminant co-aggregates. To ensure a minimal effect of detergent on enzyme activity, the lowest possible concentration of sarcosine, which resulted in solubilised protein, needed to be determined. As the most significant difference in solubilised protein was observed between 0.3% and 1% sarcosine (Fig 2.20), a refined concentration range of 0.5-1% was investigated. The effect of

the detergent extraction on the *E. coli* proteins was also investigated to ensure that no enrichment of *E. coli* proteins at ~60kDa occurred. An empty pET22b vector was expressed in Star cells and the insoluble fraction of *E. coli* proteins obtained was used for detergent extraction. Results are indicated in Figure 2.21 A. Samples were loaded qualitatively and stained with Colloidal Coomassie. Band analysis of each fraction of the ~60kDa protein was performed using the Quantity One software program. The relative quantity obtained form the Quantity One program was used in the comparative band analysis without normalisation (refer to Figure 2.21 B).
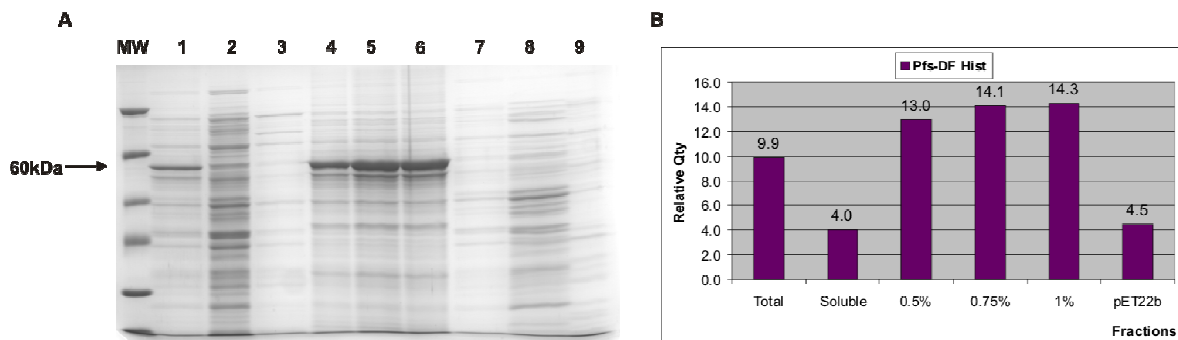
Figure 2-21: A) Comparison of *Pfs*-DF HisT and pET22b vector expression levels obtained with detergent extraction. MW: Protein molecular weight marker, Lane 1-6: Histag expression and Lane 7-9: pET22b expression. Lane 1,7: Total protein fraction, Lane 2,8: Soluble protein fraction, Lane 3: glucopyranoside fraction, Lane 4,9: 0.5% Sarcosine fraction, Lane 5: 0.75% Sarcosine fraction, Lane 6: 1% Sarcosine fraction. The arrow indicates the position of the expected 60kDa band. **B) Comparative band analysis of *Pfs*-DF HisT and pET22b vector expression levels obtained with detergent extraction.** Total and soluble refer to lane 1 and 2, the percentage indicated refer to the amount of sarcosine used to obtain solubilised protein (lane 4-6) and the pET22b refer to lane 9 (solubilised *E. coli* proteins).

Analysis of the relative quantities of the ~60kDa protein band obtained from Quantity One program indicated the solubilised fraction to be enriched 3.2-fold (13 vs 4) with 0.5% sarcosine and 3.5-fold (~14 vs 4) with 0.75% and 1% sarcosine compared with normal soluble protein expression (Figure 2.21 B). The initial evaluation of equal amounts of protein produced with 0.75% and 1% sarcosine were confirmed. Results indicated that possible *E. coli* proteins present in the insoluble fraction used for detergent extraction of DHFS-FPGS is minimal. This can be ignored as solubilised *Pfs*-DF HisT will be purified using Ni^{2+} affinity chromatography. Based on the comparative band analysis, it was decided to use the 0.5% sarcosine concentration in further experiments to purify the protein (refer to section 2.2.17).

2.3.9 Protein purification by affinity chromatography

Solubilised protein obtained from detergent extraction (161.5 µg) was loaded onto a Ni^{2+} column for purification. The purification results are indicated in Figure 2.22. Samples were loaded qualitatively on a 12% SDS-PAGE gel and stained using Colloidal Coomassie Blue.
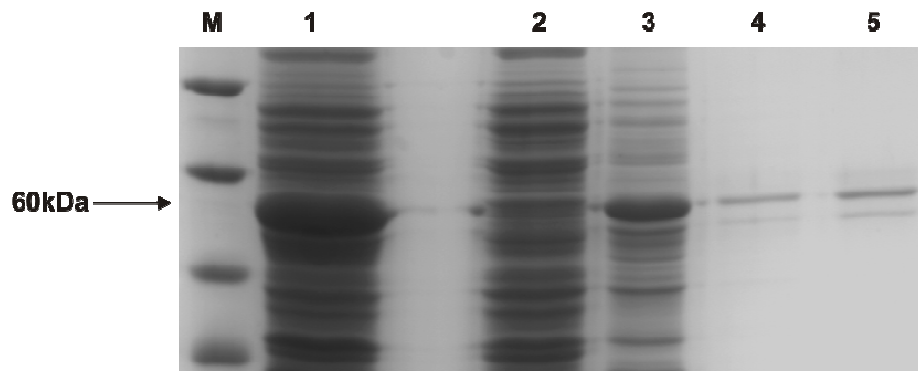
Figure 2-22: Affinity purification of detergent extracted *Pfs*-DF HisT obtained from Star cells on a Ni^{2+} column. M: Protein Molecular Weight marker, Lane 1: Total protein fraction, Lane 2: glucopyranoside fraction, Lane 3: 0.5% Sarcosine fraction, Lane 4: Wash steps 1 and 2, Lane 5: Eluted protein. The arrow indicates the position of the expected 60kDa band.

The *Pfs*-DF HisT was purified to homogeneity at ~5.02 µg/ml (7.53 µg in 1.5 ml elution fraction) as visualised by Coomassie on the SDS-PAGE gel (Figure 2.22). As can be seen, the second N-terminal truncated product is present in a much lower concentration than observed with the refolding of the inclusion body (Figure 2.13). The purified protein, as well as the fractions collected during the detergent extraction, were used to determine the activity of the DHFS-FPGS protein (refer to section 3.2.2.4) as well as in initial kinetic determinations (refer to section 3.2.2.5).

2.4 Discussion

2.4.1 Comparative protein expression of DHFS-FPGS constructs

Expression of recombinant proteins in sufficient amounts is desirable for biochemical characterisation. Over-expression of *P. falciparum* proteins in *E. coli* remains a challenge due to several contributing factors such as protein disorder, basic pI and protein size influencing both expression and solubility (Mehlin C. *et al.*, 2006; Vedadi M. *et al.*, 2007). Previously, a codon-optimised *dhfs-fpgs* gene was constructed to improve recombinant protein expression levels. Results indicated that the codon-optimised gene complemented a *E. coli* K-12 mutant strain, SF4 (F⁻ *folC strA recA tn10: srlC*) better than the native construct (section 2.1.6) (Coetzee L., 2003). Based on these results, the comparative expression levels between the synthetic and native *dhfs-fpgs* gene constructs were investigated. The full-length native *Pf. dhfs-fpgs* gene (~1573bp) was successfully amplified, cloned and sequenced to verify identity. The recombinant clones were subsequently used in a comparative expression study of p*Pfn-HisT/TagI* in the *E. coli* system. In-depth analyses showed that *Pfs*-DF TagI expression resulted in the highest quantity of total protein in BL21 Star cells followed by *Pfs*-DF HisT expression (Refer to Table 2.7). The highest fold difference in total protein expression was found in the BL21 Gold cells expressing *Pfn*-DF TagI compared to the *Pfs*-DF TagI total protein expression (Refer to Figure 2.11). However, comparison of the relative quantity of total protein expression levels indicated lower expression of the *Pfn*-DF TagI in the Gold cells than the *Pfs*-DF TagI / *Pfs*-DF HisT in the Star cells (Refer to Table 2.7). Thus, even though the fold difference indicated the native construct to produce the best expression levels, the overall yield of protein is much lower. Comparison of the soluble and insoluble protein fractions obtained from each construct indicated that distribution of soluble and insoluble DHFS-FPGS protein expression is ~50%. The most soluble protein was found with both *Pfn*-DF TagI in BL21 Star cells and *Pfn*-DF HisT in BL21 cells (Refer to Table 2.7). Based on the results it seems that the native gene gives a better yield of soluble protein however, upon LC-MS/MS analysis, a majority of *E. coli* proteins were identified from the soluble fractions for both the native as well as synthetic soluble fractions. Due to the high levels of background *E. coli* proteins identified comparisons of protein expression in the soluble fractions are inconclusive. Therefore, it was decided that the *pPfs-HisT* would be the best construct to use in subsequent solubilisation experiments, due to the high levels of total protein expression found in the BL21 Star cells, with the added advantage of one-step purification. The *pPfs-HisT* were also chosen based on the complementation studies that indicated the synthetic construct to complement better than the native construct.

2.4.2 Solubilisation and purification of *Pfs*-DF HisT

2.4.2.1 Unfolding/refolding of protein in inclusion bodies

High concentrations of active recombinant proteins are found in inclusion bodies (Villaverde A. and Carrio M.M., 2003). Due to *Pfs*-DF HisT yielding the highest total protein expression (mainly insoluble) the refolding of inclusion bodies was investigated as a source of active protein. Purified inclusion bodies were resolubilised with the aid of detergents such as urea and guanidinium chloride. The protein was then refolded by dilution (Carrio M.M. and Villaverde A., 2001; Singh S.M. and Panda A.K., 2005). The purified *Pfs*-DF HisT present in inclusion bodies was solubilised according to the method of Sirawaraporn *et al* (Sirawaraporn W. *et al.*, 1993). The refolding of the protein resulted in the production of about 7.04 μg/ml (63.39 μg in 9 ml) of pure protein after purification. However, electrophoretic analysis showed the presence of a second band. This was identified to be an N-terminal truncated product of the DHFS-FPGS protein. Investigation of the DNA sequence indicates that there are two extra start codons (AUG) present at position 51bp (Met-17) and 123bp (Met-41), which could have an influence on protein folding and the start of translation. Analysis of the ORF at 51bp with the Sequence Manipulation Site (Stothard P., 2000) calculated the *Pfs*-DF HisT MW to be ~58.32kDa and corresponds to the value predicted for the truncated protein (~57kDa) and have a predicted pI value of 6.65. The ORF at 123bp calculates the predicted MW to be ~55.32kDa and a pI of 6.94. The ~60kDa protein have a predicted pI value of 6.69. The reason for the truncation is unclear, but a possible explanation is that a hairpin loop forms at the beginning of the sequence resulting in translation only starting at position 51bp of the sequence. Alternatively, slippage of the translational machinery occurs due to incorrect pausing. This could be resolved by re-synthesis of the codon-optimised gene to a codon-harmonised sequence, which would aid the translational machinery in pausing at the correct sites and prevent slippage or the formation of hairpin loops (Kincaid R.L. *et al.*, 2002). Activity determination of the refolded protein indicated very low specific activities (refer to section 3.2.2.2). Based on these results several other methods were investigated to improve the yield of soluble *Pfs*-DF HisT expression.

2.4.2.2 Auto-induction studies

Grossman *et al.* and Studier showed that enhanced leaky expression of the pET system at the stationary phase of growth, results in higher expression of proteins than found with IPTG induction (Grossman T.H. *et al.*, 1998; Studier F.W., 2005). This leaky expression is partially

regulated by nutrient availability and metabolic balancing of pH, which consequently requires complex media for growth (Grossman T.H. *et al.*, 1998; Studier F.W., 2005). Based on their conclusions several auto-inducing media complexes (Table 2.5) were investigated for the increased soluble expression of *Pfs*-DF HisT as initial refolding protocols were not successful. Initial experiments indicated that both the 6% YE and 4xYT media showed increased levels of protein expression. However, reproducible expression levels could not be obtained with independent experiments. The auto-induction of cultures were ultimately performed using the method of Studier (Studier F.W., 2005). Even though both of these methods resulted in increased levels of protein expression, no visible enhancement of soluble protein expression was observed. Increased levels of total protein expression, is thus proportional to increased insoluble expression as no increase in soluble expression were observed. Consequently, even with the slower leaky expression of *Pfs*-DF HisT in *E. coli* there still seems to be translational problems resulting in incorrectly folded protein that aggregates in inclusion bodies.

2.4.2.3 Chaperone proteins

E. coli utilises three chaperones systems for *de novo* folding of proteins. These are TF, DnaK-DnaJ-GrpE and GroEL-GroES (Baneyx F. and Mujacic M., 2004). The co-expression of *Pfs*-DF-HisT with DnaK-DnaJ-GrpE and GroEL-GroES was investigated. Results indicated a shift to soluble protein expression, but at a ~ 2-fold lower total protein expression rate. The lower expression levels can be attributed to the chaperones redirecting the protein for folding, thus slowing down both the expression and folding systems. They may also possibly stabilise off-pathway intermediates, thus preventing the precipitation of proteins (Baneyx F. and Mujacic M., 2004). Even though soluble protein was obtained, the overall efficiency of co-expression of the chaperones was not beneficial, due the overall lower total protein expression levels. This decreased level of total protein can also be due to the over-expression of the chaperones as other researchers have observed a smaller increase in yield with the co-expression of GroEL-GroES (Dale G.E. *et al.*, 1994; Wall J.G. and Pluckthum A., 1995). Similar results were found with the DnaK-DnaJ-GrpE chaperones co-expressed with a tyrosine kinase, which led to an increase in solubility but a decrease in protein levels (Caspers P. *et al.*, 1994; Wall J.G. and Pluckthum A., 1995). Thus even though chaperones are widely used to improve soluble expression, optimisation is required to yield optimum protein levels. Possible strategies would include the presence of *P. falciparum* chaperones (PfHsp90 and PfHsp70) in the expression system (Banumathy G. *et al.*, 2003; Shonhani A. *et al.*, 2005).

2.4.2.4 Detergent extraction

An alternative strategy for the solubilisation of *Pfs*-DF-HisT was investigated. This involved the extraction of protein from the inclusion bodies using non-ionic (n-Octyl-ß-glucopyranoside) and ionic (Lauroyl sarcosine) detergents. (Angov E, Walter Reed Army Institute of Research, USA, personal communication). Initial investigations of sarcosine concentrations showed that as little as 0.3% increased the solubilised fraction of the protein. To ensure that the sarcosine does not interfere with activity assays, the lowest possible concentration of sarcosine giving soluble protein was used in further experiments. The protein was solubilised and purified to ~5.02 µg/ml (7.53µg in 1.5ml elution fraction) of pure protein visualised with Colloidal Coomassie Blue. This method also resulted in a lower concentration of the truncated protein being present, which is advantageous for the kinetic analysis. During the refolding of the solubilised protein, the refolding pathway competes with the aggregation pathway (nonnative intermolecular hydrophobic interactions). Thus, prevention of these interactions is crucial for successful formation of the native structure (Vallejo L.F. and Rinas U., 2004). Less truncated protein is refolded, as it would form more nonnative interactions compared to the full-length protein. The background *E. coli* protein levels were also investigated. Results indicated that low levels of *E. coli* proteins enrich during detergent extraction. This was expected as minimal *E. coli* proteins are found in the inclusion bodies (most of the *E. coli* proteins are found in the soluble fraction). In recombinant expression systems the main integral protein found in the inclusion body are the recombinant protein. Possible contaminant proteins are found due to co-precipitation of insoluble cell material during inclusion body recovery. Thus, increased washing and isolation of inclusion bodies ensure the removal of these contaminants and solubilisation and refolding of only the recombinant protein (Vallejo L.F. and Rinas U., 2004; Ventura S. and Villaverde A., 2006). Thus even though a minimal amount of *E. coli* proteins do enrich during the solubilisation of the target protein it can be disregarded as the protein was purified further using affinity chromatography. Purified solubilised protein was shown to be active and used in further activity assays and for kinetic evaluation (refer to section 3.2.2.4 and 3.2.2.5).

In summary, the detergent extraction protocol yields the best protein levels for subsequent analyses, and was used in activity and kinetic determinations as described in Chapter 3.

Chapter 3:
Activity and kinetic properties of the *P. falciparum* DHFS-FPGS

3. Introduction

The reaction mechanism of bifunctional DHFS-FPGS is given in Figure 3.1. Dihydropteroate is converted to dihydrofolate after the addition of the first glutamate residue through an amide linkage to the carboxyl end of the *p*-amino benzoic acid (*p*ABA) group, catalysed by DHFS. The double bond in dihydrofolate is reduced to give tetrahydrofolate, a reaction catalysed by DHFR. This reduction takes place in a NADPH/H$^+$-dependent reaction at position 5/6 and 7/8 of the pyrazine ring (Bollheimer L.C. *et al.*, 2005). In the final polyglutamation reaction, additional glutamate residues are added by FPGS to the γ-position (Hyde J.E., 2005; Mathieu M. *et al.*, 2005). This takes place in a K$^+$- and Mg^{2+}-ATP-dependent manner. Through the formation of an acyl phosphate intermediate the phosphate group of ATP activates the γ-carboxyl group of the dihydropteroate substrate (from DHFS) or the glutamate group of folic acid (FPGS). This is followed by a nucleophillic attack by the amino group of the incoming glutamate to form the glutamylated folate (Sheng Y. *et al.*, 2003). The polyglutamate tail length (n) varies between 1 and 10 and is usually shorter in prokaryotes than in eukaryotes (Bognar A.L. *et al.*, 1986). Rapidly dividing cells mostly have polyglutamates as the predominant folate cofactors for one-carbon metabolism (Stanger O., 2002). The increased cellular retention of folates and affinity for folate-utilising enzymes, as well as the accumulation of folates in the mitochondria where glycine synthesis takes place requires polyglutamylation (McGuire J.J. and Bertino J.R., 1981; Moran R.G., 1999). The reactive parts of the polyglutamates are the nitrogen atoms at position N$^{5/10}$ where the displacement of hydrogen atoms by a one-carbon unit generates the different, biologically active folates (Bollheimer L.C. *et al.*, 2005).
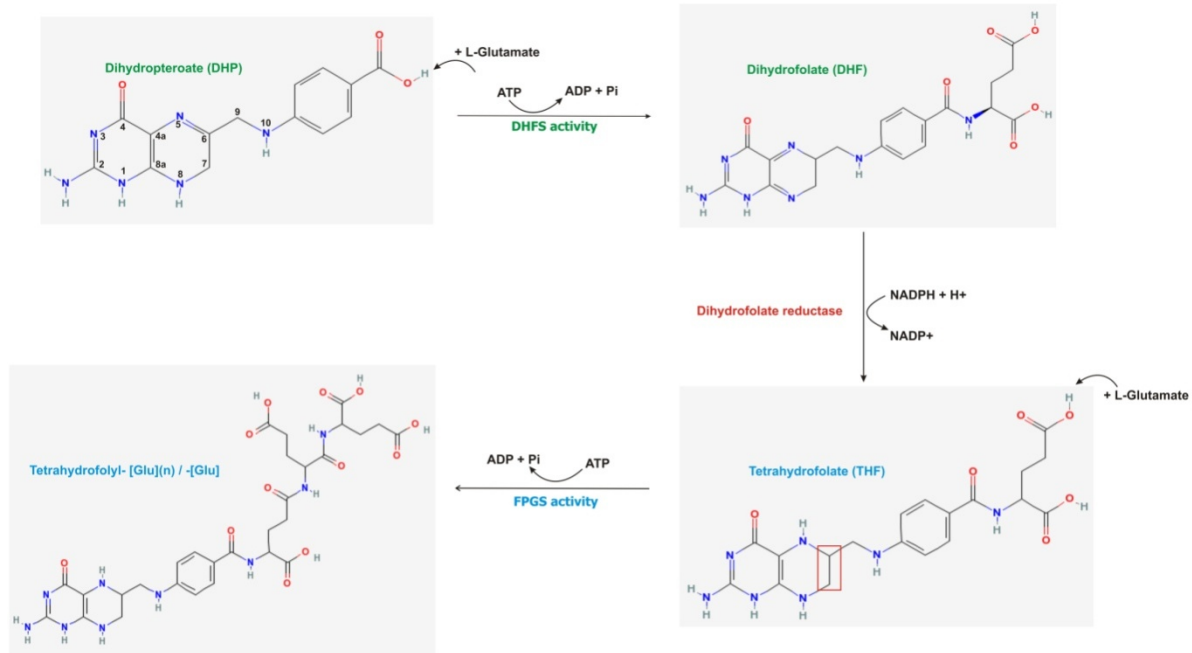
Figure 3-1: The DHFS-FPGS reaction mechanism. Dihydropteroate is converted to dihydrofolate by the addition of an L-glutamate (DHFS). The double bond is reduced to give tetrahydrofolate (DHFR) indicated by red square. Polyglutamation occur at the γ-linked position of THF (FPGS). Glutamate additions occur in an MgATP dependent manner. Adapted from (Mathieu M. *et al.*, 2005).

Several studies have investigated the mechanism of formation of pteroylpoly-γ-glutamates in organisms such as *Lactobacillus casei*, *Streptococcccus faecalis* and *Corynebacterium* species (Shane B., 1980a; Shane B., 1980b; Shane B. *et al.*, 1983). The major folates found for *L. casei* and S*. faecalis* were octa- and nonaglutamates, respectively. Results also indicated higher activity in the presence of the diglutamate substrates than with the monoglutamates (Shane B. *et al.*, 1983). Analysis of the FPGS in *Corynebacterium* species showed the enzyme to be most stable in the presence of KCl and phosphate and the presence of the K^+ ion is essential for enzyme activity (Shane B., 1980a). Other requirements were identified to be $MgATP^{2-}$, as a nucleotide substrate due to the ineffectiveness of $CoATP^{2-}$ and $MnATP^{2-}$ to act as nucleotide substrates. The enzyme also showed a much higher Km for the diglutamate substrates than for the monoglutamates (Shane B., 1980a; Shane B., 1980b). The mechanism of folylpolyglutamate synthase was found to be sequential and an ordered Ter-Ter mechanism (Shane B., 1980b). Thus, the order of substrate addition was MgATP, THF and glutamate and the author suggested that the mechanism proceeds via a pteroyl-γ-glutamyl phosphate intermediate. The same requirement was also observed for the *E. coli*, *L. casei*, *Neurospora crassa,* human and rat liver enzymes. Thus in general, for all folylpolyglutamates synthetases, the presence of a folate substrate, MgATP, L-glutamate, as well as the monovalent cation K^+ are indispensable

(Bognar A.L. *et al.*, 1986; Chen L. *et al.*, 1996; McDonald D. *et al.*, 1995; McGuire J.J. *et al.*, 1980; Shane B., 1980a; Shane B. and Bognar A.L., 1983).

