UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Optimization of the heterologous expression of folate metabolic enzymes of *Plasmodium falciparum*

By

Shivani Goolab

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

December 2010

© University of Pretoria

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:…….

Plagiarism Declaration

UNIVERSITY OF PRETORIA

FACULTY OF NATURAL AND AGRICULTURAL SCIENCES

DEPARTMENT OF BIOCHEMISTRY

Full name: _____ Student number: _____

Title of the work: _____

Declaration

1. I understand what plagiarism entails and am aware of the University's policy in this regard.
2. I declare that this _____ (e.g. essay, report, project, assignment, dissertation, thesis etc) is my own, original work. Where someone else's work was used (whether from a printed source, the internet or any other source) due acknowledgement was given and reference was made according to departmental requirements.
3. I did not make use of another student's previous work and submit it as my own.
4. I did not allow and will not allow anyone to copy my work with the intention of presenting it as his or her own work.

Signature _____ Date _____

Acknowledgements

This thesis would not have been possible without my supervisor, Prof Louw, whose continuous feedback, guidance and patience allowed me to delve deeper into analyzing the subject.

I am to grateful to my co-supervisor, Prof Birkholtz, whose encouragement, guidance and support from the beginning right through the end allowed me to develop an understanding and skills of the subject.

I would like to thank Dr Stephens, from Rhodes University for her continuous input and help with my experimental work.

I am indebted to all the staff, students of various research groups at the University of Pretoria for their support especially Esmare Human for her enthusiasm and help throughout my postgraduate experience.

I would like to acknowledge the National Research Foundation for awarding me the UK-SA bursary and the Department of Science and Technology of South Africa for funding the SAMI programme.

I owe my deepest gratitude to my family for their blessings, patience and understanding to allow me to follow through on my career, to my loved ones for their faith, love and support on this journey.

To my dearest Guruji, “Words are too small to convey the Truth yet they can trigger the flames of wisdom, shower comfort on the intense longing, raise cheers of joy and make you know I treasure Your love for me. Like God his creation, the root its trunk, unknown, unheard, unseen, in the silent chamber of heart I am in You as You, You are in me as me.”

Summary

Malaria is a fatal tropical disease affecting billions of people in impoverished countries world-wide. An alarming fact is that a child in Africa dies of malaria every 30 seconds that amounts to 2500 children per day (www.who.int/features/factfiles). Malaria is caused by the intraerythrocytic forms of *Plasmodium* species, notably *P. falciparum*, P. *vivax, P. ovale* and *P. malariae* (Hyde 2007). The spread of drug-resistant strains, failure of vector control programs, rapid growth rate of the parasite, and lack of a vaccine have further exacerbated the effects of malaria on economic development and human health. It is therefore imperative that novel drug targets are developed or current antimalarial drugs optimized (Foley and Tilley 1998).

One such target is folate biosynthesis, given that folates and their derivatives are required for the survival of organisms (Muller *et al.* 2009). DHFR and DHPS are currently the only folate targets exploited however, their antifolate drugs are almost useless against parasite resistant strains. As such, guanosine-5'triphosphate cyclohydrolase I (GTPCHI) among other antifolate candidates are considered for intervention (Lee *et al.* 2001). Knock-out studies (of *P. falciparum gtpch*I) resulted in the suppression of DHPS activity (Nzila *et al.* 2005). Additionally, *gtpch*I amplified 11-fold in *P. falciparum* strains resistant to antifolates due to mutations in *dhps* and *dhfr* and this may be a mechanism for the compensation of reduced flux of folate intermediates (Kidgell *et al.* 2006; Nair *et al.* 2008).

Over-expression of *P. falciparum* proteins in *E. coli* remains a challenge mainly due to the A+T rich *Plasmodium* genome resulting in a codon bias. This results in the expression of recombinant proteins as insoluble proteins sequestered in inclusion bodies (Carrio and Villaverde 2002; Mehlin *et al.* 2006; Birkholtz *et al.* 2008a).

Comparative expression studies were conducted of native GTPCHI (nGTPCHI), codon optimized GTPCHI (oGTPCHI) and codon harmonized (hGTPCHI) in various *E. coli* cell lines, using alternative media compositions and co-expression with *Pf*hsp70. The nGTPCHI protein did not express because the gene consisted of codons rarely used by

E. coli (codon bias). The expression levels of purified hGTPCHI were a greater in comparison to oGTPCHI using the different expression conditions. This is because codon-harmonization involves substituting codons to replicate the codon frequency preference of the target gene in *P. falciparum*, as such the translation machinery matches that of *Plasmodium* (Angov *et al.* 2008). Furthermore, greater expression levels of GTPCHI were achieved in the absence of *Pf*hsp70 due to expression of a possible N-terminal deletion product or *E. coli* protein. Purification conditions could be improved to obtain homogenous GTPCHI and further analysis (mass spectrometry and enzyme activity assays) would be required to determine the nature of soluble GTPCHI obtained. To improve the expression of soluble proteins the wheat germ expression system was used as an alternate host. However, GTPCHI expression was not effective, possibly due to degradation of mRNA template or the absence of translation enhancer elements.

Index

List of Figures.....ix

List of Tables.....xi

Abbreviations.....xii

Chapter 1: Introduction

1.1 MALARIA: A LIABILITY TO THE WORLD.....2

1.2 PARASITE LIFE-CYCLE.....3

1.3 PATHOGENICITY OF *Plasmodium*.....4

1.4 RESTRAINING MALARIA.....6

1.4.1 Vector control.....6
1.4.2 Vaccines.....8
1.4.3 Diagnostics.....9
1.4.4 Malaria control via chemotherapy.....10
1.4.4.1 Drug combinations.....11
1.4.4.2 Quinoline.....14
1.4.4.2.1 Chloroquine.....14
1.4.4.3 Artemisinin derivatives.....16
1.4.4.4 Atovaquone (naphtoquinone).....17
1.4.4.5 Antifolates.....17

1.5 ADVANCES IN ANTIMALARIAL TARGETS:.....19

1.5.1 New permeability pathways.....19
1.5.2 Mitochondrion.....19
1.5.3 Apicoplast.....20
1.5.4 Food vacuole.....20
1.5.5 Microtubulins.....21
1.5.6 Oxidative stress.....21
1.5.7 Polyamine biosynthesis.....22

1.6 NUCLEIC ACID METABOLISM.....23

1.6.1 Thymidine synthesis via the folate pathway.....25
1.6.1.1 Folate salvage.....27
1.6.1.2 Antifolate targets worth probing.....29
1.6.1.2.1 6-pyruvoyltetrahydrobiopterin synthase (PTPS).....29
1.6.1.2.2 Hydroxymethylpterin pyrophosphokinase (HPPK, E.C. 2.7.6.3).....29
1.6.1.2.3 Dihydrofolate synthase (DHFS, E.C. 6.3.2.12).....30
1.6.1.2.4 Serine hydroxymethyltransferase (SHMT, E.C. 2.1.2.1).....31
1.6.2 Guanosine triphosphate cyclohydrolase (GTPCHI, E.C. 3.5.4.16).....32
1.6.2.1 Characteristics of the active proteins GTPCHI in other species.....32
1.6.2.2 Copy number variation and antifolate drug resistance.....34

RESEARCH AIMS.....35

Chapter 2: Heterologous expression and optimization of native, codon optimized and harmonized *P.falciparum* GTPCHI using *E. coli*

2.1 INTRODUCTION:....36

2.1.1 *E. coli* as an expression host....36
2.1.2 Optimization of recombinant protein expression....38
2.1.2.1 Expression vector system....38
2.1.2.2 Codon harmonization....41
2.1.2.3 Co- expression: Molecular chaperone proteins....43
2.1.2.4 Auto-induction....45

2.2 MATERIALS AND METHODS....46

2.2.1 Introduction of *P. falciparum gtpch*I....46
2.2.2 Primer design....46
2.2.3 Nucleic acid concentration determinations....46
2.2.4 PCR gene amplification....47
2.2.4.1 n*gtpch*I....47
2.2.4.2 o*gtpch*I....48
2.2.5 Cloning....48
2.2.5.1 The pGEM® T-Easy vector system....48
2.2.6 Preparation of competent cells....49
2.2.6.1 Calcium chloride heat shock cells....49
2.2.6.2 Electrocompetent cells....49
2.2.7 Transformation....50
2.2.7.1 Heat shock method....50
2.2.7.2 Electrocompetent cells....51
2.2.8 Identification of recombinant clones....51
2.2.8.1 Colony PCR n*gtpch*I and o*gtpch*I....51
2.2.8.2 Plasmid isolation....52
2.2.8.3 Screening of recombinant clones....52
2.2.9 Sub-cloning into the pET22b vector system....53
2.2.9.1 Restriction enzyme digestion....53
2.2.9.2 Cloning into pET22b....53
2.2.9.3 Screening of recombinant clones....54
2.2.10 Codon harmonization of *P. falciparum gtpch*I....54
2.2.10.1 RNA structure predictions and final alterations to h*gtpch*I....55
2.2.11 Sub-cloning into the pET22b vector system....56
2.2.11.1 Restriction enzyme digestion....56
2.2.11.2 Cloning into pET22b....57
2.2.12 Identification of recombinant clones....57
2.2.12.1 Colony PCR h*gtpch*I....57
2.2.12.2 Screening of recombinant clones....58
2.2.13 Nucleotide sequencing....58
2.2.14 Protein expression....59
2.2.14.1 *E. coli* cell lines....59
2.2.14.2 Cell membrane disruption and protein isolation....60
2.2.14.3 Co-expression: Molecular chaperone proteins....61
2.2.14.4 Auto-induction media....62
2.2.15 Protein purification....63
2.2.15.1 Affinity chromatography....63
2.2.16 Protein concentration determination....64
2.2.17 SDS-PAGE analysis....65
2.2.18 Western blotting....66