Due to limited information regarding the *P. falciparum* DHFS-FPGS enzyme, structural and biochemical investigations are urgently needed to aid in the design of mechanism-based inhibitors capable of arresting parasite growth. For this, sufficient amounts of correctly folded, soluble DHFS-FPGS are needed to determine the kinetic properties of the protein and ultimately the three-dimensional structure of the enzyme.

Chapter 2 described the optimisation of expression of *P. falciparum* DHFS-FPGS from the synthetic *dhfs-fpgs* gene. This chapter focuses on the protein obtained from the synthetic *P. falciparum dhfs-fpgs* gene and determination of its functional activity and kinetic properties.

3.1 Methods

3.1.1 Glutamate standard curve

A stock solution of 12.5 mM glutamate (25 mM glutamate and 1μCi L-[U-^{14}C] glutamic acid) was diluted with 100 mM Tris, 50 mM Glycine buffer (pH 9.75-10) to give the following final concentration range: 1.25 mM, 625 μM, 312.5 μM, 156.25 μM, 62.5 μM, 31.25 μM, and 15.625 μM Glutamate. 10 or 15 μl of each concentration was spotted and dried onto DE-81 Whatman anion-exchange paper (pretreated with 2.5 mM EDTA, pH 8 for 1 hour and dried). The glutamate binds to the positively charged paper via ionic interactions. The DE-81 paper was exposed onto a Phosphor imaging screen (Bio-Rad, Hercules, USA) for 24 hours and scanned in at a 100 microns using the Personal Molecular Imager$^®$ FX (Biorad, Hercules, USA).

3.1.2 Protein concentration determination

Protein concentrations were determined according to the Folin-Lowry method (Lowry O.H. *et al.*, 1951). The determination involves two distinct steps: 1) The Copper(II)ion in an alkaline solution reacts with the protein to form complexes with functional groups of tyrosine, tryptophan and cysteine, known as the Biuret reaction. 2) These complexes react with the Folin-Ciocalteu reagent. The product becomes reduced to molybdenum/tungsten blue and can be detected calorimetrically by absorbance at 500-750 nm (Lowry O.H. *et al.*, 1951). The Biuret reaction alone is insensitive but addition of the Folin-Ciocalteu reagent to detect the reduced copper increases the assays sensitivity. Presence of strong acids or ammonium sulfate can interfere with the assay. BSA was diluted to obtain a standard concentration range of 300, 240, 210, 180, 150, 120, 90, 60 and 30 μg/ml protein. To each BSA standard and protein sample, 300 μl of Solution ABC (20:1:1) was added (Solution A: 2% Na_2CO_3 in 1 M NaOH; Solution B: 1% $CuSO_4$. H_2O; Solution C: 2% potassium tartrate (($C_4H_4K_2)_6.5H_2O$)). Samples were incubated for 15 minutes. After addition of 900μl of a 10% 2N Folin-Ciocalteu (Sigma-Aldrich, St Louis, USA) (mixture of phosphotungstate and phosphomolybdate in phenol reagent), samples were incubated for an additional 45 minutes in the dark. The absorbance of the samples was measured at 690 nm in a 96 well microtitre plate (Bibby Sterilin Ltd., Stone Staffordshire, UK) on a Multiskan Ascent spectrophotometer (Thermo Labsystems).

3.1.3 Activity determinations

Tetrahydrofolate (10 mM THF, 0.1 M potassium phosphate buffer pH 7, 0.1% (w/v) ascorbic acid) and 7-8 Dihydropteroate (5 mM DHT, 25 mM potassium phosphate buffer pH 7, 0.1% (w/v) ascorbic acid) were used as substrates to determine the activities of FPGS and DHFS, respectively. Both substrates were dissolved in phosphate buffer containing ascorbic acid to prevent the oxidation of the folates. To further protect the folates, the solutions were flushed with liquid nitrogen, aliquoted and stored at -20℃ in the dark (Kallen R.G. and Jencks W.P., 1966; Kok R.M. *et al.*, 2004; Ramasastri B.V. and Blakley R.L., 1964). The enzyme activity assay measures the incorporation of [^{14}C] glutamate into folylpolyglutamates using THF (FPGS assay) and DHT (DHFS assay) as substrates. The assays were adapted from Bognar *et al.*, 1986. Reaction mixtures contained Tris-Glycine buffer (100 mM Tris, 50 mM Glycine, pH 9.75), 100 μM THF, 50 μM DHT, 0.25 μCi L-[U-^{14}C] Glutamic acid, 5 mM ATP, 10 mM MgCl$_2$, 200 mM KCl, 5 mM DTT, Dimethyl sulfoxide (DMSO) (10 μl), 10 μg Bovine serum albumin (BSA), 0.25 mM unlabelled glutamate and enzyme (~100 ng – 5 μg) in a final volume of 100 μl. The reaction tubes were capped and incubated at 37℃ for 2 hours (Bognar A.L. *et al.*, 1986). The reaction was stopped by spotting 10 or 15 μl of the reaction mixture onto DE-81 Whatman paper (pretreated with 2.5 mM EDTA pH 8 for 1 hour and dried). The excess reagents were washed off using an 80 μM NaCl/10 mM Tris, pH 7.5, wash buffer for 20 min. The anion-exchange paper was subsequently exposed onto a Phosphor imaging screen (Bio-Rad, Hercules, USA) for 24 hours and scanned in at a 100 microns using the Personal Molecular Imager® FX (Biorad, Hercules, USA). The purified DHFS-FPGS protein from *Toxoplasma gondii* (Dr P. Sims, University of Manchester, UK) was used as a positive control in the assays. The protein was assayed as mentioned above.

3.1.4 Detection of Radioactivity

The phosphor screen is composed of a barium fluorobromide matrix doped with europium (BaFBr:Eu^{2+}) and has a sensitivity of 1.0-^{14}C dpm/mm^2/hour. When the radioactive decay emission collides with the BaFBr:Eu^{2+} screen, phosphor oxidation of Eu^{2+} occurs and a high energy state is formed due to the electron movements. The Eu^{2+} is oxidised to Eu^{3+} and the BaFBr is reduced to BaFBr$^-$. These ions remain in the higher energy state when the screen is removed from the sample thus storing the energy from ionising radiation (Figure 3.2, stage 1). When the internal laser, which emits light at a wavelength of 635nm, moves over the activated site the BaFBr$^-$ absorbs the light energy freeing electrons and reducing Eu^{3+} to a lower energy state Eu^{2+} (stage 2). This reduction of trapped energy releases photons that are captured by a photomultiplier tube and seen as a digital file. The image displayed shows

the darkness of each pixel at a specific sample location, which is proportional to the signal intensity of the radioactive sample (stage 3).
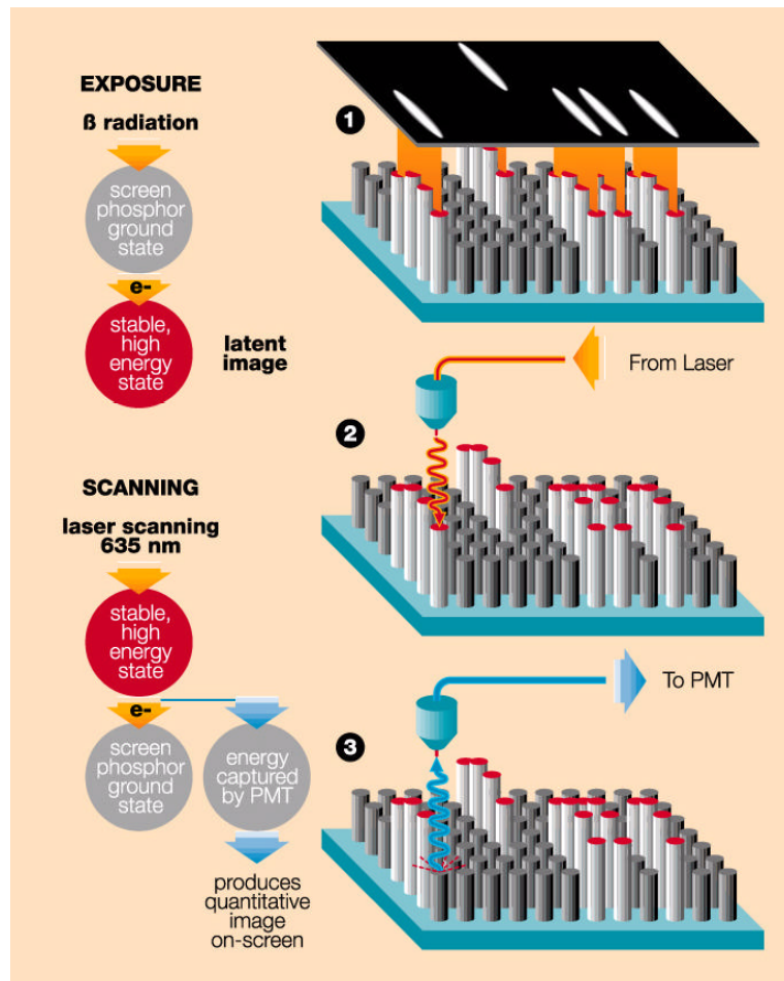
Figure 3-2: **The storage phosphor detection mechanism** (BioRad).1.Exposure of irradiation to phosphor screen, 2: Scanning of screen, 3: Image of radioactive sample.

The image obtained from the Quantity One program was analysed using the volume detection function, (Volume = Sum of the intensities of the pixels within the volume boundary x pixels (intensity units x mm^2)). The intensity of the background volume was subtracted from each spot intensity using the global background subtraction function of the Quantity One program. The adjusted intensities were used to calculate the counts per hour, which were used to obtain the number of pmoles of product in the reaction from the glutamate standard curve. The pmoles of product is used in further calculations to determine the activity of the enzyme (section 3.1.5).

3.1.5 Calculations

1 Unit of enzyme activity: = $\dfrac{\text{Amount of product}}{\text{time}}$

= pmol/h

- Time = 2 hours (incubation period of enzyme assay)

Specific activity: = Unit/ng = pmol/h/ng

- The amount of ng used in calculation are calculated from the amount of protein (ng) present in the specific spot volume

Fold purification: = $\dfrac{\text{specific activity of fraction}}{\text{specific activity of crude}}$

3.2 Results

3.2.1 Glutamate standard curve

The glutamate assay was performed to obtain a standard curve (nmoles product/ counts/hour) from which the activities in pmoles/h were determined using the linear equation of the graph. Reactions were stopped by spotting (10 or 15 µl) onto the pre-treated anion-exchange DE-81 Whatman paper. The assays were exposed for 24 hours onto a Phosphor imaging screen. The image obtained from the Quantity One program (Figure 3.3 A) was analysed using the volume detection function. Intensities for each specific concentration of the glutamate assay were used to calculate the counts per hour. The counts per hour were plotted against the nmoles of glutamate in the reaction to obtain the standard curve as given in Figure 3.3 B.
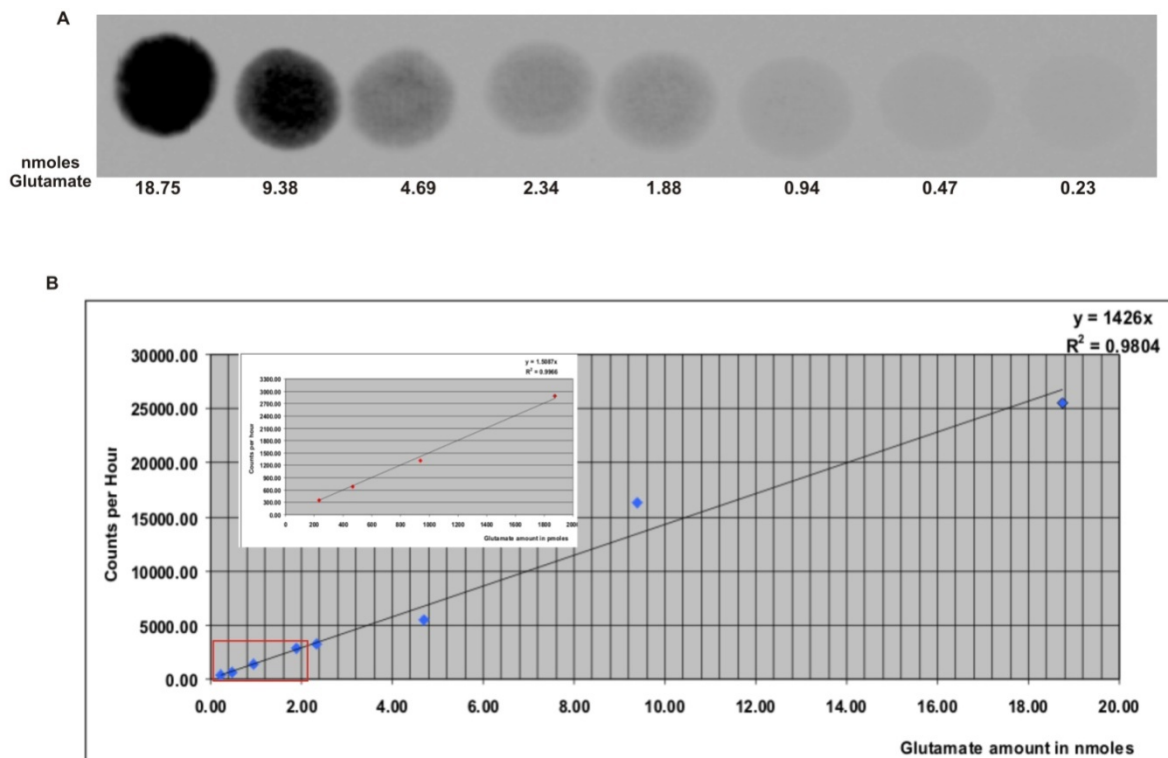
Figure 3-3: Glutamate standard curve. A: Intensities of concentration range of glutamate in nmoles, B: Standard curve of glutamate intensities. The inset indicates an enlargement of the lower end of the standard curve indicated by the red square.

The equation of the standard curve was used in the activity assays to determine the amount of product produced from the counts obtained after 24 hours. The pmoles of product produced are used in further calculations (section 3.1.5).

3.2.2 Activity determinations

3.2.2.1 Determination of *E. coli* background proteins

A control experiment was performed to determine if background *E. coli* protein activity influences the activity determination of *Pfs*-DF HisT in the total and soluble samples. Protein expression of cells only (Star cells), Star cells with empty vector (pET22b) and *pPfs-HisT* in Star cells (*Pfs*-DF HisT) were performed as in section 2.2.14. Protein concentrations of the total (soluble + insoluble) as well as soluble fractions were determined using the Lowry method (section 3.1.2). Each fraction was used for the assays. Assays were performed as given in section 3.1.3. 10 or 15 µl of reactions were spotted in triplicate onto the pre-treated anion-exchange DE-81 Whatman paper (Figure 3.4).
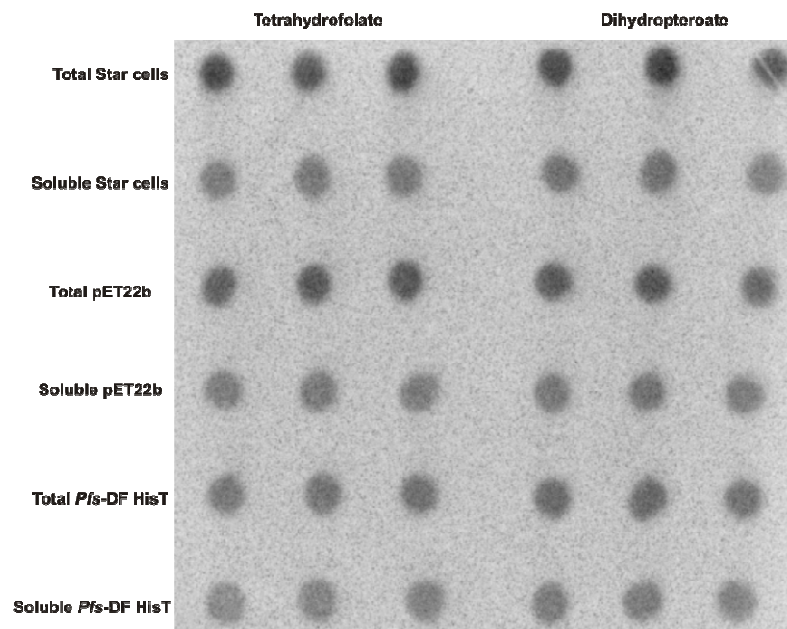
Figure 3-4: Activity determination of *E. coli* background proteins in total and soluble fractions. Samples used in the assay are indicated on the left and the specific substrate on the top of the figure. Star cells indicate the activity of Star cells with no vector, pET22b indicates activity of Star cells with empty vector, and *Pfs*-DF HisT indicates activity of *pPfs-HisT* in Star cells.

The image obtained from the Quantity One program (Figure 3.4) was analysed using the volume detection function to determine the counts per hour. Each spot was encircled as a volume to determine specific intensity. The counts obtained from this intensity after 24 hours for activity assays were converted to amount of pmoles using the glutamate standard curve equation. The number of pmoles determined was used in the calculation of specific activity of each fraction (section 3.1.5) and are indicated in Figure 3.5. The ng or µg of protein in a specific spot volume was used in the calculations. For the total samples, the values were 39

µg, 31 µg and 30 µg for the Star cells, pET22b and *Pfs*-DF HisT, respectively. The amount of protein used in calculations for the soluble samples are 207 ng, 198 ng and 364 ng for the Star cells, pET22b and *Pfs*-DF HisT, respectively.
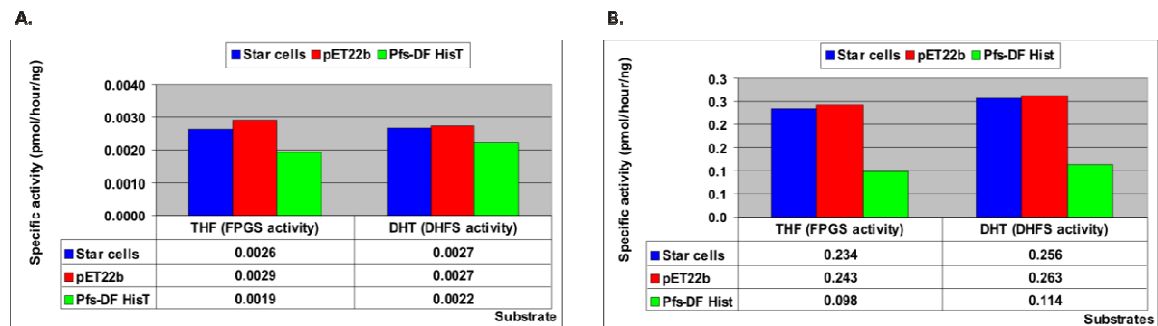
Figure 3-5: Specific activities of protein from the A) total fractions and B) soluble fractions. Star cells indicate the expression of Star cells with no vector, pET22b indicates expression of Star cells with empty vector, and *Pfs*-DF HisT indicates expression of *pPfs-HisT* in Star cells.

The results indicate that the higher specific activity was obtained in the total protein fraction of the pET22b expression followed by expression from the Star cells and then in the *Pfs*-DF HisT (Figure 3.5 A). The same trend was observed with the activity (pmoles/hour) for each of the constructs (results not shown). It is difficult to draw a conclusion from the specific activities obtained for the total protein fraction because of the presence of the insoluble proteins and aggregates interfering with the assay. Results from the soluble protein fraction (Figure 3.5 B) indicate almost similar specific activities for proteins expressed from Star cells and the pET22b vector. Inclusion of *pPfs-HisT* in the Star cells lowers the specific activity of proteins obtained from the soluble fraction with about 2.3 fold. Thus the *E. coli* background proteins do seem to influence the activity determination of *Pfs*-DF HisT in the total and soluble fractions. From these results it seems that the expression of *pPfs-HisT* might be toxic to the cells since activities declined in comparison to the other constructs. After two independent experiments, analysis of the OD_{600nm} measurements obtained following overnight induction indicated expression levels as follows: *Pfs*-DF HisT/Star cells > Star cells > pET22b/Star cells with a ratio of 3:2:1. This is contradictory to toxic proteins, which would decrease cell growth. The presence of *Pfs*-DF HisT in the cells probably decreases soluble expression overall due to metabolic stress, thus lower specific activities are detected in the soluble samples. Furthermore there could be an interaction between the *E. coli* DHFS-FPGS and *Pfs*-DF HisT that results in the *E. coli* DHFS-FPGS co-aggregating with the insoluble *Pfs*-DF HisT.

Due to the presence of the *E. coli* DHFS-FPGS activity in the total and soluble fractions it is necessary to purify *Pfs*-DF HisT to determine correct activities for this enzyme. In several of the experiments it was impossible to purify *Pfs*-DF HisT due to the very low concentration of the protein in the soluble sample. In such cases the specific fractions (total or soluble) of each experiment were compared with each other to deduce a conclusion and activity of *Pfs*-DF HisT.