2.3 RESULTS....68

2.3.1 Primers....68
2.3.2 Amplication and cloning of n*gtpch*I and o*gtpch*I into pGEM®-T Easy vector....69
2.3.3 Selection of positive recombinant clones of n*gtpch*I and o*gtpch*I- pGEM®-T Easy....69
2.3.4 Isolation and restriction enzyme digestion of pET22b....71
2.3.5 Selection of positive recombinant clones of n*gtpch*I and o*gtpch*I- pET22b....71
2.3.6 Comparison of codon frequency of the n*gtpch*I and h*gtpch*I....73
2.3.7 RNA structure predictions of h*gtpch*I....74
2.3.8 Isolation of h*gtpch*I-pUC57 and pET22b....75
2.3.9 Selection of positive recombinant clones of h*gtpch*I- pET22b....76
2.3.10 Nucleotide sequencing....76
2.3.11 Comparative protein expression....76
2.3.11.1 Comparative growth study....76
2.3.11.2 Protein expression levels as a function of induction time....77
2.3.11.3 BL21 (DE3) star cells....80
2.3.11.4 M15 pREP4 cells....82
2.3.11.5 Chaperone expression....84
2.3.11.6 *P. falciparum* GTPCHI His-tag protein purification....86
2.3.12 Auto-induction media....88

2.4 DISCUSSION....91

2.4.1 Codon harmonization and RNA structure prediction....91
2.4.2 Expression: Codon harmonization opposed to codon optimization....93
2.4.3 *E. coli* cell lines....95
2.4.4 *P. falciparum* molecular chaperone....96
2.4.5 Auto-induction media....98

Chapter 3: Cell-free expression of native, codon optimized and codon harmonized *P. falciparum* GTPCHI

3.1 INTRODUCTION:....100

3.1.1 Expression host used for Plasmodium protein synthesis....100
3.1.2 Yeasts....101
3.1.3 Baculovirus....102
3.1.4 Plant cells....103
3.1.5 Mammalian cells....104
3.1.6 Cell-free expression system....105

3.2 METHODS AND MATERIALS....110

3.2.1 Linearization of the pET22b- *Pfgtpch*I....110
3.2.2 Creating blunt ended plasmid DNA....110
3.2.3 *In vitro* transcription....111
3.2.4 High pure RNA cleanup....111
3.2.5 Cell-free protein expression....112
3.2.6 SDS-PAGE analysis....113

3.3 RESULTS....113

3.3.1 Linearization of the pET22b- *Pfgtpch*I....113
3.3.2 RNA yield and purity....114
3.3.3 Cell free expression system....115

3.4 DISCUSSION....116

Chapter 4: Concluding discussion and future perspectives

4.1 INSIGHTS INTO *Plasmodium* BIOLOGY....121

4.2 TARGETING FOLATE METABOLISM....122

4.3 PROBING INTO *P. falciparum* GTPCHI....123

4.4 EXPRESSION OF SOLUBLE GTPCHI....124

4.5 FUTURE PERSPECTIVES....128

4.5.1 Improving the yield of soluble proteins....128
4.5.2 Downstream use of GTPCHI....129

Appendices....130
References....151

LIST OF FIGURES

Chapter 1

Figure 1-1: People of Africa at risk of contracting malaria……..…2
Figure 1-2: The life-cycle of *P. falciparum*……..…4
Figure 1-3: The morphological organization of a *P. falciparum* merizoite……6
Figure 1-4*:* The *Plasmodium* life cycle including phases targeted by antimalarial drugs ……10
Figure 1-5: A representation of the structure of a parasite-infected red blood cell and novel antimalarial drug targets in *Plasmodium* parasite……..……..21
Figure 1-6: The interconnectivity of purine salvage, pyrimidine synthesis and folate metabolism in *P. falciparum*……..……....25
Figure 1-7: a) The synthesis of reduced folates via *de novo* and b) salvage pathways (thymidylate cycle)……..…….…...27
Figure 1-8: The first reaction of folate biosynthesis in *P. falciparum* catalyzed by GTPCHI……...33