3.2.2.2 Unfolding/refolding of protein in inclusion bodies

Purified *Pfs*-DF HisT obtained from the refolding of protein from the inclusion bodies (section 2.2.13) and Ni^{2+} affinity chromatography (section 2.2.17) was used to determine the activity of the purified *Pfs*-DF HisT protein (Sirawaraporn W. *et al.*, 1993). The *Pfs*-DF HisT fraction after refolding and purification was used for the assays. Assays were performed as given in section 3.1.3. 10 or 15 µl of reactions were spotted in triplicate onto the pre-treated anion-exchange DE-81 Whatman paper (Figure 3.6). The positive control included in the assay was the His-tagged purified DHFS-FPGS protein of *T. gondii* to determine the effectiveness of the assay as well as to compare activities of *Pfs*-DF HisT to that of *T. gondii.* Protein concentrations of all the fractions during refolding were determined using the Bradford method (section 2.2.11).
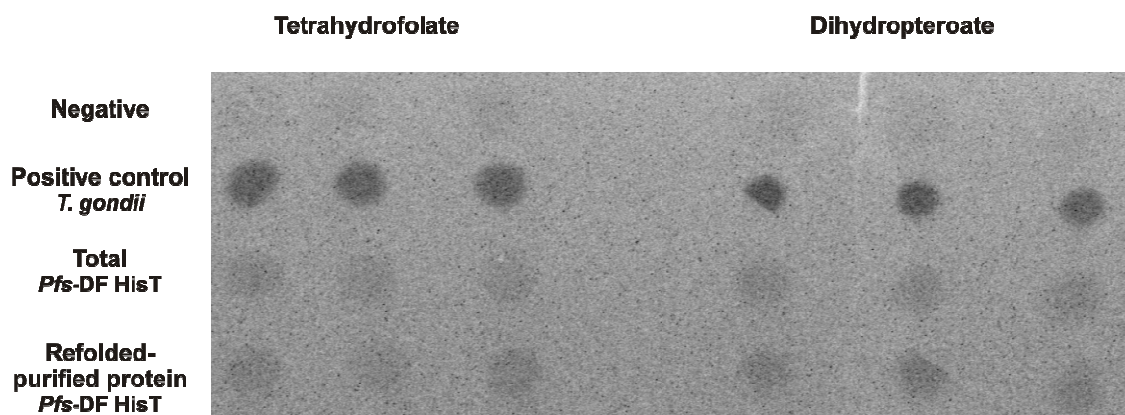
Figure 3-6: Activity determination of *Pfs*-DF HisT isolated with the unfolding/refolding protocol and Ni^{2+} affinity chromatography. Samples used in assay are indicated on the left and the specific substrate on the top of the figure.

The image obtained from the Quantity One program (Figure 3.6) was analysed using the volume detection function and glutamate standard curve as mentioned before. To determine the activity of the refolded purified protein, the activities as well as specific activities of the total and the refolded protein were determined using the pmoles of product obtained and the ng of enzyme in the spot volume (section 3.1.5) (Table 3.1). The amount of total *Pfs*-DF HisT

protein and refolded purified *Pfs*-DF HisT protein in the enzyme spot volume were calculated to be 16 ng and 108 ng, respectively. As can be seen from Table 3.1 the refolded purified protein shows a much lower specific activity (0.84 and 0.77 pmol/h/ng) than the total sample (6.02 and 6.2 pmol/h/ng) for the FPGS and DHFS activities, respectively. From the results, it appears that the *Pfs*-DF HisT is inactive, as insignificant enrichment of activity is taking place after purification. However these low values of *Pfs*-DF HisT may indicate active protein as much lower levels of specific activities were found in soluble and total samples with expression of *Pfs*-DF HisT in a recombinant system (Figure 3.5). Therefore, it was initially concluded that the refolding protocol yields low levels of active protein and other methods were investigated to obtain higher levels of pure active protein.

Table 3-1: Activities of *Pfs*-DF HisT isolated with the unfolding/refolding protocol (Sirawaraporn W. *et al.*, 1993) and Ni^{2+} affinity chromatography. The average was determined from three spot volumes and experiment was performed once.

THF (FPGS activity)							
Samples	Average	Standard deviation	Counts per hour a	pmoles b	Activity pmoles/h c	Specific activity pmol/h/ng d	Protein present
Positive control	211343.67	3934.771	8805.99	256.49	128.25		*T. gondii*
Total	158504.00	10300.886	6604.33	192.37	96.18	6.02	*E. coli* proteins + *Pfs*-DF HisT
Refolded purified protein	148977.33	5886.811	6207.39	180.80	90.40	0.84	*Pfs*-DF HisT
DHT (DHFS activity)							
Samples	Average	Standard deviation	Counts per hour a	pmoles b	Activity pmoles/h c	Specific activity pmol/h/ng d	Protein present
Positive control	176904.67	7930.161	7371.03	214.70	107.35		*T. gondii*
Total	163436.33	4420.678	6809.85	198.35	99.18	6.20	*E. coli* proteins + *Pfs*-DF HisT
Refolded purified protein	137823.00	10684.807	5742.63	167.27	83.63	0.77	*Pfs*-DF HisT

a. Counts per hour = Average/24 (24 hours exposed on the Phosphor screen)

b. pmoles = Determined with the glutamate standard curve using the following equation:

$$\text{pmoles product} = (y = 1426x) * 1000; \quad y = \text{counts per hour and } x = \text{nmoles}$$

c. h= incubation time of 2 hours for assay

d. ng = specific protein amount (ng) in spot volume

3.2.2.3 Inducing media

The activity of *Pfs*-DF HisT was measured in the total and soluble fractions obtained from expression in 4xYT media, as well as auto-inducing media (section 2.2.14). *Pfs*-DF HisT was not purified due to low levels in the soluble fraction (refer to section 2.3.6, Figure 2.17) and the refolding protocol yielding levels of active protein that were too low (refer to section 3.2.2.2). Each *Pfs*-DF HisT fraction during expression was used for the assays. Assays were performed as given in section 3.1.3. Reaction mixtures were performed in duplicate and 10 or 15 µl of each reaction mixture was spotted in duplicate for the 4xYT media and 4-times each for the auto-inducing media onto the pre-treated DE-81 Whatman paper (Figure 3.7). Protein concentrations of all the fractions were determined using the Bradford method (section 2.2.11).
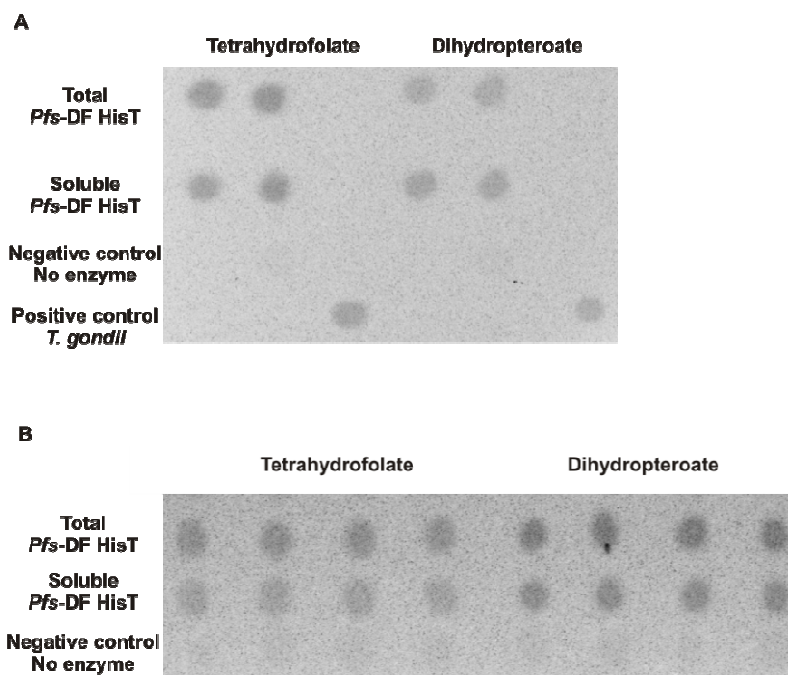
Figure 3-7: Activity determination of *Pfs*-DF HisT isolated from A) 4xYT media and B) auto-inducing media. Samples used are indicated on the left and the specific substrate on the top of the figure.

The samples were analysed as before using the Quantity One program and the glutamate standard curve. The amount of product in pmoles for the 4xYT media (Table 3.2) and auto-inducing media (Table 3.3) were used in further calculations to determine the activity as well as specific activity of *Pfs*-DF HisT from both media (section 3.1.5). The amount of protein of the total fractions in the spot volume was 36 ng and 509 ng for the 4xYT media and auto-inducing media, respectively. The amount of protein of the soluble fractions in

the spot volumes was 16 ng and 382 ng for the 4xYT media and auto-inducing media, respectively.

Table 3-2: Activities of *Pfs*-DF HisT in 4xYT media. The average was determined from two spot volumes and the experiment was performed once.

THF (FPGS activity)							
Samples	**Average**	**Standard deviation**	**Counts per hour** [a]	**pmoles** [b]	**Activity pmoles/h** [c]	**Specific activity pmol/h/ng** [d]	**Protein present**
Positive control	224915.34 (1)	N/A	9371.47	272.97	136.48		*T. gondii*
Total	274211.80	56818.21	11425.49	332.79	166.40	4.68	*E. coli* proteins + *Pfs*-DF HisT
Soluble	155313.38	42798.66	6471.39	188.49	94.25	5.87	*E. coli* proteins + *Pfs*-DF HisT
DHT (DHFS activity)							
Samples	**Average**	**Standard deviation**	**Counts per hour** [a]	**pmoles** [b]	**Activity pmoles/h** [c]	**Specific activity pmol/h/ng** [d]	**Protein present**
Positive control	141906.47 (1)	N/A	5912.77	172.22	86.11		*T. gondii*
Total	323733.84	55157.54	13488.91	392.90	196.45	5.53	*E. coli* proteins + *Pfs*-DF HisT
Soluble	331685.31	14696.32	13820.22	402.55	201.27	12.54	*E. coli* proteins + *Pfs*-DF HisT

Table 3-3: Activities of *Pfs*-DF HisT in auto-inducing media. The average was determined from four spot volumes and the experiment was performed once.

THF (FPGS activity)							
Samples	**Average**	**Standard deviation**	**Counts per hour** [a]	**pmoles** [b]	**Activity pmoles/h** [c]	**Specific activity pmol/h/ng** [d]	**Protein present**
Total	160653.68	6571.89	6693.90	194.98	97.49	0.19	*E. coli* proteins + *Pfs*-DF HisT
Soluble	139382.58	6356.19	5807.61	169.16	84.58	0.22	*E. coli* proteins + *Pfs*-DF HisT
DHT (DHFS activity)							
Samples	**Average**	**Standard deviation**	**Counts per hour** [a]	**pmoles** [b]	**Activity pmoles/h** [c]	**Specific activity pmol/h/ng** [d]	**Protein present**
Total	242980.00	10053.52	10124.17	294.89	147.45	0.29	*E. coli* proteins + *Pfs*-DF HisT
Soluble	175391.74	12736.76	7307.99	212.86	106.43	0.28	*E. coli* proteins + *Pfs*-DF HisT

The activities of total proteins (*E. coli* and *Pfs*-DF HisT) (166.4 and 196.45 pmoles/h) are in the same magnitude to the activities found with purified *Toxoplasma gondii* DHFS-FPGS (136.48 and 86.11 pmoles/h) in the 4xYT media (Table 3.2). The 4xYT media shows a 1.25-fold increase in FPGS activity and a 2.27-fold increase in DHFS activity in the soluble sample (Figure 3.8). The auto-inducing media shows a small increase in activity in the soluble sample (Table 3.3 and Figure 3.8). Comparison of protein expression levels in the 4xYT media reveal much lower levels of *Pfs*-DF HisT than in the auto-inducing media (refer to section 2.3.6, Figure 2.17). Thus the higher specific activities of the soluble proteins (*E. coli* and *Pfs*-DF HisT) in the 4xYT media could be due to the lower levels of *Pfs*-DF HisT, which was shown to have a negative effect on the expression of *E. coli* enzymes (section 3.2.2.1). This negative effect is substantiated with the auto-inducing media, which expresses high amounts of *Pfs*-DF HisT, but result in very low levels of specific activities for the soluble proteins (*E. coli* and *Pfs*-DF HisT) (Figure 3.8).
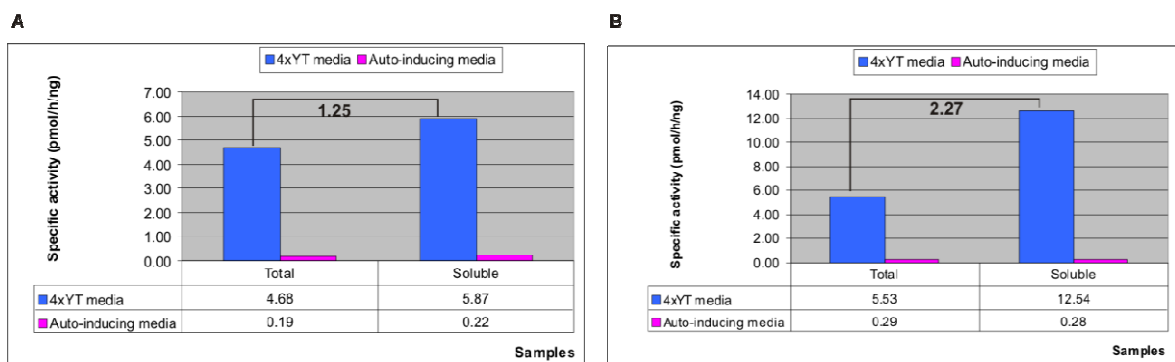
Figure 3-8: Relative fold changes of *Pfs*-DF HisT specific activity in 4xYT media and auto-inducing media. **A) FPGS activity, B) DHFS activity** Fold differences are indicated above the bars.

Based on the expression results (Figure 2.17), which indicate no increase in soluble expression, together with the activity determinations (Table 3.2 and 3.3), detergent extraction was subsequently investigated to increase the soluble protein for purification and determination of the *Pfs*-DF HisT activity.

3.2.2.4 Detergent extraction

Initial detergent extraction analysis of *Pfs*-DF HisT indicated that 0.3% sarcosine yielded solubilised protein (section 2.3.8). To determine the effect of the sarcosine on *Pfs*-DF HisT activity, the isolation protocol included 0.5% sarcosine and 0.75% sarcosine and both the isolations were used in the enzyme assays. Activity assays were performed (section 3.1.3) on the samples obtained during the detergent extraction of the protein (refer to section

2.2.16) and after Ni^{2+} purification (refer to section 2.2.17). *Pfs*-DF HisT reaction mixtures were performed in duplicate and 10 or 15 µl of each reaction was spotted in duplicate onto the pre-treated anion-exchange DE-81 Whatman paper (Figure 3.9).
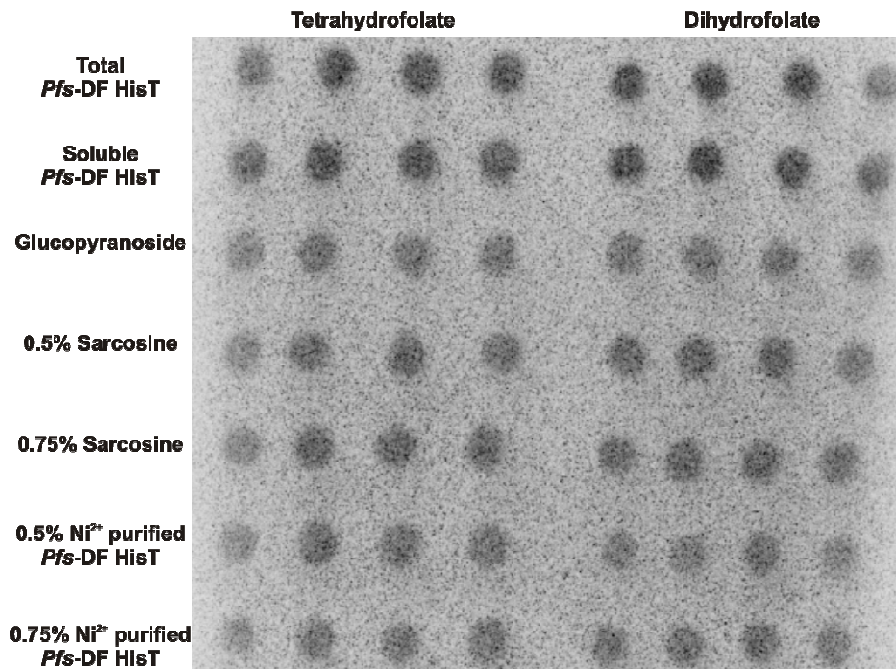
Figure 3-9: Activity determination of *Pfs*-DF HisT isolated with detergent extraction (0.5% and 0.75% sarcosine) and Ni^{2+} affinity chromatography. Samples used in the assay are indicated on the left and the specific substrate on the top of the figure. Assays in duplicate and spotted in duplicate. Total and Soluble *Pfs*-DF HisT: samples taken after normal expression; glucopyranoside: soluble fraction obtained after treatment of pellet with glucopyranoside; 0.5%/0.75% sarcosine: solubilised protein after sarcosine treatment of pellet with specific sarcosine concentration; 0.5%/0.75% Ni^{2+} purified *Pfs*-DF HisT: purified solubilised Histag protein.

Samples were analysed as mentioned previously using the glutamate standard curve and the protein quantities as given in Table 3.4. Analysis indicated that both concentrations of sarcosine yielded active *Pfs*-DF HisT after Ni^{2+}-protein purification (Table 3.5).

Table 3-4: Protein concentrations used for calculating the specific activity of *Pfs*-DF HisT during detergent extraction. Determined using the Bradford method (section 2.2.11)

Sample	Protein amount in spot volume
Total	654 ng
Soluble	690 ng
Glucopyranoside	158 ng
0.5% Sarcosine	704 ng
0.75% Sarcosine	683 ng
0.5% Ni^{2+}Purified *Pfs*-DF HisT protein	5.2 ng
0.75% Ni^{2+}Purified *Pfs*-DF HisT protein	6.7 ng

Table 3-5: Activities of *Pfs*-DF HisT during detergent extraction using 0.5% and 0.75% sarcosine and Ni^{2+} affinity chromatography. The average was determined from four spot volumes (duplicate reactions) and the experiment was performed once.

Samples	Average	Standard deviation	Counts per hour [a]	pmoles [b]	Activity pmoles/h [c]	Specific activity pmol/h/ng [d]	Protein present
Total	10543.45	2408.36	159.61	111.93	55.96	0.09	*E. coli* proteins + *Pfs*-DF HisT
Soluble	10967.52	574.82	177.28	124.32	62.16	0.09	
glucopyranoside	9440.25	249.34	113.64	79.69	39.85	0.25	
0.5% sarcosine	9838.75	314.89	130.25	91.34	45.67	0.06	
0.75% sarcosine	10271.31	594.39	148.27	103.98	51.99	0.08	
0.5% Ni^{2+}Purified *Pfs*-DF HisT protein	9859.95	707.43	131.13	91.96	45.98	8.82	*Pfs*-DF HisT
0.75% Ni^{2+}Purified *Pfs*-DF HisT protein	8940.06	408.23	92.80	65.08	32.54	4.84	*Pfs*-DF HisT
DHT (DHFS activity)							
Samples	Average	Standard deviation	Counts per hour [a]	pmoles [b]	Activity pmoles/h [c]	Specific activity pmol/h/ng [d]	Protein present
Total	10958.37	2101.49	172.16	120.73	60.37	0.09	*E. coli* proteins + *Pfs*-DF HisT
Soluble	11853.03	1547.83	209.44	146.87	73.44	0.11	
glucopyranoside	10673.19	951.16	160.28	112.40	56.20	0.36	
0.5% sarcosine	11705.66	1123.45	203.30	142.57	71.28	0.10	
0.75% sarcosine	12049.19	688.45	217.61	152.61	76.30	0.11	
0.5% Ni^{2+}Purified *Pfs*-DF HisT protein	10896.64	934.64	169.59	118.93	59.46	11.41	*Pfs*-DF HisT
0.75% Ni^{2+}Purified *Pfs*-DF HisT protein	10964.29	894.30	172.41	120.91	60.45	9.00	*Pfs*-DF HisT

Analysis of quantity of product formed (tetrahydrofolyl-[Glu](n) and dihydrofolate, respectively) with *Pfs*-DF HisT indicated that 91.34 and 142.57 pmoles of product are formed when protein is extracted and solubilised with 0.5% sarcosine, compared to 103.98 and 152.61 pmoles of product when protein is extracted and solubilised with 0.75% sarcosine (Table 3.5). Analysis of quantity of product formed with purified *Pfs*-DF HisT indicated the same trend, where more pmoles of product are formed using protein extracted and solubilised from 0.75% sarcosine (Table 3.5). From these analyses it seems that the 0.75% sarcosine protocol yields more active *Pfs*-DF HisT based on the amount of pmoles of product formed. The DHFS/FPGS specific activities of *Pfs*-DF HisT were determined after extraction, solubilisation and Ni^{2+} purification using 0.5% and 0.75% sarcosine to confirm these results (Table 3.5 and Figure 3.10). The specific activity of solubilised *Pfs*-DF HisT with 0.5% and 0.75% indicates approximately the same values of ~0.06 and ~0.1 pmol/h/ng for FPGS and DHFS activities, respectively. The specific activities of purified solubilised *Pfs*-DF HisT indicated the enzyme lost activity as the sarcosine concentration in the extraction protocol increases. Specific activity decreases for the FPGS and DHFS activities from 8.82 (0.5%) to 4.84 (0.75%) and from 11.41 (0.5%) to 9 (0.75%), respectively (Figure 3.10). Thus even though the *Pfs*-DF HisT from solubilisation with 0.75% showed higher pmoles of product formed, the 0.5% sarcosine concentration yields more active, purified, solubilised *Pfs*-DF HisT based on the specific activities. The loss in activity of purified *Pfs*-DF HisT after 0.75% solubilisation could be due to the 0.75% sarcosine interfering with the Ni^{2+} affinity column, however approximately the same concentration of protein is purified with the 0.5% and 0.75% sarcosine sample (Table 3.4). Thus, it seems that the 0.75% sarcosine might rather have a detrimental effect on the *Pfs*-DF HisT activity during extraction and solubilisation than interfering with the Ni^{2+} affinity column.
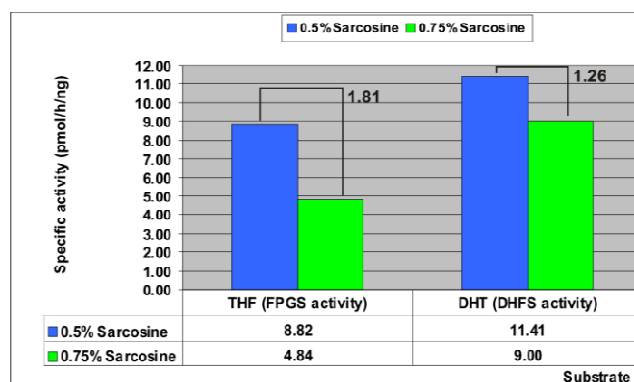
Figure 3-10: Fold differences in specific activity of purified solubilised *Pfs*-DF HisT. Comparison between 0.5% sarcosine and 0.75% sarcosine specific activities. Fold differences are indicated above the bars.

Analysis of Table 3.5 indicates the enrichment of specific activity occurrs at two steps of the detergent protocol. The first enrichment is at the glucopyranoside fraction (0.25 and 0.36 pmol/h/ng) and then secondly at the purification step of *Pfs*-DF HisT (8.82 and 11.41 pmol/h/ng).The fold differences in specific activity between the soluble fraction and the glucopyranoside fraction (partially solubilised protein) are indicated in Figure 3.11 A. The fold differences are 2.7-fold and 3.3-fold for the DHFS and FPGS specific activities respectively. From these results, it seems that one could have discontinued the protocol at the glucopyranoside step as it already yields higher levels of active *Pfs*-DF HisT. As can be seen from Figure 2.20, lane 4 and Figure 2.21, lane 3, the glucopyranoside fraction contains several contaminating proteins and the *Pfs*-DF HisT is present at a much lower concentration than found with the sarcosine fractions. Purification of *Pfs*-DF HisT in the glucopyranoside fraction by means of Ni^{2+} affinity chromatography thus needs to be performed to rule out the effect of the background *E. coli* proteins, as well as to address important questions such as: can the protein be purified successfully and if so, is the specific activity still higher than the soluble sample as well as the purified sarcosine fraction? As mentioned in section 2.3.8 some of the background *E. coli* proteins can be removed by initial washing of the inclusion bodies before detergent extraction. The fold differences in specific activity were also compared between the purified solubilised *Pfs*-DF HisT and the solubilised protein from the 0.5% sarcosine fraction. *Pfs*-DF HisT showed a fold increase of 114-fold and 147-fold for DHFS and FPGS specific activities, respectively compared to the solubilised protein obtained from the 0.5% sarcosine fraction (Figure 3.11 B). Thus higher levels of active protein were obtained after purification.
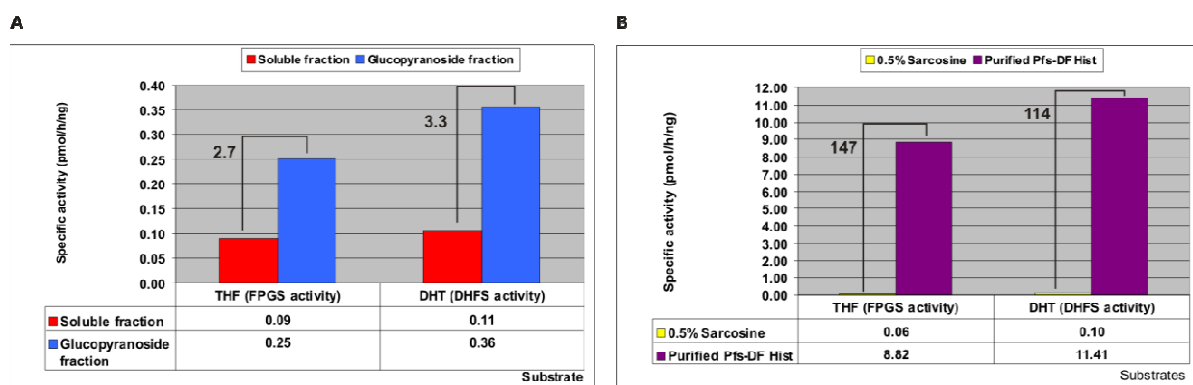
Figure 3-11: Fold differences in specific activity of *Pfs*-DF HisT. A) Comparison of specific activities between soluble fraction and glucopyranoside fraction of *Pfs*-DF HisT. **B)** Comparison of specific activities between 0.5% sarcosine fraction and purified refolded-soluble *Pfs*-DF HisT. Fold differences are indicated above the bars.

Comparison between the specific activities of the purified solubilised *Pfs*-DF HisT obtained from the detergent extraction (8.82 and 11.41 pmol/h/ng) and the refolded purified *Pfs*-DF HisT from the refolding protocol (0.84 and 0.77 pmol/h/ng) (section 3.2.2.2) suggest that the use of non-ionic and ionic detergents yields more active protein. Thus the fold purification is actually much higher because the soluble fraction contains both *E. coli* and *P. falciparum* proteins, while the solubilised pellet only contains *Pfs*-DF HisT. From the refolding experiments the optimum concentration of sarcosine to obtain active *Pfs*-DF HisT protein is 0.5%. The 0.5% sarcosine concentration protocol was used to obtain pure active *Pfs*-DF HisT for subsequent kinetic analysis. The percentage substrate to product conversion was determined to be 0.09% and 0.23% for FPGS and DHFS activity, respectively indicating the substrate to be in excess, which is essential for kinetic analysis.

3.2.2.5 Kinetic analysis

For any enzyme kinetic experiment the linearity of the enzyme assay with respect to the enzyme concentration is essential. This is determined at low enzyme concentration compared to substrate concentration. An enzyme concentration is chosen from this linear range and the effects of varying substrate concentrations are determined. The data obtained would result in a Michealis Menten graph from which the Km and Vmax values can be determined. Due to DHFS-FPGS being a two substrate enzyme, the experiment should be done twice, firstly with a constant high pterin concentration and varying the glutamate concentration and secondly by using a constant glutamate concentration and varying the pterin concentration. Using both dihydropteroate and tetrahydrofolate as pterin substrates the kinetic constants for both the DHFS-FPGS activities can be determined. The purified solubilised *Pfs*-DF HisT obtained after treatment with 0.5% sarcosine and Ni^{2+} affinity chromatography was used in initial kinetic experiments to determine the linearity of the assay. An enzyme concentration range of 0.2, 0.5, 0.75, 1.00, 2.5 ng/µl were assayed as described in section 3.1.3. 10 or 15 µl of reaction mixtures were spotted in duplicate onto the pre-treated anion-exchange DE-81 Whatman paper (Figure 3.12).
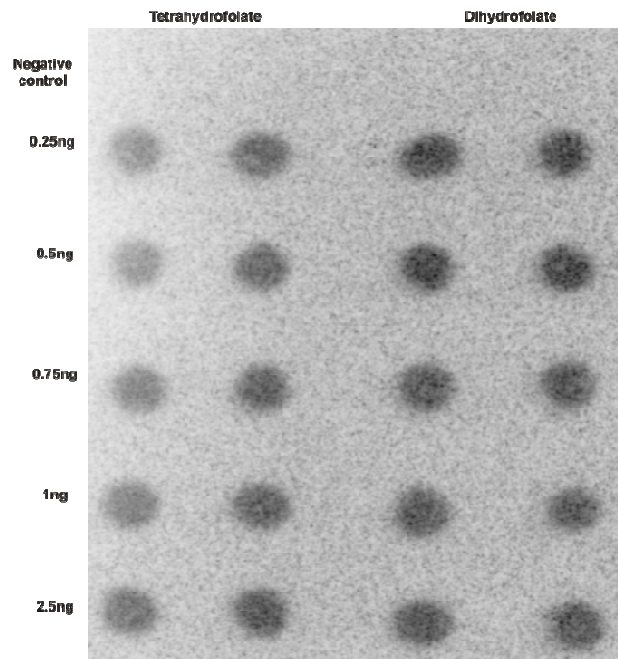
Figure 3-12: *Pfs*-**DF HisT (DHFS/FPGS) assay vs. enzyme concentration.** Concentrations of *Pfs*-DF HisT used in the assay are indicated on the left and the specific substrate on top of the figure. Samples were spotted in duplicate.

The samples were analysed as before using the Quantity One program and the glutamate standard curve. The amount of product in pmoles was plotted against enzyme concentration to determine linearity of enzyme (Figure 3.13).
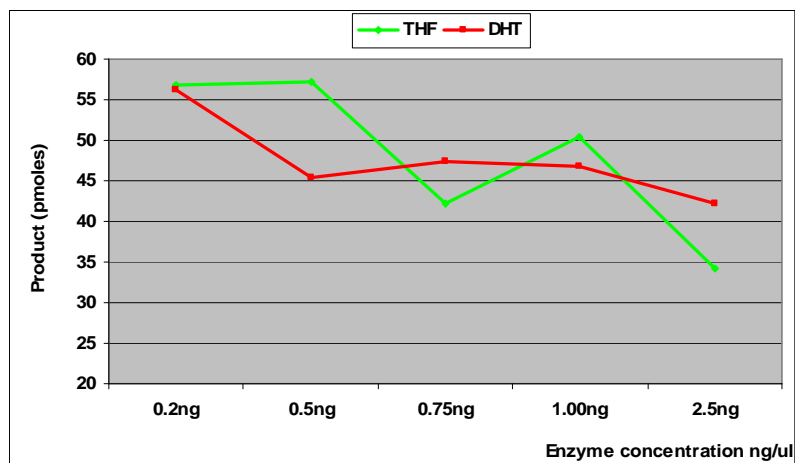
Figure 3-13: Analysis of linearity of enzyme assay. The amount of product in pmoles is plotted against enzyme concentration.

From the results obtained (Figure 3.13) it is clear that there is no linear increase in product formation with increasing *Pfs*-DF HisT concentration. Initially it was thought that the imidazole present in the purified fraction might interfere with the enzyme assay.

Solubilisation and purification was performed as before and the purified fraction dialysed to remove the imidazole. The same trend of non-linearity was observed (results not shown). The same trend was also observed with *Pfs*-DF HisT obtained by affinity purification of the soluble sample (results not shown). It is unclear why this trend is observed. It could be due to increments between the *Pfs*-DF HisT concentrations that were too small. Increasing the increments requires higher concentrations that were not possible due to the low yield of *Pfs*-DF HisT obtained after solubilisation and purification. The very low concentrations of purified *Pfs*-DF HisT (~4.57 ng/µl), could also result in the destabilisation of the enzyme. Another possible explanation could be the removal of stabilising factors during the purification, which also affects the activity of *Pfs*-DF HisT. Due to these unexpected results a control experiment was performed to determine if the DHFS-FPGS activity assay is effective. Total protein fractions obtained from protein expression of Star cells (no vector) and *pPfs-HisT* in Star cells (*Pfs*-DF HisT) were used in an assay with increasing volumes (section 3.1.3). A doubling in activity was expected with doubling of protein. 10 or 15 µl reactions were spotted onto the pre-treated anion-exchange DE-81 Whatman paper and analysed with the Quantity One program. The amounts of pmoles of product are indicated in Figure 3.14 A and the activities (pmol/hour) are indicated in Figure 3.14 B.
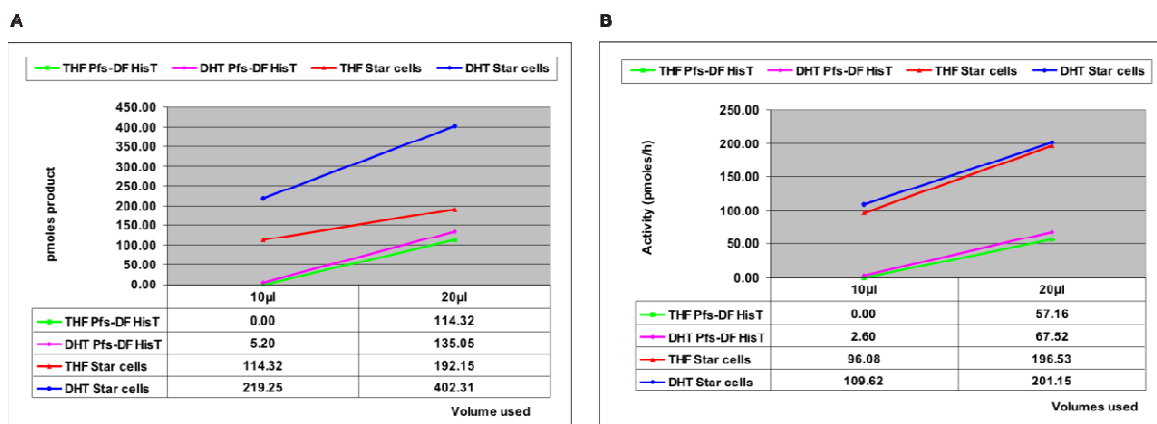
Figure 3-14: Determination of effectiveness of DHFS-FPGS assay. A) Comparison of pmoles product formed between total fractions of Star cells and *Pfs*-DF HisT. **B)** Comparison of activity (pmoles/h) between total fractions of Star cells and *Pfs*-DF HisT. Star cells indicate the expression of Star cells with no vector, and *Pfs*-DF HisT indicates expression of *pPfs-HisT* in Star cells. THF= FPGS activity, DHT= DHFS activity.

The results indicate a doubling of pmoles product (Figure 3.14 A) and activity (Figure 3.14 B) with doubling of the total fraction of Star cells only. However, Star cells with the synthetic gene construct (*Pfs*-DF HisT) shows only a detectable enzyme activity and pmoles product formed at 20 µl. No discernable activity or pmole product formation was

observed at the lower volume of protein fraction (10 µl). Furthermore, as observed before, the activities of DHFS-FPGS were again much lower than the corresponding activities of the Star cells alone. As mentioned earlier, the much lower activities of *Pfs*-DF HisT might possibly be due to other contributing factors such as additional proteins being expressed to protect the *E. coli* from the high levels of *Pfs*-DF HisT present, co-precipitation of the *E. coli* DHFS-FPGS or lower soluble protein levels (refer to section 3.2.2.1 and Figure 3.5). Thus it is clear that the expression of *Pfs*-DF HisT is not toxic to the *E. coli* cells; however it results in much lower activities being observed. To determine the true extent of the problem at the protein level would be difficult as the normal *E. coli* DHFS-FPGS represents about 0.01% of the soluble proteins expressed (Bognar A.L. *et al.*, 1986).

Due to these unanticipated results, the kinetic determinations could not be completed for *Pfs*-DF HisT. Several strategies are currently being investigated to improve protein yield, as well as optimisation of the assay conditions for kinetic determinations.

3.3 Discussion

3.3.1 Background *E. coli* proteins

To determine the effect of *E. coli* background proteins on the enzyme assays, activity determinations were performed on proteins isolated from empty Star cells, Star cells containing an empty pET22b vector and Star cells containing *pPfs-HisT.* The results indicated that the Star cells and pET22b proteins to have approximately the same specific activities in the total and soluble fractions. The inclusion of the *pPfs-HisT* construct to the expression system gave much lower specific activities for the proteins obtained from the total and soluble fractions (refer to Figure 3.5). The decrease in specific activity in the total fractions could be attributed to higher levels of inclusion bodies present, as most of the *Pfs*-DF HisT is expressed mainly as insoluble aggregates. The insolubility of *Pfs*-DF HisT could result in proteins such as the *E. coli* DHFS-FPGS to co-precipitate during isolation of the proteins. The actual determination of this would be difficult as the *E. coli* express very low levels of DHFS-FPGS. It seems from these results that the *Pfs*-DF HisT protein may possibly be toxic to the cells. Fully induced bacteria are said to behave strangely when trying to over-express recombinant proteins. They grow slower, produce higher amounts of heat shock proteins and may eventually stop growing altogether (Kurland C.G. and Dong H., 1996). The authors suggest that in a bacterial system that over-expresses recombinant protein or gratuitous protein it will limit the growth rate, because the overproduction of this component is associated with the underproduction of other native components (Kurland C.G. and Dong H., 1996). Thus if *Pfs*-DF HisT is toxic to the cells one would see a decrease in growth rate resulting in a lower optical density. Analysis of OD_{600nm} measurements of overnight auto-inducing cultures indicated a higher growth rate for *pPfs-HisT* in Star cells compared to the other constructs. Thus it seems that the *Pfs*-DF HisT is not toxic to the cells as the growth rate was not influenced. Another possible explanation could be more on the protein level with translational crashes being involved. An assumption of the protein-burden model is that there is a depression of normal protein production upon the increase of recombinant protein (Kurland C.G. and Dong H., 1996). There is thus a proportional reduction of the fraction of active protein in the soluble samples resulting in the observed lower specific activities. This de-repression may be due to a response of the translational system to the unequal competition for ribosomes by the recombinant protein and normal proteins exhibiting weaker or stronger affinities for ribosomes (Kurland C.G. and Dong H., 1996). It was also found that over-expression of proteins could lead to the degradation of rRNA's and intact ribosomes leading to a crash of the translational system (Baneyx F., 1999; Dong H. *et al.*, 1995). This could possibly

explain the observed results. Lower levels of active *E. coli* are thus produced resulting in lower specific activity with increased levels of *Pfs*-DF HisT in the expression system. Even though it's not clear why there is a decrease in specific activity the presence of the *E. coli* proteins in the total and soluble fractions have to be taken into account to make a conclusive assessment of the activity of *Pfs*-DF HisT. Due to the high level of *Pfs*-DF HisT found in the inclusion bodies, a refolding protocol was utilised to refold and solubilise *Pfs*-DF HisT (Sirawaraporn W. *et al.*, 1993). The refolded-soluble *Pfs*-DF HisT was used in activity determinations to determine if active protein was obtained.

3.3.2 Activity determinations of *Pfs*-DF HisT from different solubilisation methods

Activity determinations performed on purified, refolded soluble *Pfs*-DF HisT from the inclusion bodies indicated that the protein has a specific activity of ~0.84 pmol/h/ng and 0.77 pmol/h/ng for both the FPGS and DHFS activities, respectively. These activities were lower than the specific activity found for the total protein fraction obtained directly after protein expression. Previous experiments using the same protocol indicated the *P. falciparum* DHFR to be active (Sirawaraporn W. *et al.*, 1993), and the protein refolding protocol to be a good source of native protein. The solubilisation of a *P. falciparum* 42kDA merozoite surface protein 1 (MSP1$_{42}$) was also successfully obtained using a similar refolding protocol (Shimp Jr R.L. *et al.*, 2006). From the results of this study, it is apparent that the method could not be extrapolated to DHFS-FPGS as the purified refolded-soluble *Pfs*-DF HisT showed very little activity. A possible explanation for this could be non-optimal refolding resulting in inactive protein or a small percentage of active protein. The presence of the detergent guanidium chloride could be the contributing agent for loss of activity as high concentrations of chaotropic reagents results in the loss of secondary structure leading to the random coil formation of the protein (Singh S.M. and Panda A.K., 2005). The guanidium chloride acts by disrupting the hydrogen bonds and the hydrophobic interactions and results in complete unfolding of the protein. Therefore, residual reagent could interfere with the reassembly of the protein. The refolding of the protein could also result in a suboptimal fold not showing any activity. The loss of secondary structure during solubilisation is considered to be the main reasons for the loss of bioactive proteins from inclusion bodies (Singh S.M. and Panda A.K., 2005). Due to low yield of active protein obtained after refolding and solubilisation of *Pfs*-DF HisT, several other solubilisation methods were investigated to increase the specific activities.

The different inducing media showed the *Pfs*-DF HisT to be active in both the 4xYT and auto-inducing media taking into account the background *E. coli* levels (Refer to Table 3.2 and 3.3). Higher specific activities are observed for the soluble proteins from the 4xYT media. This might be due to the lower expression levels of *Pfs*-DF HisT in the 4xYT media (refer to section 2.3.6), resulting in a less negative effect on the recombinant expression of *E. coli* proteins (refer to section 3.2.2.1). The opposite effect was observed for the auto-inducing media that produces high levels of *Pfs*-DF HisT, resulting in much lower specific activities for the soluble proteins. Thus elevated levels of *Pfs*-DF HisT have a greater effect on the normal *E. coli* protein expression. Both media indicated the production of active protein, however it was decided to use the auto-inducing media as preferred culturing method due to its reproducibility and high levels of *Pfs*-DF HisT obtained (refer to section 2.3.6). The extraction and solubilisation of *Pfs*-DF HisT from inclusion bodies with ionic and non-ionic detergents was investigated to obtain pure active protein.

Initial solubilisation indicated that low levels of sarcosine are required to increase solubilised *Pfs*-DF HisT (refer to section 2.3.8). To determine which concentration of sarcosine would have the least effect on the *Pfs*-DF HisT activity, comparative activity assays of *Pfs*-DF HisT after extraction using both 0.5% and 0.75% sarcosine, were investigated. Fractions obtained during the solubilisation and purification of each sarcosine concentration was used in the determination of specific activity as well as fold purification. The specific activities obtained for solubilised *Pfs*-DF HisT after treatment with 0.5% and 0.75% sarcosine were compared. The results indicated a decrease in specific activity of solubilised *Pfs*-DF HisT with increasing sarcosine concentration (refer to Figure 3.10). Thus the 0.5% sarcosine concentration would be the optimum concentration to use for obtaining active *Pfs*-DF HisT.

Comparison of the purified *Pfs*-DF HisT to the solubilised *Pfs*-DF HisT (before purification) from the 0.5% sarcosine concentration showed an increase in specific activity for the purified protein (refer to Table 3.5 and Figure 3.11 B). The purified *Pfs*-DF HisT indicated a 147-fold (FPGS activity) and 114-fold (DHFS activity) difference in specific activity compared to the solubilised *Pfs*-DF HisT. These results thus indicate that pure active *Pfs*-DF HisT can be obtained from the detergent extraction using a 0.5% sarcosine concentration followed by Ni^{2+} affinity purification. Comparison of the solubilised *Pfs*-DF HisT from the glucopyranoside fraction to the soluble fraction (before solubilisation) indicated an unexpected 2.7-fold (FPGS activity) and 3.3-fold (DHFS activity) increase in specific activity (refer to Figure 3.11 A). Sarcosine is a much harsher detergent, interacting with the ionic and hydrogen bonds of proteins in solution, which could denature the protein

completely. Sarcosine has a low critical micelle concentration (CMC), which makes the micelle more stable and molecules such as proteins are slowly incorporated or removed from the micelle. The high micelle molecular weight (293.4) and low aggregation number (AN=2) of sarcosine favour its removal from protein complexes via dialysis due to its interference with assay conditions (www.sigmaaldrich.com, 2007). Due to sarcosine interfering with protein purification, activity determinations and possibly its denaturing effect on the protein, it would be beneficial to exclude the sarcosine from the detergent extraction protocol, thereby decreasing the possible loss of activity of *Pfs*-DF HisT. Thus a comparative study between the purified *Pfs*-DF HisT from the 0.5% sarcosine fraction and the glucopyranoside fraction need to be performed to determine if more active protein can be obtained from the glucopyranoside fraction. Importantly, optimisation experiments would be required to determine the optimal concentration of glucopyranoside for solubilisation of *Pfs*-DF HisT as 1% of glucopyranoside showed very little enrichment of *Pfs*-DF HisT (refer to Figure 2.20, Lane 4 and Figure 2.21, Lane 3). The effect of glucopyranoside on the Ni^{2+} affinity purification, as well as activity assay, would also be required to ensure pure active protein can be obtained.