Chapter 2

Figure 2-1 Strategies for expression of *Plasmodium* proteins……38
Figure 2-2: The codon harmonization algorithm……..…….42
Figure 2-3: PCR amplification of the full length n*gtpch* and o*gtpch*I gene……69
Figure 2-4: Screening of *Pf* n*gtpch*I- pGEM®-T easy vector……70
Figure 2-5: Screening of *Pf* o*gtpch*I- pGEM®-T Easy vector……...70
Figure 2-6: The *BamH*I and *Nde*I restriction vector map of pET22b……...71
Figure 2-7: Screening of recombinant clone of n*gtpch*I- pET22b……72
Figure 2-8: Screening of recombinant clone of o*gtpch*I- pET22b……72
Figure 2-9: The average codon frequency of the wild type *gtpch*I gene in *P. falciparum* and *E. coli*, including the h*gtpch*I sequence in *E. coli* (10 codons)……..……..73
Figure 2-10: A mountain plot representation of 17 predicted secondary structures of the RNA of the h*gtpch*I gene in *P. falciparum* and the base pairing…..……....74
Figure 2-11: Restriction enzyme digestion of pUC57-h*gtpch*I……..……..…….....75
Figure 2-12: Screening of recombinant clone of h*gtpch*I- pET22b……..……..76
Figure 2-13: Growth curves used to monitor the nGTPCHI, oGTPCHI and hGTPCHI his tag protein expression……..……..…77
Figure 2-14: The level of expression of the nGTPCHI, oGTPCHI and hGTPCHI his tag proteins in different expression cell lines as a function of the induction time……..…78

Figure 2-15: The relative quantity measured for total fractions of the A) nGTPCHI, B) oGTPCHI and C) hGTPCHI his tag proteins obtained as pre-induced, post induced……....79
Figure 2-16: Expression of the nGTPCHI, oGTPCHI and hGTPCHI proteins in star cell line……..…..80
Figure 2-17: The relative quantity measured for the total (green bars), soluble (blue bars) and insoluble (red bars) fractions of the oGTPCHI and hGTPCHI proteins in star cells……81
Figure 2-18: Expression of the nGTPCHI, oGTPCHI and hGTPCHI in m15 cells……82
Figure 2-19: The relative quantity measured for total (green bars), soluble (blue bars) and insoluble (red bars) of the nGTPCHI, oGTPCHI and hGTPCHI proteins in m15 cells……..……....83
Figure 2-20: Expression of the nGTPCHI, oGTPCHI and hGTPCHI in star 70 cell line……..84
Figure 2-21: The relative quantity measured for total (green bars), soluble (blue bars) and insoluble (red bars) of the oGTPCHI and hGTPCHI in star 70 cells……..…..…85
Figure 2-22: SDS-PAGE and western blot analysis of Ni^{2+} ion affinity purification of the nGTPCHI, oGTPCHI and hGTPCHI his tag proteins……...86
Figure 2-23: The relative quantity measured for the western blot analysis of purified fractions of oGTPCHI and hGTPCHI his tag proteins……87
Figure 2-24: Expression of the nGTPCHI, oGTPCHI and hGTPCHI his tag proteins in A) star 70 and B) m15 cells……..…..88
Figure 2-25: A) SDS-PAGE and B) western blot analysis of Ni^{2+} ion affinity purified nGTPCHI, oGTPCHI and hGTPCHI his tag proteins from star 70 and m15 cell lines, respectively……..…...89
Figure 2-26: The relative quantity measured for the western blot analysis of purified fractions of nGTPCHI, oGTPCHI and hGTPCHI his tag proteins……90

Chapter 3

Figure 3-1: The process of cell-free synthesis and analysis of proteins……..107
Figure 3-2: Protein production route using the wheat germ cell-free system……..109
Figure 3-3: The linearization of h*gtpch*I-pET22b, o*gtpch*I-pET22b and n*gtpch*I-pET22b……114
Figure 3-4: Expression of the nGTPCHI, oGTPCHI and hGTPCHI proteins in wheat germ extracts……..…115

LIST OF TABLES

Chapter 1

Table 1-1: Currently used antimalarial drug combinations……..……12

Table 1-2: Currently developed commercial antimalarial drugs and targets……..……..13

Table 1-3:The comparison of folate metabolism in *Plasmodium* and humans……..26

Chapter 2

Table 2-1: attributes of *P. falciparum* genome and reasons for poor expression……37

Table 2-2: The strategy of optimizing recombinant expression of *P. falciparum* proteins in *E.coli*……..……39

Table 2-3: vector-construct and the gene and protein annotations utilized…….....46

Table 2-4: Vector-construct, host expression lines and annotations utilized in text……..…..60

Table 2-5: Composition of media used for auto induction studies……....63

Table 2-6: Primers used for the subcloning of GTPCHI domains for expression of proteins in the pET22b His-tag expression system……..……68