A study on the effect of non-ionic detergents on solubilisation suggested that inhibition of enzyme activity was not related to solubilisation by these detergents (Guillon G. *et al.*, 1978). The authors suggested that two different actions of non-ionic solubilising detergent contribute to overall changes in activity. Enzymes treated with detergent undergo rapid and partially reversible changes with an increase in temperature. This suggests changes in the hydrophobic interactions between soluble protein and bound detergent molecules. Thus some enzyme catalytic activity might be highly dependent on the nature of the hydrophobic environment of the enzyme (Guillon G. *et al.*, 1978). Secondly, the inhibition of enzyme activity could probably be due to the micellar form of the detergent. Thus reduction of total detergent concentration below the critical micellar concentration increases enzyme activity (Guillon G. *et al.*, 1978). All of these factors thus need to be taken into account when working with detergents for the solubilisation of proteins. Thus the use of non-ionic detergents is still more beneficial than ionic detergents due to the non-denaturing effects of these former types of detergents on proteins.

3.3.3 Kinetic analysis

Initial kinetic analysis indicated a non-linear response in product formation with increasing purified *Pfs*-DF HisT concentrations. A possible reason could be the low concentrations of protein obtained after detergent extraction. Firstly, the enzyme might be unstable at low concentrations and secondly, the increments between the different concentrations are too small to reflect a significant increase in enzymatic activity. Due to low protein yields, higher concentrations of protein could not be utilised. The possible solution to this problem would be a larger scale extraction and solubilisation of *Pfs*-DF HisT or by concentrating the protein by means of spin columns. To ensure that the assay was functioning optimally, increasing volumes of total fractions (soluble + insoluble proteins) from Star cells (no vector) and *Pfs*-DF HisT (*pPfs-HisT* in Star cells) was used to observe an increase in enzyme activity. Results indicated a doubling in the pmoles of product produced as well as in activity (pmol/hour) with increasing volumes of protein from the Star cells only. The protein activities for the Star cells were much higher than the activity found for the synthetic gene construct proteins. Significantly, detectable enzyme activity of Star cells expressing the synthetic gene was only observed at 20 µl of total protein fraction and virtually no detectable activity at 10 µl. The undetectable activity at the lower volume of total protein fraction reinforces the suggestion that the protein is unstable at low protein concentrations. Stabilising factors possibly also play a role in the increased activity observed at the higher volume of total protein.

Even though active protein was obtained with the detergent extraction, the concentration of *Pfs*-DF HisT is still too low for the determination of kinetic properties. Thus the initial issue, which is the low soluble expression of DHFS-FPGS in the *E. coli* host, needs to be resolved for optimum activity determinations. Several strategies can be followed to improve soluble expression in addition to those investigated in this study. These include the use of solubility enhancing tags, codon-harmonisation, and cell-free expression systems (Endo Y. and Sawasaki T., 2006; Kincaid R.L. *et al.*, 2004; Mudeppa D.G. *et al.*, 2007; Shimizu Y. *et al.*, 2006; Terpe K., 2003; Waugh D.S., 2005).

Chapter 4:
Concluding Discussion

4. The "omics" of *Plasmodium falciparum*

In 1996, an international venture was launched to sequence the complex genome of *P. falciparum*. It was anticipated that these efforts would provide new insights into understanding this unique parasite's biology for the development of vaccines, drugs and diagnostics (Doolan D.L. *et al.*, 2003; Hoffman S.L *et al.*, 2002). Gene ontology analysis revealed that 1.3% of genes are involved in cell-to-cell adhesion and or the invasion of host cells, while 3.9% are implicated in the evasion of the host immune system (Gardner M.J *et al.*, 2002a). The *P. falciparum* genome appears to be under represented in biological processes such as cell cycle, cell organisation, and biogenesis, as well as transcription factors (Kooij T.W.A. *et al.*, 2006). About 60% of the predicted proteins showed no similarity to other proteins and were termed hypothetical proteins (Gardner M.J *et al.*, 2002b). These results suggest that a smaller percentage of the genome is committed to enzymes or that the enzymes are increasingly difficult to identify (Gardner M.J *et al.*, 2002a). The high A+T richness of the genome, the presence of parasite-specific inserts in proteins and the great evolutionary distance of the parasite from other organisms contributes to the difficulty in assigning a function to malarial proteins (Gardner M.J *et al.*, 2002a).

Since the publication of the malaria parasite genome, numerous "-omics" studies has been initiated to uncover the cellular roles and complex biology of the hypothetical proteins in *P. falciparum*. In 2003, Le Roch *et al.* determined the comparative level and temporal pattern of expression of more than 95% of the predicted genes as the parasite moves through the IDC (Le Roch K.G *et al.*, 2003). Expression profiling of the IDC showed that the majority of gene expression occurs during the IDC and that 75% of these genes are activated only once (Bozdech Z *et al.*, 2003). This cascade of transcriptional activation correlates with the need for the particular protein's biological function. Inhibition of key transcription factors of this activation provides a means of arrest of parasite development (Kooij T.W.A. *et al.*, 2006). This expression profiling behaviour can thus be exploited to prioritise novel antigens and epitopes that may be targets for protective immunity against the parasite (Doolan D.L. *et al.*, 2003).

Analysis of the proteomes illustrates that there is a distinctive and highly coordinated expression of the proteome for the asexual blood stages (Florens L *et al.*, 2002) and the sexual stages of the parasite (Lasonder E *et al.*, 2002). *P. falciparum* also clusters the co-regulated genes in its genome to specific chromosomes as is the case in other eukaryotes such as *Saccharomyces cerevisiae* and *H. sapiens* (Florens L *et al.*, 2002). These results were supported by the studies of Le Roch *et al.* who showed that genes with similar functions have related expression profiles. Both these studies indicated that genes and/or hypothetical proteins could be clustered together with known genes and known protein function. This allows researchers to assign preliminary functions to these hypothetical proteins and to characterise proteins as new targets in the search for vaccines and drug targets.

The proteins secreted by *P. falciparum* are of great interest as these proteins remodel the host cell for survival and cause the disease pathologies. Several recent studies identified a second functional amino-terminal signal sequence on the proteins exported into the erythrocyte (Hiller N.L. *et al.*, 2004; Marti M. *et al.*, 2004). This signal predicts a secretome of 300-400 proteins for *P. falciparum*, greatly increasing the number of potential vaccine and drug targets. The secretome has opened up new avenues for malaria research, namely understanding the molecular mechanism of protein export and together with new 'permeome' results could provide new targets for drug and vaccine research. The permeome is defined as the total complement of proteins involved in membrane permeability within a given organism (Martin R.E. *et al.*, 2005). New transport proteins were identified as possible nutrient transporters for sugars, amino acids, nucleosides, and vitamins (Martin R.E. *et al.*, 2005). Several transport proteins were also identified which is involved in the maintenance of ionic composition and metabolic waste removal. Six putative amino-acid transporters were also identified (Martin R.E. *et al.*, 2005). Comparison of the *P. falciparum* permeome to other organisms indicates that the parasite is minimalistic with regard to transporters (4.7 transport genes per Mb of DNA). This implies that the parasite might not have numerous transporters for a specific role, thus an inhibitor of a single transporter may prove to be an effective antimalarial drug as there will probably be no alternative transporter to utilise (Martin R.E. *et al.*, 2005).

As previously mentioned, there are a large number of *P. falciparum* proteins of unknown function. These proteins are very important as they could provide new targets for vaccine and drug target research. Date *et al.* wanted to deduce protein function on a genome wide scale using computational models. The interactome of *P. falciparum* was obtained from a

network of pair-wise functional linkages and reconstructed by integrating data from transcriptome profiling studies and *in silico* linkages. The interactome revealed that ~ 60% of proteins are currently annotated as hypothetical and that ~95% of these are linked to other known proteins. These linkages could suggest functional characteristics providing a means of elucidating protein function of the large number of hypothetical proteins present in *P. falciparum* (Date S.V. and Stoechert Jr., 2006).

Even though several new targets have been identified through the "-omics" of *P. falciparum* and advances in understanding the biology have been made, we are not closer to knowing how to eradicate this burden. The use of the "-omics" technologies are increasing our understanding of key differences in parasite-host interactions and are paving the way for a permanent solution (Hayes C.N. *et al.*, 2006). It seems that the words from Erwin Chargaff: "Science is wonderfully equipped to answer the question "How?" but it gets terribly confused when you ask the question "Why?", is as true for malaria as it is for science (Erwin Chargaff, Columbia Forum, Summer, 1969).

4.1 DHFS-FPGS as putative drug target

The only two activities currently being targeted in the folate pathway are DHFR and DHPS. This is problematic as emergence of resistance to antifolates targeting these enzymes is increasing. Other promising drug targets identified in the folate pathway are; cyclohydrolase (GTP-CH), serine hydroxymethyltransferase (SHMT), and the bifunctional dihydrofolate synthase/folylpolyglutamate synthase (DHFS-FPGS) (Lee C.S. *et al.*, 2001). DHFS-FPGS was identified to be a bifunctional protein involved in both the *de novo* (addition of L-glutamate moieties to dihydropteroic acid) and salvage pathways (extension of the glutamate chain) (Lee C.S. *et al.*, 2001). The importance of the FPGS activity is underlined by the fact that all organisms have FPGS activity to synthesise folylpolyglutamates. DHFS is only present in organisms that synthesise their own folates and much less is known about DHFS activity. Humans do not have a DHFS homologue, which makes this enzyme an appealing drug target. A single inhibitor directed against DHFS-FPGS will have a tremendous effect on blocking both routes of folate metabolism, which is essential for parasite growth. This inhibition would achieve the same results as observed with the synergy between the antifolates sulfadoxine (SDX) and pyrimethamine (PYR) inhibiting dihydropteroate synthase (DHPS) and dihydrofolate reductase (DHFR), respectively. The potential of DHFS-FPGS as a drug target needs to be assed by kinetic and structural studies, which require large amounts of active enzyme.

4.2 The aims of this study: obtaining active *P. falciparum* DHFS-FPGS

Due to the codon bias and high A+T-rich genome of the *P. falciparum* parasite, over-expression of parasite proteins in heterologous expression systems is problematic (Baca A.M. and Hol W.G.J., 2000). Codon-optimisation strategies can be utilised to produce sufficient amounts of recombinant proteins (Gustafsson C. *et al.*, 2004). A previously synthesised codon-optimised *dhfs-fpgs* gene (Coetzee L., 2003) was used in comparative expression studies and solubilisation experiments as described in Chapter 2. For the comparative expression studies, the full-length native *Pf. dhfs-fpgs* gene was isolated successfully. The isolated native *Pf. dhfs-fpgs* gene (~1573bp) was cloned and sequenced to verify identity. The expression levels of the native *Pf. dhfs-fpgs* gene and the synthetic *Pf. dhfs-fpgs* gene were compared in three different *E. coli* cell lines. The highest level of gene expression was obtain for the synthetic *Pf. dhfs-fpgs* gene in fast growing BL21 Star (DE3) cells, however insoluble protein was produced in inclusion bodies. Analysis indicated that the synthetic Tagless construct showed higher total protein levels than the synthetic Histag construct in the Star cells. *E. coli* DHFS-FPGS normally represents approximately 0.01% of the soluble proteins expressed (Bognar A.L. *et al.*, 1986), which might explain the insolubility observed for the synthetic gene product. Lower expression levels were observed for both the native and the synthetic *Pf. dhfs-fpgs* genes in the slower growing BL21 (DE3) pLysS and the BL21 Gold (DE3) pLysS. Nevertheless, the expression in these two cell lines yielded increased soluble levels. Both the native Tagless construct in BL21 Star cells and the native Histag construct in BL21 cells indicated the most soluble protein, however upon LC-MS/MS analysis of the soluble fractions, high levels of *E. coli* proteins were identified. Based on these results, the tagged synthetic *Pf. dhfs-fpgs* gene product was used in further solubilisation experiments to increase the yield of soluble DHFS-FPGS as it has an added advantage of one-step purification and an improved complementation of a *E.coli* K-12 mutant strain, SF4 (F$^-$ *folC strA recA tn10: srlC*) compared to the native construct.

Isolated DHFS-FPGS from solubilisation experiments was assayed as described in Chapter 3 to determine the activity and kinetic properties of the enzyme. Using the method of Sirawaraporn *et al.* (Sirawaraporn W. *et al.*, 1993), the refolding and solubilisation of DHFS-FPGS resulted in pure protein after purification however; activity determinations indicated that the protein has very little activity. The loss of activity is possibly due to the loss of secondary structure resulting in sub-optimal refolding of the protein (Singh S.M. and Panda A.K., 2005). An N-terminal truncated product of the DHFS-FPGS protein

(~57kDa) was identified after electrophoretic analysis. The ORF of DHFS-FPGS contains two in-frame ATG codons at Met-17 (51bp) and Met-41 (123bp). Complementation studies using truncated DHFS-FPGS at these positions indicated that the DHFS and FPGS activities are independent of the first 40 residues at the N-terminal (Salcedo E. *et al.*, 2001). Analysis of the ORF using the Sequence Manipulation Site (Stothard P., 2000) indicated the start codon at 51bp to be responsible for the predicted truncated protein. Possible explanations for the truncation would be hairpin loop formation resulting in downstream translation or incorrect pausing which causes translational slippage. Introducing the correct pause sites by means of codon-harmonisation would decrease the formation of hairpin loops as well as the incorrect pausing of the translational machinery.

The inclusion of chaperone proteins during protein expression resulted in a small increase of soluble protein but with lower total protein yield. The chaperones redirect proteins for folding thus slowing down both the expression and folding system which contributes to the lower expression level (Baneyx F. and Mujacic M., 2004). Due to the much lower yield of DHFS-FPGS with the chaperone system, alternative solubilisation methods were investigated. Improved expression levels are obtained when cells are grown in complex auto-inducing media than with IPTG induction due to the increased leaky expression (Grossman T.H. *et al.,* 1998; Studier F.W., 2005). Investigation of the protein expression levels of DHFS-FPGS in auto-inducing media complexes indicated increased protein levels. However, protein expression levels from the 6% YE and the 4xYT media were not reproducible with independent experiments. The use of auto-inducing media resulted in higher total protein levels but no increase in soluble protein (Studier F.W., 2005). Activity determinations on DHFS-FPGS from these media indicated the protein to be active.

The extraction and solubilisation of synthetic DHFS-FPGS from the inclusion bodies using non-ionic (n-Octyl-ß-glucopyranoside) and ionic (Lauroyl sarcosine) detergents were investigated as an alternative strategy (Angov E, Walter Reed Army Institute of Research, personal communication). The use of these detergents indicated that 0.5% sarcosine increases the solubilised fraction of the protein. A minimal amount of *E. coli* proteins does enrich during the refolding of the target protein but these levels were inconsequential due the subsequent purification of the protein using Ni^{2+}-affinity chromatography. A comparative activity assay of synthetic DHFS-FPGS after detergent extraction using both 0.5% and 0.75% sarcosine was investigated to determine which concentration would have the least effect on the activity determinations. The results indicated that with an increase in sarcosine concentration the activity of solubilised DHFS-FPGS decreases, thus the 0.5% sarcosine concentration was used in subsequent experiments. Purified DHFS-FPGS

indicated a 100-fold increase in activity compared to the solubilised DHFS-FPGS. Thus, the use of the auto-inducing media for high-level protein production in combination with the isolation and solubilisation of DHFS-FPGS by means of detergents resulted in pure, active protein. Interestingly the solubilised DHFS-FPGS from the glucopyranoside fraction to the soluble fraction (before solubilisation) indicated a ~3-fold increase in specific activity. Due to the denaturing effect of sarcosine on proteins, it would be worthwhile to investigate if the use of glucopyranoside as the only detergent for solubilisation could yield higher levels of active DHFS-FPGS.

Due to the *E. coli* background proteins, being present in the soluble samples as analysed by LC-MS/MS the effect of these proteins on the enzyme assay was also investigated. Results indicated that proteins (total and soluble fractions) isolated from empty Star cells, and Star cells containing an empty pET22b vector to have similar specific activities. However, the inclusion of *P. falciparum* DHFS-FPGS results in a decrease of specific activities from the total and soluble fractions. Initially it seemed that the DHFS-FPGS must be toxic to the cells, but optical density measurements indicated no effect on growth rates. Thus, the presence of *P. falciparum* DHFS-FPGS in the expression system must have some other detrimental effect on the *E. coli* proteins to result in the decrease of specific activity. These effects could include the co-precipitation of *E. coli* DHFS-FPGS as well as the decrease of the soluble protein levels.

Initial kinetic analysis involved the determination of the linearity of the assay with increasing enzyme concentrations. The purified *Pfs*-DF HisT obtained after treatment with 0.5% sarcosine and Ni^{2+} affinity chromatography was used for the kinetic analysis. Results indicated a non-linear response that was unexpected. Very low concentration of DHFS-FPGS was obtained after purification, which could possibly destabilise and cause the inactivity of the enzyme. The purification of DHFS-FPGS could also remove a stabilising factor needed for enzyme activity. The non-linear response is most probably due to the small increments between the enzyme concentrations resulting in no significant effect in product formation being observed. To ensure that results obtained was not due to assay conditions increasing volumes of total fractions from empty Star cells and Star cells with DHFS-FPGS were used to determine the linearity of the assay. Increasing volumes of proteins resulted in a doubling of product formation as well as activity (pmol/hour) for proteins form the Star cells. However, a significant enzyme activity for DHFS-FPGS could only be detected at the higher protein concentration at much lower activity levels than in the Star cells. These results strengthen the suggestion that the enzyme is unstable at low

protein concentrations as well as some stabilising factors playing a role during the activity determinations.

The primary aims of the thesis were achieved by the successful isolation of active DHFS-FPGS however; the isolated active DHFS-FPGS were in too low concentrations for the subsequent kinetic analysis. The enigma of the lower expression of the *E. coli* proteins in the presence of *P. falciparum* DHFS-FPGS has not been resolved.

4.3 Future perspective

Additional experiments that need to be performed include the activity determinations of DHFS-FPGS co-expressed with the chaperones as well as optimising the detergent extraction using only glucopyranoside. Initially it was decided to not proceed with the chaperone proteins as much lower levels of protein expression were obtained even though more soluble protein was produced. However if DHFS-FPGS could be purified in high enough levels from these soluble samples it could prove invaluable for the kinetic analysis. A comparative assay between purified DHFS-FPGS from the soluble sample and purified DHFS-FPGS from the detergent extraction would provide a comprehensible conclusion on the true effect of the detergent on the activity of DHFS-FPGS. The use of glucopyranoside as the only detergent during solubilisation would be beneficial as non-ionic detergents has little or no denaturing effect on proteins at/below CMC. However several optimisation steps are required to determine the correct amount of glucopyranoside that would yield high levels of active proteins without interfering with the assay or the Ni^{2+} affinity chromatography. Even though the above mentioned experiments would assist in the answering of important questions raised during the study, improved soluble expression of DHFS-FPGS would be more beneficial.

As mentioned, several strategies can be followed to improve the solubility of DHFS-FPGS such as: solubility enhancing tags, codon-harmonisation, and cell-free expression systems. Solubility-enhancing tags such as the maltose binding protein (MBP) and the N-utilisation substance A (NusA) have the ability to promote the solubility of their fusion partners (Waugh D.S., 2005). The MBP protein is also naturally an affinity tag and could thus be utilised as both (Terpe K., 2003; Waugh D.S., 2005). Another alternative is the use of cell-free protein synthesis, which is based on the cellular ribosomal protein synthesis system (Shimizu Y. *et al.*, 2006). This system is composed of cell extract from either *E. coli*, wheat germ or rabbit reticulocytes, containing all the components for proteins synthesis such as ribosomes, translation factors, aminoacyl-tRNA and tRNAs.

Proteins are produced by addition of template DNA or mRNA to the mixture, followed by incubation for several hours (Endo Y. and Sawasaki T., 2006; Shimizu Y. *et al.*, 2006). A wheat germ cell free system was shown to have little preference for codon usage and was able to express A+T-rich cDNAs thus providing a means for soluble expression of *Plasmodium* proteins (Endo Y. and Sawasaki T., 2006). The wheat germ *in vitro* translation system was successfully used to express functional *P. falciparum* DHFR-TS (Mudeppa D.G. *et al.*, 2007). Higher expression levels of DHFR-TS were obtained with this system compared to codon-optimised expression (Mudeppa D.G. *et al.*, 2007).

The recent introduction of codon-harmonisation instead of codon-optimisation could also provide more soluble expression of DHFS-FPGS. This method as mentioned in Chapter 2 takes into account not only the codon bias, but also the frequency of codons on the sequence (Kincaid R.L. *et al.*, 2002). Initial analysis of the *dhfs-fpgs* gene indicated that harmonisation of the first few codons (~150bp) at the 5' end would increase the soluble expression of DHFS-FPGS (Appendix A) (de Ridder J., Angov E., Personal communication). The harmonisation of the first 195bp of *dhfs-fpgs* gene is currently being investigated as a means to improve the expression of this enzyme. Even though this might be the solution, it seems that not only the high A+T-richness of the *Plasmodium* genome needs to be taken into account for optimal protein expression. Two independent studies performed by Vedadi *et al.* and Mehlin *et al.* suggested that there is a weak association between the %A+T content and protein expression (Mehlin C. *et al.*, 2006; Vedadi M. *et al.*, 2007). The main factors identified to influence expression and solubility were categorised to be protein characteristics (molecular weight, pI and predicted disorder) and gene composition (stretches of A and T, number of introns and %A+T) (Mehlin C. *et al.*, 2006). Analysis of these factors showed that pI is associated with both expression and solubility. The authors also indicated that protein size plays an important role in expression as larger proteins have more opportunities to aggregate (Mehlin C. *et al.*, 2006). However, it was suggested that the effect of %A+T is more likely due to its confounding relationship with pI, size and protein disorder (Mehlin C. *et al.*, 2006; Vedadi M. *et al.*, 2007). Both studies found a high correlation between protein disorder and soluble expression (Mehlin C. *et al.*, 2006; Vedadi M. *et al.*, 2007). Disordered proteins are characterised by having relatively low complexity in their sequences. Low complexity means a bias in favour of only a few amino acids with a larger degree of repetition (Lovell S.C., 2003). *Plasmodium* genomes feature low complexity regions and parasite specific inserts. These regions maintain a A+T-rich base composition as well as recurring $(AT)_n$ tracts (Pizzi E. and Frontali C., 2000). Low complexity segments contain mostly asparagine and lysine, and are proposed to form non-globular domains looping out of the

protein (Pizzi E. and Frontali C., 2001). Disordered loops are directly problematic for heterologous expression of soluble proteins (Vedadi M. *et al.*, 2007). Hopefully the use of codon-harmonisation would address the problem of low complexity by means of pause sites. The introduction of pause sites would allow the translational machinery to correctly fold the protein of interest.

Large quantities of correctly folded active DHFS-FPGS could be used for X-ray crystallography and NMR analysis for the determination of the three dimensional structure. Three-dimensional structures enhance the ability to design novel inhibitors and modifying existing drugs. The design of novel inhibitors directed against the folate metabolism would aid the eradication of the malaria burden. These inhibitors could also be used synergistically with current antifolates to decrease the rate of resistance development of the parasite to antimalarials in use.

Summary

Malaria is a life-threatening parasitic disease that causes at least 300 million acute illnesses annually, of which at least one million infected humans die, mainly children under the age of 5 years. This overwhelming burden is due to the most virulent causative agent, *Plasmodium falciparum*, as a result of its prevalence in sub-Saharan Africa, as well as its resistance to nearly all anti-malarials in use. There is thus an urgent need to discover and characterise new novel parasitic targets for chemotherapeutic intervention.

Folate metabolism is the target of several anti-malarials such as pyrimethamine and sulfadoxine. These drugs cause a decrease in parasite growth since *Plasmodia* have a high rate of replication and demand for nucleotides during DNA synthesis. The parasite is almost totally resistant to the current antifolates. Further insights into the folate pathway and its drugs are essential for the understanding of the resistance mechanism and to identify/characterise new drug targets for inhibition. Three possible new drug targets were identified and characterised in the folate pathway (Lee C.S. *et al.*, 2001). One of these targets is the bifunctional enzyme, dihydrofolate-synthase folylpolyglutamate synthase (DHFS-FPGS).

The bifunctionality and activity of the *dhfs-fpgs* gene in *Plasmodium* was confirmed by functional complementation in yeast and bacteria and shown to be unique to *Plasmodia* and bacteria. This gene indicated only a 15-17% homology to other organisms; DHFS activity is usually only present in prokaryotes but not in humans or other eukaryotes (Salcedo E. *et al.*, 2001). Although part of a bifunctional protein and having closely related catalytic functions, the DHFS and FPGS activities have distinct roles to play in both the *de novo* and salvage pathways of folate metabolism. These characteristics indicate DHFS-FPGS as a potentially good drug target since a single inhibitor is likely to have a drastic effect on both routes and consequently arrest DNA synthesis in the malaria parasite. This could prove to be a very effective and novel antimalarial strategy.

Comparative expression studies of synthetic and native DHFS-FPGS presented here indicated that the highest quantity of protein is expressed from the synthetic gene. However, results indicated that most of the recombinant protein expressed in various *E.*

coli cell lines, produced insoluble protein aggregates. Various strategies were employed in an attempt to improve recombinant soluble expression including auto-induction of T7 promoter activity. However, this did not result in an increased percentage of soluble protein expression even though improved total protein expression was observed. The inclusion of chaperone proteins resulted in a minor change in soluble expression. Activity assays of the DHFS-FPGS from these two methods indicated that active protein was produced in a correctly folded manner. Due to the high amount of recombinant protein present in the inclusion bodies, various methods were investigated to isolate and refold the DHFS-FPGS protein. The use of a non-ionic and ionic detergent for refolding resulted in pure, solubilised, active protein. Activity assays of the refolded, soluble protein indicated that the protein is active. Preliminary kinetic analysis was unsuccessful and requires further investigations.

References

Abbot A. (2001) **Earliest malaria DNA found in Roman baby graveyard**. *Nature*, 412, 847.

Anonymous. (2002) **Malaria after the genomes**. *The Lancet*, 360, 1107.

Arav-Boger R. and Shapiro T.A. (2005) **Molecular mechanisms of resistance to antimalarial chemotherapy: The unmet challenge**. *Annual Review of Pharmacology and Toxicology*, 45, 565-585.

Baca A.M. and Hol W.G.J. (2000) **Overcoming codon bias: A method for high-level over expression of *Plasmodium* and other AT-rich parasite genes in *Escherichia coli***. *International Journal for Parasitology*, 30, 113-118.

Balint G.A. (2001) **Artemisinin and its derivatives. An important new class of antimalarial agents**. *Pharmacology and Therapeutics*, 90, 261-265.

Baneyx F. (1999) **Recombinant protein expression in *Escherichia coli***. *Current Opinion in Biotechnology*, 10, 411-421.

Baneyx F. and Mujacic M. (2004) **Recombinant protein folding and misfolding in *Escherichia coli***. *Nature Biotechnology*, 22, 1399-1408.

Bannister L.H., Hopkins J.M., Fowler R.E., Krishna S. and Mitchell G.H. (2000) **A Brief Illustrated Guide to the Ultrastructure of *Plasmodium falciparum* asexual blood stages**. *Parasitology Today*, 16, 427-433.

Banumathy G., Singh V., Pavithra S.R. and Tatu U. (2003) **Heat shock protein 90 function is essential for *Plasmodium falciparum* growth in human erythrocytes**. *The Journal of Biological Chemistry*, 278, 18336-18345.

Baton L.A. and Ranford-Cartwright L.C. (2005) **Spreading the seeds of million-murdering death: metamorphoses of malaria in the mosquito**. *Trends in Parasitology*, 21, 573-579.

Baum J., Maier A.G., Good R.T., Simpson K.M. and Cowman A.F. (2005) **Invasion by *P. falciparum* merozoites suggests a hierarchy of molecular interactions**. *PLOS Pathogens*, 1, 299-309.

Belli S.I., Walker R.A. and Flowers S.A. (2005) **Global protein expression analysis in apicomplexan parasites: Current status**. *Proteomics*, 5, 918-924.

Biagini G.A., O'Neill P.M., Nzila A., Ward S.A. and Bray P.G. (2003) **Antimalarial chemotherapy: young guns or back to the future?** *Trends in Parasitology*, 19, 479-487.

Bloland P.B. (2001) **Drug resistance in malaria**. *World Health Organization*, 1-27.

Bognar A.L., Osborne C., Shane B., Singer S.C. and Ferone R. (1986) **Folyl-poly-γ-glutamate synthetase -dihydrofolate synthetase-cloning and high expression of the *Escherichia coli folC* gene and purification and properties of the gene product.** *Methods in Enzymology*, 122, 349-359.

Bollheimer L.C., Buettner R., Kullmann A. and Kullmann F. (2005) **Folate and its preventive potential in colorectal carcinogenesis. How strong is the biological and epidemiological evidence**. *Critical Reviews in Oncology/Hematology*, 55, 13-36.

Bozdech Z, Llinas M, Pulliam B.L, Wong E.D, Zhu J and DeRisi J.L. (2003) **The Transcriptome of the Intraerythrocytic developmental Cycle of plasmodium falciparum**. *PLoS Biology*, 1, 085-100.

Bradford M.M. (1976) **A rapid and sensitive method for the Quantitation for Microgram Quantities of protein utilizing the principle protein-dye binding.** *Analytical Biochemistry*, 72, 248-254.

Carrio M.M. and Villaverde A. (2001) **Protein aggregation as bacterial inclusion bodies is reversible**. *FEBS Letters*, 489, 29-33.

Carrio M.M. and Villaverde A. (2002) **Construction and deconstruction of bacterial inclusion bodies.** *Journal of Biotechnology*, 96, 3-12.

Caspers P., Stieger M. and Burn P. (1994) **Overproduction of bacterial chaperones improves the solubility of recombinant protein tyrosine kinases in *Escherichia coli*.** *Cellular Molecular Biology*, 40, 635-644.

Chen L., Qi H., Korenberg J., Garrow T.A., Choi Y.J. and Shane B. (1996) **Purification and properties of human cytosolic folylpoly-γ-glutamate synthetase and organization, localization, and differential splicing of its gene**. *The Journal of Biological Chemistry*, 271, 13077-13087.

Cinquin O., Christopherson R.I. and Menz R.I. (2001) **A hybrid plasmid for expression of toxic malarial proteins in *Escherichia coli***. *Molecular and Biochemical Parasitology*, 117, 245-247.

Clark I.A., Budd A.C., Alleva L.M. and Cowden W.B. (2006) **Human malarial disease: a consequence of inflammatory cytokine release.** *Malaria Journal*, 5, 1-32.

Coetzee L. (2003) *Plasmodium falciparum* dihydrofolate synthase-folylpolyglutamate synthase (DHFS-FPGS): Gene synthesis and recombinant expression. *Department of Biochemistry*. University of Pretoria, Pretoria, Vol. Magister Scientiae, p. 119.

Cossins E.A. (2000) **The fascinating world of folate and one-carbon metabolism**. *Canadian Journal of Botany*, 78.

Costi M.P. and Ferrari S. (2001) **Update on antifolate drugs targets**. *Current Drug Targets*, 2, 135-166.

Cowman A.F. (2001) **Functional analysis of drug resistance in *Plasmodium falciparum* in the post-genomic era**. *International Journal for Parasitology*, 31, 871-878.

Cowman A.F. and Crabb B.S. (2006) **Invasion of red blood cells by malaria parasites**. *Cell*, 124, 755-766.

Dale G.E., Broger C., Langen H., D'Arcy A. and Stuber D. (1994) **Improving protein solubility through rationally designed amino acid replacements: solubilization of the trimethoprim-resistant type S1 dihydrofolate reductase.** *Protein Engineering*, 7, 933-939.

Darko C.A., Angov E., Collins W.E., Bergmann-Leitner E.S., Girouard A.S., Hitt S.L., McBride J.S., Diggs C.L., Holder A.A., Long C.A., Barnwell J.W. and Lyon J.A. (2005) **The Clinical-Grade 42-Kilodalton Fragment of Merozoite Surface Protein 1 of *Plasmodium falciparum* strain FVO expressed in *Escherichia coli* protects *Aoutus nancymai***

against challenge with homologous erythrocytic-stage parasites. *Infection and Immunity*, 73, 287-297.

Date S.V. and Stoechert Jr., C.J. (2006) **Computational modeling of the *Plasmodium falciparum* interactome reveals protein function on a genome-wide scale**. *Genome Research*, 16, 542-549.

De Beer T.A.P., Louw A.I and Joubert F. (2006) **Elucidation of sulfadoxine resistance with structural models of the bifunctional *Plasmodium falciparum* dihydropterin pyrophosphokinase-dihydropteroate synthase.** *Biorganic and Medicinal Chemistry*, 14, 4433-4443.

Dong H., Nilsson L. and Kurland C.G. (1995) **Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction.** *Journal of Bacteriology*, 177, 1497-1504.

Doolan D.L., Aguiar J.C., Weiss W.R., Sette A., Felgner P.L., Regis D.P., Quinones-Casas P., Yates J.R., Blair P.L., Richie T.L., Hoffman S.L. and Carucci D.J. (2003) **Utilization of genomic sequence information to develop malaria vaccines**. *The Journal of Experimental Biology*, 206, 3789-3802.

Duraisingh M.T. and Cowman A.F. (2005) **Contribution of the *pfmdr1* gene to antimalarial drug-resistance**. *Acta Tropica*, 94, 181-190.

Endo Y. and Sawasaki T. (2006) **Cell-free expression systems for eukaryotic protein production**. *Current Opinion in Biotechnology*, 17, 373-380.

Fichera M.E. and Roos D.S. (1997) **A plastid organelle as a drug target in apicomplexan parasites**. *Nature*, 390, 407-409.

Flick K., Ahuja S., Chene A., Bejarano M.T. and Chen Q. (2004) **Optimized expression of *Plasmodium falciparum* erythrocyte membrane protein I domains in *Escherichia coli.*** *Malaria Journal*, 3, 1-8.

Florens L, Washburn M.P, Rain J.D, Anthony R.M, Grainger M, Haynes J.D, Moch J.K, Muster N, Sacci J.B, Tabb D.L, Witney A.A, Wolters D, Wu Y, Gardner M.J, Holder A.A, Sinden R.E, Yates JR and Carucci D.J. (2002) **A proteomic view of the *Plasmodium falciparum* life cycle**. *Nature*, 419, 520-526.

Francis S.E., Sullivan Jr. D.J. and Goldberg D.E. (1997) **Hemoglobin metabolism in the malaria parasite *Plasmodium falciparum***. *Annual Review Microbiology*, 51, 97-123.

Gardner M.J, Hall N, Fung E, White O, Berriman M, Hyman R.W, Carlton J.M, Pain A, Nelson K.E, Bowman S, Paulsen I.T, James K, Eisen J.A, Rutherford K, Salzberg S.L, Craig A, Kyes S, Chan M-S, Nene V, Shallom S.J, Suh B, Peterson J, Angiuoli S, Pertea M, Allen J, Selengut J, Haft D, Mather M.W, Vaidya A.B, Martin D.M.A, Fairlamb A.H, Fraunholz M.J, Roos D.S, Ralph S.A, McFadden G.I, Cummings L.M, Subramanian G.M, Mungall C, Venter J.C, Carucci D.J, Hoffman S.L, Newbold C, Davis R.W, Fraser C.M and B., B. (2002a) **Genome sequence of the human malaria parasite *Plasmodium falciparum***. *Nature*, 419, 498-511.

Gornicki P. (2003) **Apicoplast fatty acid biosynthesis as a target for medical intervention in apicomplexan parasites.** *International Journal for Parasitology*, 33, 885-896.

Greenwood B.M., Bojang K., Whitty C.J.M. and Targett G.A.T. (2005) **Malaria**. *Lancet*, 365, 1487-1498.

Gregson A. and Plowe C.V. (2005) **Mechanisms of resistance of malaria parasites to antifolates**. *Pharmacological Reviews*, 57, 117-145.

Grossman T.H., Kawasaki E.S., Punreddy S.R. and Osborne M.S. (1998) **Spontaneous cAMP-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability.** *Gene*, 209, 95-103.

Guillet P., Guessan R.N., Darriet F., Traore-Lamizana M., Chandre F. and Carnevale P. (2001) **Combined pyrethroid and carbamate 'two-in-one' treated mosquito nets: field efficacy against pyrethroid-resistant *Anopheles gambiae* and *Culez quinquefasciatus*.** *Medical and Veterinary Entomology*, 15, 105-112.

Guillon G., Roy C. and Jard S. (1978) **A systematic study of effects on non-ionic detergents on solubilisation and activity of pig kidney Adenylate Cyclase.** *European Journal of Biochemistry*, 92, 341-348.

Gustafsson C., Govindarajan S. and Minshull J. (2004) **Codon bias and heterologous protein expression.** *Trends in Biotechnology*, 22, 346-353.

Hanahan D., Jessee J. and Bloom F.R. (1991) **Plasmid Transformation of *Escherichia coli* and other bacteria.** *Methods in Enzymology*, 204, 63-113.

Hayes C.N., Wheelock A.M., Normark J., Wahlgren M., Goto S. and Wheelock C.E. (2006) **Enlistment of omics technologies in the fight against malaria: Panacea or Pandora's Box?** *Journal of Pesticide Science*, 31, 263-272.

Haynes R.K and Krishna S. (2004) **Artimisinins: activities and actions.** *Microbes and Infection*, 2004, 1339-1346.

Heddini A. (2002) **Malaria pathogenesis: a jigsaw with an increasing number of pieces.** *International Journal of Parasitology*, 32, 1587-1598.

Hemingway J., Beaty B.J., Rowland M., Scott T.W. and Sharp B.L. (2006) **The Innovative Vector Control Consortium: improved control of mosquito-borne diseases.** *Trends in Parasitology*, 22, 309-312.

Hiller N.L., Bhattacharjee S., van Ooij C., Liolios K., Harrison T., Lopez-Estrano C. and Haldar K. (2004) **A Host-targeting signal in virulence proteins reveals a secretome in malarial infection.** *Science*, 306, 1934-1937.

Hillier C., J., Ware L., A., Barbosa A., Angov E., Lyon J., A., Heppner D.G. and Lanar D.E. (2005) **Process development and analysis of liver-stage antigen 1, a preerythrocyte-stage protein-based vaccine for *Plasmodium falciparum*.** *Infection and Immunity*, 73, 2109-2115.

Hoffman S.L, G.M, S., Collins F.H and Venter J.C. (2002) ***Plasmodium*, human and *Anopheles* genomics and malaria.** *Nature*, 415, 702-709.

Holmes D.S. and Quigley M. (1981) **A rapid boiling method for the preparation of bacterial plasmids.** *Analytical Biochemistry*, 114, 193-197.

http://sites.huji.ac.il.

Hyde J.E. (2005) **Exploring the folate pathway in *Plasmodium falciparum*.** *Acta Tropica*, 94, 191-206.

Jana S. and Deb J.K. (2005) **Strategies for efficient production of heterologous protein in *Escherichia coli*.** *Applied Microbiology and Biotechnology*, 67, 289-298.

Joet T. and Krishna S. (2004) **The hexose transporter of *Plasmodium falciparum* is a worthy drug target.** *Acta Tropica*, 89, 371-374.

Johnson D.J., Fidock D.A., Mungthi M., Lakshmanan V., Sidhu A.B.S., Bray P.G. and Ward S.A. (2004) **Evidence for a central role for PfCRT in conferring *Plasmodium falciparum* resistance to diverse antimalarial agents**. *Molecular Cell*, 15, 867-877.

Kallen R.G. and Jencks W.P. (1966) **The dissociation constants of tetrahydrofolic acid**. *The Journal of Biological Chemistry*, 241, 5845-5850.

Kanzok S.M. and Zheng L. (2003) **The mosquito genome-a turning point?** *Trends in Parasitology*, 19, 329-331.

Karcher S.J. (1995) *In: Molecular Biology: A Project Approach*. San Diego Academic Press Inc.

Keenan S.M. and Welsh W.J. (2004) **Characteristics of the *Plasmodium falciparum* PK5 ATP-binding site: implications for the design of novel antimalarial agents**. *Journal of Molecular Graphics and Modelling*, 22, 241-247.

Kincaid R.L., Angov E. and Lyon J.A. (2002) Protein expression by codon harmonization and translational attenuation. In United States. (ed.). Veritas, San Francisco, CA,US, p. 37.

Kincaid R.L., Angov E. and Lyon J.A. (2004) Protein expression by codon harmonization and translational attenuation. In United States. (ed.). Veritas, San Francisco, CA,US., p. 37.

Kirk K. (2004) **Channels and transporters as drug targets in the *Plasmodium*-infected erythrocyte**. *Acta Tropica*, 89, 285-298.

Kok R.M., Smith D.E.C., Dainty J.R., van den Akker J.T., Finglas P.M., Smulders Y.M. and de Meer K. (2004) **5-Methyltetrahydrofolic acid and folic acid measured in plasma with liquid chromatography tandem mass spectrometry: applications to folate absorption and metabolism**. *Analytical Biochemistry*, 326, 129-138.

Kooij T.W.A., Janse C.J. and Waters A.P. (2006) ***Plasmodium* pos-genomics: better the bug you know?** *Nature Reviews Microbiology*.

Krishna S., Woodrow C.J., Staines H.M., Haynes R.K. and Puijalon O.M. (2006) **Re-evaluation of how artemisinins work in light of emerging evidence of *in vitro* resistance**. *Trends in Molecular Medicine*, 12, 200-205.

Krungkrai J., Webster H.K. and Yuthavong Y. (1989) ***De novo* and salvage biosynthesis of pteroylpentaglutamates in the human malaria parasite, *Plasmodium falciparum***. *Molecular and Biochemical Parasitology*, 32, 25-38.

Kurland C.G. and Dong H. (1996) **Bacterial growth inhibition by overproduction of protein**. *Molecular Microbiology*, 21, 1-4.

Lasonder E, Ishihama Y, Andersen J.S, Vermunt A.M.W, Pain A, Sauerwein R.W, Eling W.M.C, Hall N, Waters A.P, Stunnenberg H.G and Mann M. (2002) **Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry**. *Nature*, 419, 537-542.

Le Roch K.G, Zhou Y, Blair P.L, Grainger M, Moch J.K, Haynes J.D, De La Vega P, Holder A.A, Batalov S, D.J, C. and Winzeler E.A. (2003) **Discovery of gene function by expression profiling of the malaria parasite life cycle**. *Science*, 301, 1503-1508.

Lee C.S., Salcedo E., Wang Q., Wang P., Sims P.F.G. and Hyde J.E. (2001) **Characterization of three genes encoding enzymes of the folate biosynthesis pathway in *Plasmodium falciparum***. *Parasitology*, 122, 1-13.

Liu J., Istvan E.S., Gluzman I.Y., Gross J. and Goldberg D.E. (2006) ***Plasmodium falciparum* ensures its amino acid supply with multiple acquisition pathways and redundant proteolytic enzyme systems**. *Proceedings of the National Academy of Science*, 103, 8840-8845.

Lovell S.C. (2003) **Are non-functional, unfolded proteins ('junk proteins') common in the genome?** *FEBS Letters*, 554, 237-239.

Lowry O.H., Rosebrough N.J., Farr A.L. and Randall R.J. (1951) **Protein measurement with the Folin phenol reagent**. *The Journal of Biological Chemistry*, 193, 265-275.

Marrelli M.T., Moreira L.A., Kelly D., Alphey L. and Jacobs-Lorena M. (2006) **Mosquito transgenesis: what is the fitness cost?** *Trends in Parasitology*, 22, 197-202.