Chapter 3

Table 3-1: The concentration of purified RNA and the purity (260/280) obtained for h*gtpch*IpET22b, o*gtpch*I-pET22b and n*gtpch*I-pET22b……..114

ABBREVIATIONS

ACT	Artemisinin-based combination therapies
AS	Artemisinin
ATP	Adenosine triphosphate
ATQ	Atovaquone
Bp	Base pair
BSA	Bovine serum albumin
Bti	Bacillus thuringiensis israelensis
CDK	Cyclin-dependent protein kinase
CNP	Copy number polymorphism
CQ/R	Chloroquine/ resistance
CSP	Circumsporozoite protein
D/FV	Digestive/ food vacuole
DDT	bis(4-chlorophenyl)-1,1,1,1-trichlorethane
DHF	Dihydrohydrofolate
DHFR-TS	Dihydrofolate reductase-Thymidylate synthase
DHNA	Dihydroneopterin aldolase
DHPS	Dihydroopterate synthase
DNA	Deoxyribonucleic acid
DnaseI	Deoxyribonuclease I
DOXP	1-deoxy-D-xylulose-5-phosphate
(d)UMP	(deoxy)Uridine 5'-monophosphate
(d)UTP	(deoxy)Uridine 5'-triosphosphate
EBP	Erythrocyte-binding protein
E.C.	Enzyme Commission
E. coli	*Escherichia coli*
EDTA	(ethylenedinitrilo) tetraacetic acid
ER	Endoplasmic reticulum
EtBr	Ethidium bromide
FAS	Fatty acid synthesis
FP	ferric-protoporphyrin IX
GMMs	Genetically modified mosquitoes
GTPCHI	Guanosine triphosphate cyclohydrolase I
Hb	Haemoglobin
HF	Halofantrine
HPPK	Hydroxymethylpterin pyrophosphokinase

Hsp	Heat shock protein
IDC	Intraerythrocytic developmental cycle
IRS	Indoor residual spraying
KAHRP	Knob-associated histidine-rich protein
LLITNs	Long-lasting insecticide-treated nets
MCs	Maurer's clefts
MMV	Medicine for malaria venture
MOA	Mode of action
MQ	Mefloquine
mRNA	Messenger ribonucleic acid
MSP	Merozoite surface protein-1
NPP	New permeability pathways
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PBN	Probenecid
PC	Phosphotidyl choline
PCR	Polymerase chain reaction
P. falciparum	*Plasmodium falciparum*
*Pf*CRT	*P. falciparum* CQR transporter
*pf*mdr1	*P. falciparum* multidrug-resistant phenotype
Pgh1	P-glycoprotein
P. pastoris	*Pichia pastoris*
PM	Plasma membrane
PTPS	6-pyruvoyltetrahydrobiopterin synthase
PVM	Parasitophorous vacuolar membrane
QN	Quinine
RBC	Red blood cell
RNA	Ribonucleic acid
RNase	Ribonuclease
RNAi/a	RNA interference/ activation
rNTP	ribonucleotide phosphate
ROS	Reactive oxygen species
RTS,S	Recombinant hybrid molecule expressed in yeast
S. cerevisiae	*Sacchromyces. Cerevisiae*
SDS	Sodium dodecyl sulphate
SHMT	Serine hydroxymethyltransferase
SIT	Sterile insect technique
SP	Sulfadoxine-pyrimethamine

TEMED	N,N,N'N'-tetramethylenediamine
THF	Tetrahydrofolate
tRNA	Transfer ribonucleic acid
TS	Thymidylate synthetase
U	Enzyme units
UTR	Untranslated regions
X-gal	5-bromo-4-chloro-3-indolyl-b-D galactopyranoside
YE	Yeast extract
YT	Yeast-tryptone

Chapter 1

Introduction

“Malaria is a disease of the poor. Time and again, countries have shown that defeating malaria is simply a question of resources. It is relatively easy to prevent. It is easy to detect and cure. There is no need for a quarter of a billion people to become sick each year. There is no reason for 900,000 to die.” (UN Secretary-General, Ban Ki-Moon’s remarks on World Malaria Day 2010). The 25th of April marks World Malaria Day, a day of unified commemoration of the global effort to provide effective control of malaria around the world. The Global Malaria Action Plan is one such effort. It is based on the need to manage existing tools and to improve sustainable interventions for all populations; to eradicate all locally acquired infections; and to maintain ongoing research to create new tools, as well as inform policy and step-up operational implementation of strategies (www.un-ngls.org). Reducing the impact of malaria would significantly propel efforts to achieve the Millennium Development Goals (2000-2015), agreed by every United Nations member state. These include not only the goal of combating the disease itself, but also goals related to women's and children's rights and health, access to education and the reduction of extreme poverty.