Marti M., Baum J., Rug M., Tilley L. and Cowman A.F. (2005) **Signal-mediated export of proteins from the malaria parasite to the host erythrocyte**. *The Journal of Cell Biology*, 171, 587-592.

Marti M., Good R.T., Knuepfer E. and Cowman A.F. (2004) **Targeting malaria virulence and remodeling proteins to the host erythrocyte**. *Science*, 306, 1930-1934.

Martin R.E., Henry R.I., Abbey J.L., Clements J.D. and Kirk K. (2005) **The 'permeome' of the malaria parasite: an overview of the membrane transport proteins of *Plasmodium falciparum***. *Genome Biology*, 6, 1-22.

Martin R.E. and Kirk K. (2007) **Transport of the essential nutrient isoleucine in human erythrocytes infected with the malaria parasite *Plasmodium falciparum*.** *Blood*, 109, 2217-2224.

Mather M.W., Henry K.W. and Vaidya A.B. (2007) **Mitochondrial drug targets in Apicomplexan parasites.** *Current Drug Targets*, 8, 49-60.

Mathieu M., Debousker G., Vincent S., Viviani F., Bamas-Jacques N. and Mikol V. (2005) ***Escherichia coli* FolC structure reveals an unexpected dihydrofolate binding site providing an attractive target for anti-microbial therapy**. *The Journal of Biological Chemistry*, 280, 18916-18922.

Matuschewski K. (2006) **Getting infectious: formation and maturation of *Plasmodium* sporozoites in the *Anopheles* vector**. *Cellular Microbiology*, 8, 1547-1556.

McDonald D., Atkinson I.J., Cossins E.A. and Shane B. (1995) **Isolation of dihydrofolate and folylpolyglutamate synthetase activities from *Neurospora***. *Phytochemistry*, 38, 327-333.

McGuire J.J. and Bertino J.R. (1981) **Enzymatic synthesis and function of folylpolyglutamates**. *Molecular and Cellular Biochemistry*, 38, 19-48.

McGuire J.J., Hsieh P., Coward J.K. and Bertino J.R. (1980) **Enzymatic synthesis of folylpolyglutamates. Characterization of the reaction and its products.** *The Journal of Biological Chemistry*, 255, 5776-5788.

Medicines for Malaria Venture. (2005) Curing malaria Together.

Mehlin C., Boni E., Buckner F.S., Engel L., Feist T., Gelb M.H., Haji L., Kim D., Liu C., Mueller N., Myler P.J., Reddy J.T., Sampson J.N., Subramanian E., Van Voorhis W.C., Worthey E., Zucker F. and Hol W.G.J. (2006) **Heterologous expression of proteins from *Plasmodium falciparum*: Results from 1000 genes**. *Molecular and Biochemical Parasitology*, 148, 144-160.

Messori L., Gabbiani c., Casini A., Siragusa M., Vincieri F.F. and Bilia A.R. (2006) **The reaction of artemisinins with hemoglobin: A unified picture**. *Biorganic and Medicinal Chemistry*, 14, 2972-2977.

Miller L.H., Baruch D.I. and Doumbo O.K. (2002) **The pathogenic basis of malaria**. *Nature*, 415, 673-679.

Moran R.G. (1999) **Roles of Folylpoly-gamma-glutamate synthetase in therapeutics with tetrahydrofolate antimetabolites: An overview.** *Seminars in Oncology*, 26, 24-32.

Moreira L.A., Ghosh A.K., Abraham E.G. and Jacobs-Lorena M. (2002) **Genetic transformation of mosquitoes: a quest for malaria control**. *International Journal for Parasitology*, 32, 1599-1605.

Mudeppa D.G., Pang C.K.T., Tsuboi T., Endo Y., Buckner F.S., Varani G. and Rathod P.K. (2007) **Cell-free production of functional *Plasmodium falciparum* dihydrofolate reductase-thymidylate synthase**. *Molecular and Biochemical Parasitology*, 151, 216-219.

Narum D.L., Kumar S., Rogers W.O., Fuhrmann S.R., Liang H., Oakley M., Taye A., Sim B.K.L. and Hoffman S.L. (2001) **Codon optimization of gene fragments encoding *Plasmodium falciparum* merozoite proteins enhances DNA vaccine protein expression and immunogenicity**. *Infection and Immunity*, 69, 7250-7253.

Neuhoff V., Nobert A., Taube D. and Ehrhardt W. (1988) **Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250**. *Electrophoresis*, 9, 255-262.

Nirmalan N, Sims P.F.G and Hyde J.E. (2004) **Quantitative proteomics of the human malaria parasite *Plasmodium falciparum* and its application to studies of development and inhibition**. *Molecular and Microbiology*, 52, 1187-1199.

Nishihara K., Kanemori M., Kitagawa M., Yanagi H. and Yura T. (1998) **Chaperone coexpression plasmids: Differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen,Cryj2, in *Escherichia coli***. *Applied and Environmental Microbiology*, 64, 1694-1699.

Nzila A. (2006a) **Inhibitors of *de novo* folate enzymes in *Plasmodium falciparum***. *Drug Discovery Today*, 11, 939-944.

Nzila A. (2006b) **The past, present and future of antifolates in the treatment of *Plasmodium falciparum* infection**. *Journal of Antimicrobial Chemotherapy*, 57, 1043-1054.

Nzila A., Ward S.A., Marsh K., Sims P.F.G. and Hyde J.E. (2005a) **Comparative folate metabolism in humans and malaria parasites (partI):pointers for malaria treatment from cancer chemotherapy.** *Trends in Parasitology*, 21, 292-298.

Nzila A., Ward S.A., Marsh K., Sims P.F.G. and Hyde J.E. (2005b) **Comparative folate metabolism in humans and malaria parasites (partII): activities as yet untargeted or specific to *Plasmodium.*** *Trends in Parasitology*, 21, 334-339.

O'Donnel R.A., Hackett F., Howell S.A., Treeck M., Struck N., Krnajski Z., Withers-Martinez C., Gilberger T.W. and Blackman M.J. (2006) **Intramembrane proteolysis mediates shedding of a key adhesion during erythrocyte invasion by the malaria parasite**. *The Journal of Cell Biology*, 174, 1023-1033.

Ouelette M., Drummelsmith J., Fadili A.E., Kundig C., Richard D. and Roy G. (2002) **Pterin transport and metabolism in *Leishmania* and related trypanosomatid parasites**. *International Journal for Parasitology*, 32, 385-398.

Painter H.J., Morrisey J.M., Mather M.W. and Vaidya A.B. (2007) **Specific role of mictochondrial electron transport in blood-stage *Plasmodium falciparum***. *Nature*, 446, 88-91.

Pasvol G. (2001) **Cell-cell interaction in the pathogenesis of severe *falciparum* malaria**. *Clinical Medicine*, 1, 495-500.

Perkins D.N., Pappin D.J., Creasy D.M. and Cottrell J.S. (1999) **Probability-based protein identification by searching sequence databases using mass spectrometry data**. *Electrophoresis*, 20, 3551-3567.

Pizzi E. and Frontali C. (2000) **Divergence of noncoding sequences and of insertions encoding nonglobular domains at genomic region well conserved in *Plasmodia***. *Journal of Molecular Evolution*, 50, 474-480.

Pizzi E. and Frontali C. (2001) **Low-complexity regions in *Plasmodium falciparum* proteins**. *Genome Research*, 11, 218-229.

Prapunwattana P., Sirawaraporn R., Yuthavong Y. and Santi DV. (1996) **Chemical synthesis of the *Plasmodium falciparum* dihydrofolate reductase-thymidylate synthase gene**. *Molecular and Biochemical Parasitology*, 83, 93-106.

Ramasastri B.V. and Blakley R.L. (1964) **5,10-Methylenetetrahydrofolic dehydrogenase from bakers' yeast**. *The Journal of Biological Chemistry*, 239, 106-111.

Rasti N., Wahlgren M. and Chen Q. (2004) **Molecular aspects of malaria pathogenesis**. *FEMS Immunology and Medical Microbiology*, 41, 9-26.

Richie T.L. and Saul A. (2002) **Progress and challenges for malaria vaccines**. *Nature*, 415, 694-701.

Ridley R.G. (2002) **Medical need, scientific opportunity and the drive for antimalarial drugs**. *Nature*, 415, 686-693.

Rieckmann K.H. (2006) **The chequered history of malaria control: are new and better tools the ultimate answer?** *Annals of Tropical Medicine & Parasitology*, 100, 647-662.

Roberts D.R., Manguin S. and Mouchet J. (2000) **DDT house spraying and re-emerging malaria**. *Lancet*, 356, 330-332.

Rogan W.J. and Chen A. (2005) **Health risks and benefits of bis(4-chlorophenyl)-1,1,1-trichloroethane (DDT)**. *Lancet*, 366, 763-773.

Rychlik W. (1993) Selection of Primers for polymerase chain reaction. In *Methods in Molecular Biology*, Vol. 15: PCR Protocols: Current Methods and Applications, pp. 31-40.

Salcedo E., Cortese J.F., Plowe C.V., Sims P.F.G. and Hyde J.E. (2001) **A bifunctional dihydrofolate synthetase-folylpolyglutamate synthetase in *Plasmodium falciparum* identified by functional complementation in yeast and bacteria**. *Molecular and Biochemical Parasitology*, 112, 239-252.

Salcedo E., Sims P.F.G. and Hyde J.E. (2005) **A glycine-cleavage complex as part of the folate one-carbon metabolism of *Plasmodium falciparum***.

Sanchez C.P., Stein D.S. and Lanzer M. (2007) **Is PfCRT a channel or a carrier? Two competing models explaining chloroquine resistance in *Plasmodium falciparum***. *Trends in Parasitology*, 23, 332-339.

Sanger F. and Coulson A.R. (1975) **A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase**. *Journal of Molecular Biology*, 94, 441-448.

Sayers J.R., Price H.P, Fallon P.G. and Doenhoff M.J. (1995) **AGA/AGG Codon usage in parasite: Implications for gene expression in *Escherichia coli***. *Parasitology Today*, 11, 345-346.

Shane B. (1980a) **Pterolylpoly(γ-glutamate) synthesis by Corynebacterium species. Purification and properties of folylpoly(γ-glutamate) synthetase**. *Journal of Biological Chemistry*, 255, 5655-5662.

Shane B. (1980b) **Pterolylpoly(γ-glutamate) synthesis by Corynebacterium species. Studies on the mechanism of folylpoly(γ-glutamate) synthetase**. *The Journal of Biological Chemistry*, 255, 563-5667.

Shane B. and Bognar A.L. (1983) **Purification and properties of *Lactobacillus casei* folylpoly-γ-glutamate synthetase**. *The Journal of Biological Chemistry*, 258, 12574-12581.

Shane B., Bognar A.L., Goldfrarb R.D. and Lebowitz J.H. (1983) **Regulation of folylpoly-γ-glutamate synthesis in bacteria: *In vivo* and *in vitro* synthesis fo pteroylpoly-γ-glutamates by *Lactobacillus casei* and *Steptococccus faecalis*.** *Journal of Bacteriology*, 153, 316-325.

Sheng Y., Ip H., Liu J., Davidson A. and Bognar A.L. (2003) **Binding of ATP and well as Tetrahydrofolate induced conformational changes in *Lactobacillus casei* folylpolyglutamate synthetase in solution**. *Biochemistry*, 42, 1537-1543.

Shimizu Y., Kuruma Y., Ying B.W., Umekage S. and Ueda T. (2006) **Cell-free translation systems for protein engineering**. *FEBS Journal*, 273, 4133-4140.

Shimp Jr R.L., Martin L.B., Zhang Y., Henderson B.S., Duggan P., MacDonald N.J., Lebowitz J., Saul A. and Narum D.L. (2006) **Production and characterization of clinical grade *Escherichia coli* derived *Plasmodium falciparum* 42 kDa merozoite surface protein 1 (MSP1$_{42}$) in the absence of an affinity tag.** *Protein Expression and Purification*, 50, 58-67.

Shonhani A., Boshoff A. and Blatch G.L. (2005) ***Plasmodium falciparum* heat shock protein 70 is able to suppress the thermosensitivity of an *Escherichia coli* DnaK mutant strain.** *Molecular Genetic Genomics*, 274, 70-78.

Singh S.M. and Panda A.K. (2005) **Solubilization and refolding of bacterial inclusion body proteins.** *Journal of Bioscience and Bioengineering*, 99, 303-310.

Sirawaraporn W., Prapunwattana P., Sirawaraporn R., Yuthavong Y. and Santi DV. (1993) **The dihydrofolate reductase domain of *Plasmodium falciparum* thymidylate synthase-dihydrofolate reductase.** *The Journal of Biological Chemistry*, 268, 21637-21644.

Skorokhod O.A., Alessio M., Mordmuller B., Arese P. and Schwarzer E. (2004) **Hemozoin (malarial pigment) inhibits differentiation and maturation of human monocyte-derived dendritic cells: A peroxisome proliferator-activated receptor-γ-mediated effect.** *The Journal of Immunology*, 173, 4066-4074.

Smith C.A., Cross J.A., Bognar A.L. and Sun X. (2006) **Mutation of Gly51 to serine in the P-loop of Lactobacillus casei folylpolyglutamate synthetase abolishes activity by altering the conformation of two adjacent loops.** *Biological Crystallography*, 62, 548-558.

Soldati D., Foth B.J. and Cowman A.F. (2004) **Molecular and functional aspects of parasite invasion**. *Trends in Parasitology*, 20, 567-574.

Sorensen H.P. and Mortensen K.K. (2005) **Advanced genetic strategies for recombinant protein expression in *Escherichia coli***. *Journal of Biotechnology*, 115, 113-128.

Spikes J.D. (1998) **Photosensitizing properties of quinine and synthetic antimalarials**. *New Trends in Photobiology*, 42, 1-11.

Stanger O. (2002) **Physiology of folic acid in health and disease**. *Current Drug Metabolism*, 3, 211-223.

Stokstad E.L.R. and Koch J. (1967) **Folic acid metabolism**. *Physiology Reviews*, 47, 83-116.

Stothard P. (2000) **The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences**. *BioTechniques*, 28, 1102-1104.

Studier F.W. (2005) **Protein production by auto-induction in high-density shaking cultures**. *Protein expression and Purification*, 41, 207-234.

Sun X., Bognar A.L., Baker E. and Smith C. (1998) **Structural homologies with ATP- and folate-binding enzymes in the crystal structure of folylpolyglutamate synthetase.** *Biochemistry*, 95, 6647-6652.

Sun X., Cross J.A., Bognar A.L., Baker E.N. and Smith C.A. (2001) **Folate-binding triggers the activation of folylpolyglutamate synthetase**. *Journal of Molecular Biology*, 310, 1067-1078.

Talman A.M., Domarle O., Mckenzie F.E., Ariey F. and Robert V. (2004) **Gametocytogenesis: the puberty of *Plasmodium falciparum***. *Malaria Journal*, 3, 1-14.

Tan X-J. and Carlson H.A. (2005) **Docking studies and ligand recognition in folylpolyglutamate synthetase**. *Journal of Medicinal Chemistry*, 48, 7764-7772.

Tekwani B.L. and Walker L.A. (2005) **Targeting the hemozoin synthesis pathway for new antimalarial drug discovery: Technologies for *In vitro* β-Hematin formation assay**. *Combinatorial Chemistry & High Troughput Screening*, 8, 63-79.

Terpe K. (2003) **Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems**. *Applied Microbiology Biotechnology*, 60, 523-533.

Thanaraj T.A. and Argos P. (1996) **Ribosome-mediated translational pause and protein domain organization**. *Protein Science*, 5, 1594-1612.

Thomas C.L. (1993) *Taber's Cyclopedic Medical Dictionary*. F.A Davis Company, Philadelphia.

Thomas J.G., Ayling A. and Baneyx F. (1997) **Molecular chaperones, folding catalysts and the recovery of active recombinant proteins form *E. coli***. *Applied Biochemistry and Biotechnology*, 66, 197-238.

Toure Y.T., Oduola A.M.J. and Morel C.M. (2004) **The *Anopheles gambiae* genome: next steps for malaria vector control**. *Trends in Parasitology*, 20, 142-149.

Trenholme K.R. and Gardiner D.L. (2004) **A sticky problem in Malaria**. *Biologist*, 51, 37-40.

Treweek T.M., Morris A.M. and Carver J.A. (2003) **Intracellular protein unfolding and aggregation: The role of small heat-shock proteins**. *Aust J Chem*, 56, 357-367.

Tripathi R.P., Mishra R.C., Dwivedi N., Tewari N. and Verma S.S. (2005) **Current status of Malaria Control**. *Current Medicinal Chemistry*, 12, 2643-2659.

Turner P.C., McLennan A.G., Bates A.D. and White M.R.H. (2000) *Molecular Biology*. BIOS Scientific Publishers, Liverpool, UK.

Turusov V., Rakitsky V. and Tomatis L. (2002) **Dichlordiphenyltrichloroethane (DDT): Ubiquity, persistence, and risks**. *Environmental Health Perspectives*, 110, 125-128.

Vallejo L.F. and Rinas U. (2004) **Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins**. *Microbial Cell Factories*, 3, 11-22.

Vallely A., Vallely L., Changalucha J., Greenwood B. and Chandramohan D. (2007) **Intermittent preventive treatment for malaria in pregnancy in Africa: What's new, what's needed?** *Malaria Journal*, 6, 16-28.

Vedadi M., Lew J., Artz J., Amani M., Zhao Y., Dong A., Wasney G.A., Gao M., Hills T., Brokx S., Qiu W., Sharma S., Diassiti A., Alam Z., Melone M., Mulichak A., Wernimont A., Bray J., Loppnau P., Plotnikova O., Newberry K., Sundararajan E., Houston S., Walker J., Tempel W., Bockharev A., Kozieradzki I., Edwards A., Arrowsmith C., Roos D., Kain K. and Hui R. (2007) **Genome-scale protein expression and structural biology of *Plasmodium falciparum* and related Apicomplexan organisms**. *Molecular and Biochemical Parasitology*, 151, 100-110.

Ventura S. and Villaverde A. (2006) **Protein quality in bacterial inclusion bodies**. *Trends in Biotechnology*, 24, 179-185.

Villaverde A. and Carrio M.M. (2003) **Protein aggregation in recombinant bacteria: biological role of inclusion bodies.** *Biotechnology Letters*, 25, 1385-1395.

Wall J.G. and Pluckthum A. (1995) **Effects of overexpressing folding modulators on the *in vivo* folding of heterologous proteins in *Escherichia coli*.** *Current Opinion in Biotechnology*, 6, 507-516.

Walter S. and Buchner J. (2002) **Molecular chaperones-Cellular Machines for Protein Folding**. *Angewandte Chemie Int. Ed.*, 41, 1098-1113.

Wang P., Brobey R.K.B., Horii T., Sims P.F.G. and Hyde J.E. (1999) **Utilization of exogenous folate in the human malaria parasite *Plasmodium falciparum* and its critical role in antifolate drug synergy**. *Molecular Microbiology*, 32, 1254-1262.

Wang P., Nirmalan N., Wang Q., Sims P.F.G. and Hyde J.E. (2004a) **Genetic and metabolic analysis of folate salvage in the human malaria parasite *Plasmodium falciparum*.** *Molecular and Biochemical Parasitology*, 135, 77-87.

Wang P., Wang Q., Aspinall T.V., Sims P.F.G. and Hyde J.E. (2004b) **Transfection studies to explore essential folate metabolism and antifolate drug synergy in the human malaria parasite *Plasmodium falciparum*.** *Molecular Microbiology*, 51, 1425-1438.

Wang P., Wang Q., Sims P.F.G. and Hyde J.E. (2007) **Characterisation of exogenous folate transport in *Plasmodium falciparum*.** *Molecular and Biochemical Parasitology*, 154, 40-51.

Waters A. (2006) **Malaria: New Vaccines for Old?** *Cell*, 124, 689-693.

Waugh D.S. (2005) **Making the most of affinity tags**. *Trends in Biotechnology*, 23, 316-320.

Weatherall D.J., Miller L.H., Baruch D.I., Marsh K., Doumbo O.K., Casals-Pascual C. and Roberts D.J. (2002) **Malaria and the Red Cell**. *American Society of Hematology*.

Whitfield J. (2002) **Portrait of a serial killer**. *Nature*.

Winstanley P.A. (2000) **Chemotherapy for Falciparum malaria: The armoury, the problems and the prospects**. *Parasitology Today*, 14, 146-153.

www.allafrica.com/stories. (2007) Africa: Re-consider DDT against malaria.

www.cdc.com. (2004) **Malaria Disease**.

www.imm.ul.pt/html/uni14i.html.

www.ncbi.nlm.gov.

www.rbm.who.int.

www.sigmaaldrich.com. (2007) Biological Detergent Theory and Application.

Yadava A. and Ockenhouse C.F. (2003) **Effect of codon optimization on expression levels of a functionally folded malaria vaccine candidate in Prokaryotic and Eukaryotic expression systems.** *Infection and Immunity*, 71, 4961-4969.

Yazdani S.S., Shakri A.R., Pattnaik P., Rizvi M.M.A. and Chitnis C.E. (2006) **Improvement in yield and purity of a recombinant malaria vaccine candidate based on the receptor-binding domain of *Plasmodium vivax* Duffy binding protein by codon optimization**. *Biotechnology Letters*, 28, 1109-1114.

Yeh I. and Altman R.B. (2006) **Drug targets for *Plasmodium falciparum*: A Post-genomic review/survey**. *Mini-Reviews in Medicinal Chemistry*, 6, 177-202.

Yuthavong Y., Yuvaniyama J., Chitnumsub P., Vanichtanankul J., Chusacultanachai S., Tarnchompoo B., Vilaivan T. and Kamchonwongpaisan S. (2005) **Malarial (*Plasmodium falciparum*) dihydrofolate reductase-thymidylate synthase: structural basis for antifolate resistance and development of effective inhibitors**. *Parasitology*, 130, 249-259.

Yuvaniyama J., Chitnumsub P., Kamchonwongpaisan S., Vanichtanankul J., Sirawaraporn W., Taylor P., Walkinshaw M.D. and Yuthavong Y. (2003) **Insights into the antifolate resistance from malarial DHFR-TS structures**. *Nature Structural Biology*, 10, 357-365.