“ Malaria defeated the international community many years ago. We cannot allow this to happen again. A single global action plan for malaria control, that enjoys Partnership-wide support, is a strong factor for success."

Margaret Chan, Director-General of the World Health Organization (2008)

"I believe that if you show people a problem, and then you show them the solution, they will be moved to act. The Global Malaria Action Plan lays out an achievable blueprint for fighting malaria – now it's time for the world to take action."

Bill Gates, Co-Chair, Bill & Melinda Gates Foundation (2007)

"The Global Malaria Action Plan aims to ensure that no country is left behind in the global fight against malaria - comprehensive, continent-wide coverage is critical to long-term success."

Tedros Adhanom Ghebreyesus, Minister of Health of Ethiopia (2009)

1.1 Malaria: A liability to the world

Malaria is a serious parasitic infection caused by the unicellular eukaryotic protozoan of the genus *Plasmodium*. Four different species of *Plasmodium* that infect humans exists namely *P. malariae, P. ovale, P. vivax* and *P. falciparum*. *P. ovale* is less prevalent and cause fewer infections whereas *P. vivax, P. malariae* and the most fatal of them all *P.falciparum*, are host-specific and is life-threatening (Hyde 2007). In 2008, Cox-Singh discovered misdiagnosed cases of *P. knowlesi* (predominantly affecting monkey species) for *P. malariae* killing four patients. However, at present *P. knowlesi* cannot be considered as the fifth species infecting humans but the authors suggest that *P. knowlesi* infection be considered as a zoonosis (Galinski and Barnwell 2009). Malaria is the leading cause of morbidity and mortality worldwide (Mehlin *et al.* 2006) and despite methods to control malaria; it remains one of the most devastating endemics. Statistics from the World Health Organization in March 2009 have revealed that out of 250 million malaria cases, nearly 1 million deaths occur annually, the majority of which are children and pregnant women from sub-Saharan Africa. Furthermore, an alarming fact is that a child dies of malaria every 30 seconds, which amounts to 2500 children per day (www.who.int/features/factfiles/malaria). This is due to the lack of acquired immunity in children under 5 years of age making them vulnerable to this disease (Riley *et al.* 2000). Figure 1-1 depicts the graphical map of African people at risk of contracting malaria.

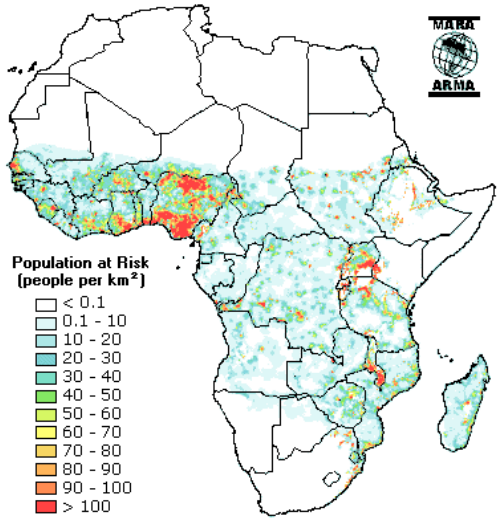


Figure 1-1: People of Africa at risk of contracting malaria. http://www.afro.who.int/malaria/

Many factors determine the severity of malaria contracted. This includes mosquito inoculation rate, the dose of sporozoites, the immunity acquired based on previous exposure to parasite virulence, host genetic polymorphisms, nutrition, housing conditions and immediate access to effective treatment (Marsh *et al.* 1995). However, manifestations of severe malaria can cause the following health risks: anemia, pulmonary edema, jaundice, renal failure and cerebral complications (Delfino *et al.* 2002). Malaria results in global economic losses of more than US$12 billion every year (even though controlling it could be possible at a fraction of the sum) as well as rising health costs (www.rollbackmalaria.org/keyfacts). The expense on health costs including preventative and treatment programs in high malaria prevalence countries amounts to 40% of public health expenditures; 30%-50% of inpatient hospital admissions; and 60% of outpatient health clinic visits. This disease limits communities of poverty-stricken people who cannot afford treatment nor have inadequate access to health (www.who.int/mediacentre/factsheets). The spread of drug-resistant strains, the failure of vector control programs, the rapid growth rate of the parasite, and the lack of a vaccine have further exacerbated the effects on economic development and human health.

1.2 Parasite life-cycle

This apicomplexan parasite has evolved to develop in both a vertebrate and an invertebrate host in a complex 3-way relationship (Figure 1-2). The *Plasmodium* parasite advances through a cycle of sexual, invasive and asexual (replicative) phases. An *Anopheles* mosquito takes a blood meal from an infected human host. As part of the blood meal gametocytes (differentiated parasites) are transferred to the mosquitos mid gut where the male and female gametocytes fuse to form zygotes and within 24 hrs transform to ookinetes and thereafter oocyts (Hayes *et al.* 2006). These oocyts bud (sporogenic cycle) to develop into highly mobile sporozoites that migrate to the salivary glands, ready for attack on a helpless human host. The mosquito transfers these sporozoites, which travel to the human liver to invade hepatocytes (pre-erythrocytic schizogony). Asexual replication occurs to discharge 1000's of merozoites into the bloodstream and this stage is commonly associated with the symptoms of fever. This maturation within the bloodstream is known as the intra-erythrocytic developmental cycle (IDC) and begins with merozoites to form rings, trophozoites and schizonts within 48-72

hrs (depending on the *Plasmodium* specie). Hereafter, schizonts rupture and release merozoites again and reinvasion of erythrocytes occurs permitting this infection cycle to repeat again. However, some parasites differentiate during the IDC into gametocytes, which are ingested by the *Anopheles* mosquito (Hayes *et al.* 2006).

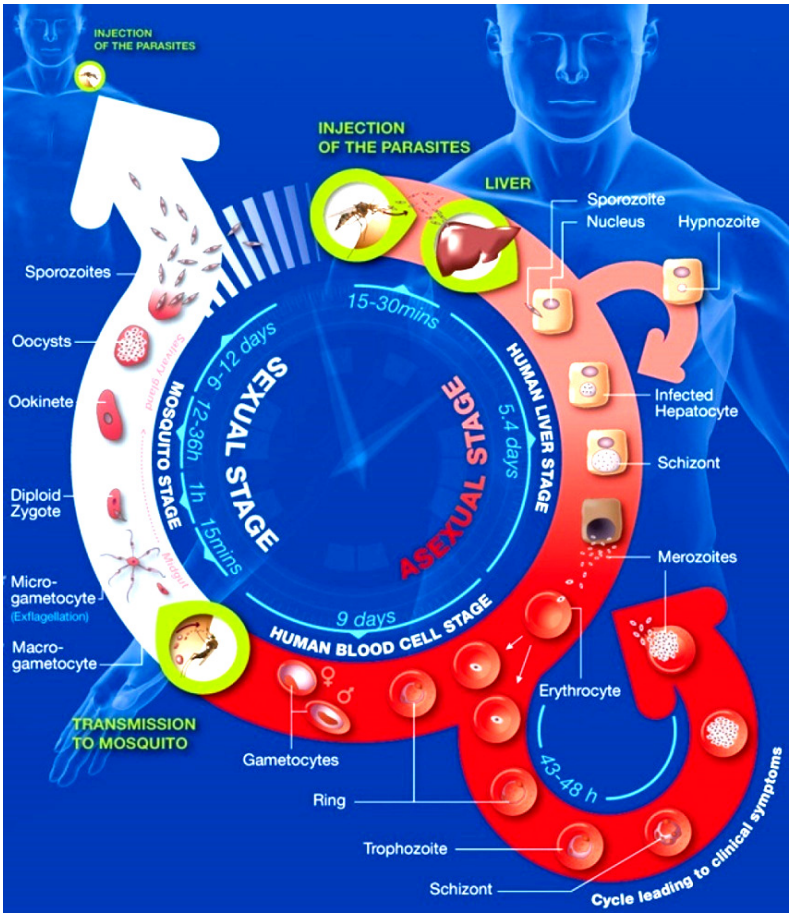


Figure 1-2: The life-cycle of *P. falciparum*. *Anopheles* mosquito takes a blood meal where gametocytes are transferred to its' mid gut where the sporogenic (sexual) cycle occurs. The mosquito transfers these sporozoites, which travel to the human liver where pre-erythrocytic schizogony (asexual cycle) occurs. Schizonts burst from hepatocytes releasing merozoites into the bloodstream and IDC takes place (Wells and Poll 2010).

1.3 Pathogenicity of *Plasmodium*

During erythrocytic schizogony the trophozoite ingests and degrades almost 80% of the host hemoglobin (Hb). As the parasite develops, the host cytoplasm is taken up via the endolysosomal system. The cytosome allows the formation of transport vehicles by invagination of the parasite plasma membrane and parasitophorous vacuolar membrane, which contains the host cytoplasm for delivery to the digestive food vacuole (DFV). Eventually these vesicles fuse with the DFV releasing its contents into the

vacuole to initiate digestion (Hempelmann *et al.* 2003). As digestion continues, proteins extracted from host Hb are also digested by the endolysosomal system. These proteins are digested by metalloproteases, cysteine and aspartic proteases in this system. The peptides are then transported to the parasite cytoplasm and hydrolyzed by cytosolic aminopeptidase. Residual heme in the form of ferriprotoporphyrin IX is produced during globin breakdown. However, this is toxic to the parasite and is therefore converted into non-toxic hemazoin (Banerjee *et al.* 2001).