Zhou Z., Schnake P., Xiao L. and Lal A.A. (2004) **Enhanced expression of a recombinant malaria candidate vaccine in *Escherichia coli* by codon optimization.** *Protein expression and Purification*, 34, 87-94.

Appendices

Appendix A

The following sequence alignment indicates the differences between codons of the native *dhfs-fpgs* gene sequence (Native), the synthetic *dhfs-fpgs* gene sequence (Synthetic) and the harmonised *dhfs-fpgs* gene sequence (Harmonised) The synthetic gene sequence was adapted to an *E. coli* codon preference (Refer to section 2.1.6).

```
                    10         20         30         40         50
           ....|....| ....|....| ....|....| ....|....| ....|....|
Native     atggaaaaaa atcaaaatga taaaagtaac aaaaatgata taattcacat
Synthetic  .....G..G. .C..G..C.. ......C... ......C.... .T........
Harmonised .......... .C..G..C.. C...TC...T ......C..C. .C........
Amino acid ~~M~~E~~K~ ~N~~Q~~N~~ D~~K~~S~~N ~~K~~N~~D~ ~I~~I~~H~~

                    60         70         80         90        100
           ....|....| ....|....| ....|....| ....|....| ....|....|
Native     gaatgataaa agtggaaatt atgataaaaa taatataaat aattttattg
Synthetic  ...C...... ..C..C..C. .......... C..C..T..C ..C.......
Harmonised ...C..C... TC...C..C. .C..C..... C..C..C..C ..C..C....
Amino acid M~~N~~D~~K ~~S~~G~~N~ ~Y~~D~~K~~ N~~N~~I~~N ~~N~~F~~I~

                   110        120        130        140        150
           ....|....| ....|....| ....|....| ....|....| ....|....|
Native     ataaaaatga tgagcatgat atgagtgaca tattacataa aataaataat
Synthetic  ....G..C.. ...A...... .....C..T. .TC.G..... ...T......
Harmonised .C.....C.. C.....C..C ...TC...T. .CC.G..C.. ...C..C..C
Amino acid ~D~~K~~N~~ D~~E~~H~~D ~~M~~S~~D~ ~I~~L~~H~~ K~~I~~N~~N

                   160        170        180        190        200
           ....|....| ....|....| ....|....| ....|....| ....|....|
Native     gaggaaaaaa aatatgaaga aataaaatct tacagtgagt gcttagaatt
Synthetic  .....G..G. .......... ...T...AGC ..T..C..A. ..C.G...C.
Harmonised .......... ...C..... ...C...AGC ..TTC..... .TC.G...C.
Amino acid ~~E~~E~~K~ ~K~~Y~~E~~ E~~I~~K~~S ~~Y~~S~~E~ ~.~~L~~E~~

                   210        220        230        240        250
           ....|....| ....|....| ....|....| ....|....| ....|....|
Native     attatataaa acacatgccc taaaattagg acttgataac ccaaaaaaat
Synthetic  GC.G...... ..C.....G. .G...C.G.. C..G...... ..G..G..GC
Harmonised GC.TC.C... ..C..C.... .C...C.G.. C..C..C..T ..G.......C
Amino acid L~~L~~Y~~K ~~T~~H~~A~ ~L~~K~~L~~ G~~L~~D~~N ~~P~~K~~K~

                   260        270        280        290        300
           ....|....| ....|....| ....|....| ....|....| ....|....|
Native     tgaacgaatc ttttggtcac ccttgtgata aatataaaac tattcatatt
Synthetic  ........AG C.....C..T ..G..C.... .......... C.........
Harmonised .C..T...AG C..C..C..T ..A..C..C. ....C..... ......C...
Amino acid ~L~~N~~E~~ S~~F~~G~~H ~~P~~C~~D~ ~K~~Y~~K~~ .~~I~~H~~I
```

```
                 310        320        330        340        350
           ....|....| ....|....| ....|....| ....|....| ....|....|
Native     gcagggacaa atgggaaagg gtctgtatgc tataaaatat atacatgtct
Synthetic  ..G..C..C. .C..C..... CAGC..G... .........T. ....C..C..
Harmonised ..T.....C. .C........ .AGC..T..T ..C.....C. .C..C..C..
Amino acid ~~.~~.~~T~ ~N~~.~~K~~ .~~S~~V~~. ~~Y~~K~~I~ ~Y~~T~~C~~

                 360        370        380        390        400
           ....|....| ....|....| ....|....| ....|....| ....|....|
Native     taaaataaaa aaattcaagg tgggtctttt ttcatcacct catatatttt
Synthetic  G..G..C... ....T..A. ....A..G.. .AGCAGC..G .....T...A
Harmonised C..C..C... ....T.... .C..C..C.. C..T..T..A ..C..C..CA
Amino acid L~~K~~I~~K ~~K~~F~~K~ ~V~~G~~L~~ F~~S~~S~~P ~~H~~I~~F

                 410        420        430        440        450
           ....|....| ....|....| ....|....| ....|....| ....|....|
Native     ctttaagaga aagaattata gtaaacgatg aaccaataag tgaaaaagag
Synthetic  GCC.GC.C.. .C.C.....T ..G....... ....G..T.. C.......A
Harmonised GCC.GC.T.. .C.T.....C ..T..T..C. ....G..CTC ..........
Amino acid ~S~~L~~R~~ E~~R~~I~~I ~~V~~N~~D~ ~E~~P~~I~~ S~~E~~K~~E

                 460        470        480        490        500
           ....|....| ....|....| ....|....| ....|....| ....|....|
Native     ttaatacatt tagtaaacga agtattaaat aaagccaaaa aattatatat
Synthetic  C.G..T...C .G..G..... ...GC.G..C .....G..G. .GC.G.....
Harmonised C.G..C..CC .G..T..T.. ...TC.G..C ........ ..C.G..C..
Amino acid ~~L~~I~~H~ ~L~~V~~N~~ E~~V~~L~~N ~~K~~A~~K~ ~K~~L~~Y~~

                 510        520        530        540        550
           ....|....| ....|....| ....|....| ....|....| ....|....|
Native     aaatccatct ttttttgaaa taattacatt agttgcattt ttacattttt
Synthetic  T..C..GAGC .......... .T.....CC. G..G..G... C.G.......
Harmonised C..C..GAGC ..C..C.... .C.....CC. G.....T..C C.G..C..CC
Amino acid I~~N~~P~~S ~~F~~F~~E~ ~I~~I~~T~~ L~~V~~.~~F ~~L~~H~~F~

                 560        570        580        590        600
           ....|....| ....|....| ....|....| ....|....| ....|....|
Native     taaataagaa ggtagattat gctataatag aaacagggat tggagggcgc
Synthetic  ....C..A.. A..G..... ..G..T..T. ....C..C.. ...C..C...
Harmonised .G..C..... ...T..C..C .....C..C. .....C.... ...C.....A
Amino acid ~L~~N~~K~~ K~~V~~D~~Y ~~A~~I~~I~ ~E~~T~~.~~ I~~G~~.~~R

                 610        620        630        640        650
           ....|....| ....|....| ....|....| ....|....| ....|....|
Native     ttagatgcaa ctaatatatt aaccaaacca gaagttattg taattacttc
Synthetic  C.G.....C. .C..C..TC. G........G .....G.... .G.....CAG
Harmonised C.G..C..T. ....C..CC. G..G.....G ......T... .G.....CAG
Amino acid ~~L~~D~~.~ ~.~~N~~I~~ L~~T~~K~~P ~~E~~V~~I~ ~V~~I~~.~~

                 660        670        680        690        700
           ....|....| ....|....| ....|....| ....|....| ....|....|
Native     cataggatat gatcatttaa atatattagg tgataatttg cctattatat
Synthetic  ...T..C... .....C.G. .C..TC.G.. C.....CC.. ..G....T.
Harmonised G..C..C..C ..C..CC.G. .C..CC.G.. C..C..CC.C ..A.....C.
Amino acid S~~I~~G~~Y ~~D~~H~~L~ ~N~~I~~L~~ G~~D~~N~~L ~~P~~I~~I~
```

```
                710        720        730        740        750
          ....|....| ....|....| ....|....| ....|....| ....|....|
Native    gtaatgaaaa aattggaatt tttaaaaaag atgctaacgt tgtaatagga
Synthetic .C.C..... ......C... .......... ....G..... G..G..T..C
Harmonised .C.C..... ......C... ..C....... .C.....T.. ...T..C..C
Amino acid ~C~~N~~E~~ K~~I~~G~~I  F~~K~~K~  ~D~A~~N~~ V~~V~~I~~G

                760        770        780        790        800
          ....|....| ....|....| ....|....| ....|....| ....|....|
Native    ccatcagtag ctatttataa aaatgttttt gataaggcaa aagaattaaa
Synthetic ..GAGC..G. .G........ ...C..G... .....A..G. .....C.G..
Harmonised ..G..T..T. ........C. ...C.....C .C.....T.. ...C.G..
Amino acid ~~P~~S~~V~ ~A~~I~~Y~~ K~~N~~V~F ~~D~~K~~.~ ~K~~E~~L~~

                810        820        830        840        850
          ....|....| ....|....| ....|....| ....|....| ....|....|
Native    ttgtactata catactgtag tacctgaacc acgaggagaa agatataatg
Synthetic C.C..C..T .....C..T. .G..G..... G..C..C... C.C.....C.
Harmonised C.C.....C ..C.....T. .T..A..... G.....C... C.T..C..C.
Amino acid N~~C~~.~~I ~~H~~.~~V ~V~P~~E~~ P~R~~G~~E ~~R~~Y~~N~

                860        870        880        890        900
          ....|....| ....|....| ....|....| ....|....| ....|....|
Native    aagaaaattc aagaatagca ttgcgcactt tagaaatatt aaatataagt
Synthetic .......CAG CC.C..T..G C.......CC .G.....TC. G..C..T..C
Harmonised .......C.. TC.T..C..T C.C..A...C .G.....CC. G..C..CTC.
Amino acid ~E~~E~~N~~ S~~R~~I~~. ~~L~~R~~.~ ~L~E~~I~~ L~~N~~I~~S

                910        920        930        940        950
          ....|....| ....|....| ....|....| ....|....| ....|....|
Native    atagattact ttttaaagtc cattatacca attaaaccac ctttaagaat
Synthetic ..T.....T. ..C.G..AAG ......T..G .........G. .GC.GC.C..
Harmonised ..C..C..T. .CC.G..... G.....C..G .........G. .AC.GC.T..
Amino acid ~~I~D~~Y~ ~F~~L~~K~~ S~~I~~I~~P ~~I~K~~P~ ~P~~L~~R~~

                960        970        980        990        1000
          ....|....| ....|....| ....|....| ....|....| ....|....|
Native    tcaatattta gctaccgagc aaatacaaca tataaagaaa aaattttcac
Synthetic ...G...C.G ..G.....A. .G..T..G.. ...T.....G ..G..CAGC.
Harmonised ...G..CC.G ....G.... .G..C..G.. C..C...... .....C..T.
Amino acid I~~Q~~Y~~L ~~A~~T~~E~ ~Q~~I~~Q~~ H~~I~~K~~K ~~K~~F~~S~

                1010       1020       1030       1040       1050
          ....|....| ....|....| ....|....| ....|....| ....|....|
Native    cagataattt agaacacaat gtacaatatc ctctagctgt tattttagat
Synthetic ....C..CC. G.....T..C ..G..G.... .G..G..G.. G...C.G...
Harmonised .G..C..CC. G.....T..C ..T..G..C. .A..C..... ....C.G..C
Amino acid ~P~~D~~N~~ L~~E~~H~~N ~~V~~Q~~Y~ ~P~~L~~A~~ V~~I~~L~~D

                1060       1070       1080       1090       1100
          ....|....| ....|....| ....|....| ....|....| ....|....|
Native    gtaggccaca atgaaacggc aatagataga ttgtgtaccg atattaatta
Synthetic ..G.....T. .C.....C.. G..T...C.C C...C..... ......C..
Harmonised ..T..G..T. .C........ T..C..CC.T C.C..C..G. .C.....C..
Amino acid ~~V~~G~~H~ ~N~~E~~T~~ .~~I~~D~~R ~~L~~C~~T~ ~D~~I~~N~~
```

131

```
                1110        1120        1130        1140        1150
           ....|....| ....|....| ....|....| ....|....| ....|....|
Native     ttttcataag ggacaaaata taaggatatg tatatctata acaaagccga
Synthetic  .........A ..C..G..C. .TC.C..T.. C..TAGC..T ..C..A...C
Harmonised C..C..C... ..C..G..C. .CC.C..C.. C..CAGC..C ..C.....CC
Amino acid Y~~F~~H~~K ~~G~~Q~~N~ ~I~~R~~I~~ C~~I~~S~~I ~~T~~K~~P~

                1160        1170        1180        1190        1200
           ....|....| ....|....| ....|....| ....|....| ....|....|
Native     gaaacttgag tgtattccat ccgttcatag ctcaatttgg cgacacctta
Synthetic  .C...C.... C..G..T... ....T..T. .G..G..... ...T...C.G
Harmonised .T..TC.CTC ...T..T..C ..C..T..C. ....G..C.. G..T..GC.G
Amino acid ~R~~N~~L~~ S~~V~~F~~H ~~P~~F~~I~ ~A~~Q~~F~~ G~~D~~T~~L

                1210        1220        1230        1240        1250
           ....|....| ....|....| ....|....| ....|....| ....|....|
Native     aaggatatat tttaccttcc atcattaaat gaaaggacct acgactttga
Synthetic  ..A.....T. ....T..G.. GAGCC.G..C ...C.C.... .T..T.....
Harmonised .....C..C. .C..T..C.. G..TC.G..C ...C.C...G. .T..T..C..
Amino acid ~~K~~D~~I~ ~F~~Y~~L~~ P~~S~~L~~N ~~E~~R~~T~ ~Y~~D~~F~~

                1260        1270        1280        1290        1300
           ....|....| ....|....| ....|....| ....|....| ....|....|
Native     ggaaatcgtt gagatgttaa ataatgaaga agaaataaaa aatgaaataa
Synthetic  A.....T..G ..A...C.G. .C..C..... ........T.... ..C.....T.
Harmonised ......A... ......C.G. .C..C..... ........C.... ..C.....C.
Amino acid E~~E~~I~~V ~~E~~M~~L~ ~N~~N~~E~~ E~~E~~I~~K ~~N~~E~~I~

                1310        1320        1330        1340        1350
           ....|....| ....|....| ....|....| ....|....| ....|....|
Native     aagagctcat attaagtagc tctaaaaagg ttggtaaatg gttagctcat
Synthetic  ....A..G.. TC.G..C... AGC.....A. .G..C..... .C.G..G...
Harmonised .....T.A.. CC.GTC.TCG AGC........ .....C.... .C.G.....C
Amino acid ~K~~E~~L~~ I~~L~~S~~S ~~S~~K~~K~ ~V~~G~~K~~ W~~L~~A~~H

                1360        1370        1380        1390        1400
           ....|....| ....|....| ....|....| ....|....| ....|....|
Native     gagaaacaag gaaatattaa tgaagaagat gcacttaaat tatacaaaag
Synthetic  ..A.....G. .C..C..... C......... ..G..G...C .G..T...C.
Harmonised ........G. .C..C..... C........C ..T..C...C .G..T...C.
Amino acid ~~E~~K~~Q~ ~G~~N~~I~~ N~~E~~E~~D ~~.~~L~~K~ ~L~~Y~~K~~

                1410        1420        1430        1440        1450
           ....|....| ....|....| ....|....| ....|....| ....|....|
Native     aggatgcata ccattaataa tcaaaaatgc atttcttgaa tgttgcaaag
Synthetic  C..C.....T ..GC.G..T. .T.....C.. G.....G..G ...........
Harmonised T..C..T..C ..GC.G..C. .A.....C.. T..C..C... ..C..T....
Amino acid R~~G~~.~~I ~~P~~L~~I~ ~I~~K~~N~~ .~~F~~L~~E ~~C~~.~~K~

                1460        1470        1480        1490        1500
           ....|....| ....|....| ....|....| ....|....| ....|....|
Native     acaatagtat attattagta tgtgggacgt ttttcgtgtt tgatgaggtt
Synthetic  .T..C..C.. TC.GC.G..T .....T..C. .C......... .......A..G
Harmonised .T..CTC... CC.GC.G..T ..C....... .C..T..C.. C..C......
Amino acid ~D~~N~~S~~ I~~L~~L~~V ~~C~~.~~T~ ~F~~F~~V~~ F~~D~~E~~V
```

132

```
              1510        1520        1530        1540        1550
           ....|....| ....|....| ....|....| ....|....| ....|....|
Native     ttaaacgttt ttgatattca ctctgatatg caagatacca tttttatgaa
Synthetic  C.G.....G. .......... TAGC...... ..G....... ..........
Harmonised C.G..T.... .C..C..... TAGC..C... ..G..C..G. ....C.....
Amino acid ~~L~~N~~V~ ~F~~D~~I~~ H~~S~~D~~M ~~Q~~D~~T~ ~I~~F~~M~~
```
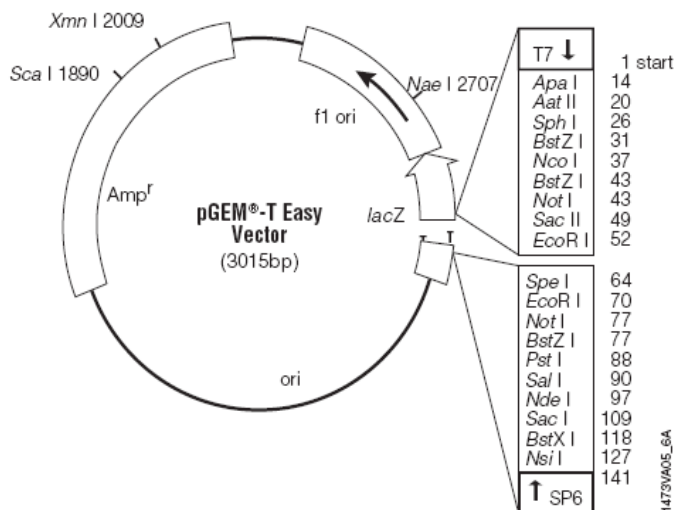
```
              1560
           ....|....| ....|....
Native     tgaaccatct cttgtttaa
Synthetic  C.....GAGC ..G..G...
Harmonised C.....GAGC ..C....G.
Amino acid N~~E~~P~~S ~~L~~V~~
```

Appendix B

pGEM® T-Easy Vector map

Xmn I 2009

Sca I 1890

Nae I 2707

f1 ori

Amp^r

pGEM®-T Easy Vector
(3015bp)

lacZ

ori

T7 ↓	1 start
Apa I	14
Aat II	20
Sph I	26
BstZ I	31
Nco I	37
BstZ I	43
Not I	43
Sac II	49
EcoR I	52

Spe I	64
EcoR I	70
Not I	77
BstZ I	77
Pst I	88
Sal I	90
Nde I	97
Sac I	109
BstX I	118
Nsi I	127
	141

↑ SP6

1473VA05_6A

pGEM®-T Easy Vector sequence reference points:

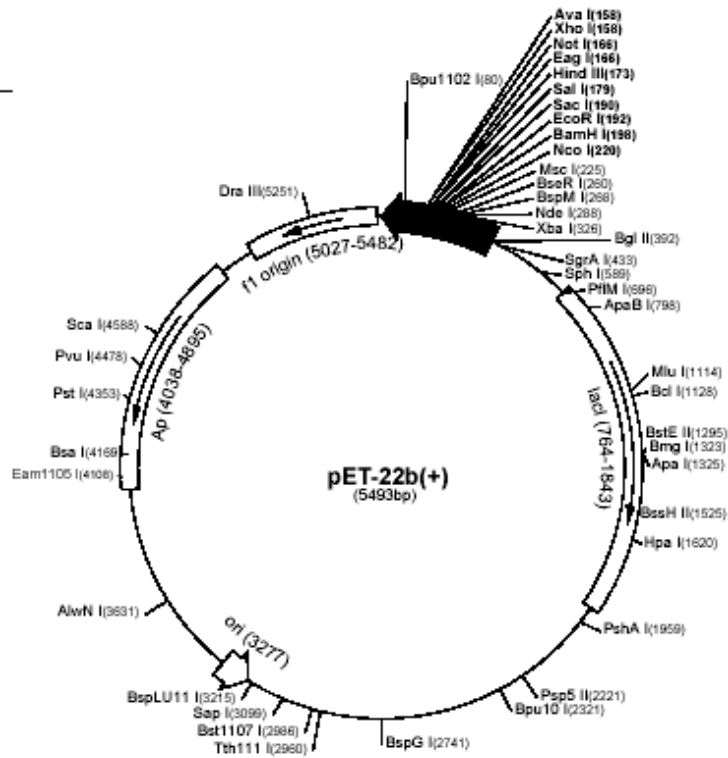|	
T7 RNA polymerase transcription initiation site	1
multiple cloning region	10–128
SP6 RNA polymerase promoter (–17 to +3)	139–158
SP6 RNA polymerase transcription initiation site	141
pUC/M13 Reverse Sequencing Primer binding site	176–197
lacZ start codon	180
lac operator	200–216
β-lactamase coding region	1337–2197
phage f1 region	2380–2835
lac operon sequences	2836–2996, 166–395
pUC/M13 Forward Sequencing Primer binding site	2949–2972
T7 RNA polymerase promoter (–17 to +3)	2999–3

Position of the A/T cloning site is indicated in the multiple cloning cassettes within the *lacZ* gene. The T7- and SP6 promoter regions indicate vector primer positions respectively. The ampicillin resistance gene (Amp^r) is also indicated.

Appendix C

pET22b vector

C-terminal His6-tagged and Tagless expression:

Appendix D

Vector map for chaperone proteins

(www.takara-bio.inc.)

Plasmid	Chaperone	Promotor	Inducer	Resistant Marker
pG-KJE8	dnaK-dnaJ-grpE groES-groEL	araB Pzt1	L-Arabinose Tetracyclin	Cm
pGro7	groES-groEL	araB	L-Arabinose	Cm
pKJE7	dnaK-dnaJ-grpE	araB	L-Arabinose	Cm
pG-Tf2	groES-groEL-tig	Pzt1	Tetracyclin	Cm
pTf16	Tig	araB	L-Arabinose	Cm