Every clinical symptom of malaria is linked to host erythrocyte rupture during the blood stage of the *Plasmodium* life- cycle when the merozoites invade and differentiate within the host erythrocyte. Figure 1-3 indicates the morphology of a merozoite from the *Plasmodium* parasite. Invasion involves many complex steps, which is mediated by precise interactions between the parasite ligands and host receptors. Classified under the phylum Apicomplexa, the *Plasmodium* parasite possesses apical membrane bound organelles (micronemes and rhoptries) that function in the invasion of the host cell and contain the parasite ligands. The formation of irreversible junctions between the merozoite and the erythrocyte causes the sequential secretion of these apical organelles into the host erythrocyte. Singh *et al*. (2010) investigated the external signals that trigger the release of the apical organelles. During the exposure of the merozoite to the low potassium concentration (present in the blood plasma), increased levels of cytosolic calcium and thereafter the release of the microneme proteins (erythrocyte binding antigen-175 and apical membrane antigen-1) to the surface of the merozoite occurred. The interaction of the erythrocyte binding antigen-175 with the host glycophorin A receptor triggers the release of the rhoptry proteins upon restoration of the calcium ion levels (Singh *et al.* 2010). As such, the micronemal proteins form the linkage between the parasite and the host erythrocyte while the rhoptries promotes vacuole formation. In order to survive and escape the host immune response, the *Plasmodium* parasite alters the erythrocyte by exporting several effector proteins beyond the enclosing parasitophorous vacuole membrane (Boddey *et al.* 2010).

During the intracellular development of the parasite, parasite proteins such as the virulence proteins (*P. falciparum* erythrocyte membrane protein 1) are secreted into the host erythrocyte (Haldar *et al.* 2007). The *P. falciparum* erythrocyte membrane protein family migrates to knob-like structures on the surface of the erythrocyte to interact with

host vascular adhesions, which cause sequestering in various organs in the host. Additionally, this protein family and the knob-associated histidine-rich proteins have cytoadhesive properties allowing the formation of rosettes since the infected erythrocyte neighbors uninfected erythrocytes. It has been suggested that the invasion and remodeling of the host erythrocyte by the malaria parasite results in devastating effects to the host considering that these events are linked to virulence and harsh pathological implications (Pasternak and Dzikowski 2009).

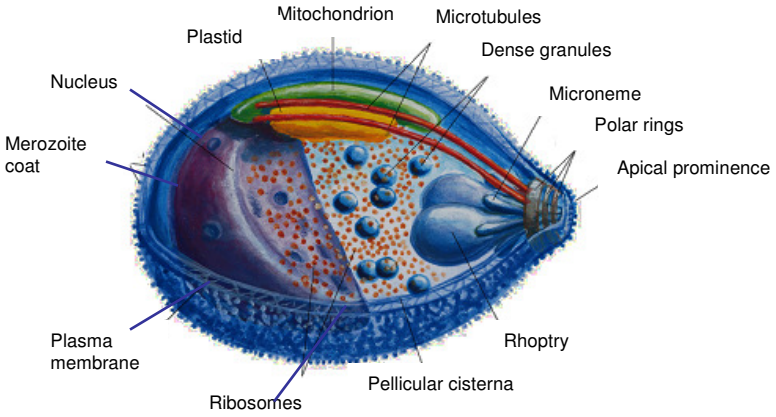


Figure 1-3: The morphological organization of a *P. falciparum* merizoite (Bannister *et al.* 2000).

1.4 Restraining malaria

The focus of restraining this disease is based on providing essential knowledge on malaria prevention, vector control, vaccines, accurate diagnosis and chemotherapy. These methods aim at eradicating the connection between the host (either human or mosquito) and the parasite (Tripathi *et al.* 2005). Each strategy has its restrictions and will be explained in detail in the next sections.

1.4.1 Vector control

Vector control programs are reliant on the use of insecticides and bed nets. In areas at greater risk of malaria infection, vector control programs have been successful in decreasing morbidity and mortality caused by malaria (Gardiner *et al.* 2005). Some techniques used for vector control include long-lasting insecticide-treated nets, which is dependent on the use of repellent or protection against biting or environmental modifications that involves preventing oviposition or larval development by stream clearing (Takken and Knols 2009). The success of insecticide treated bed nets in

conjunction with indoor residual spraying reveals vector control has potential to reduce parasite infections and distribution. One of the first insecticides used in the 1950's and 1960's to eradicate the adult mosquito was bis(4-chlorophenyl)-1,1,1,1-trichlorethane (DDT) (Rogan and Chen 2005). It gained preference for vector control due to its repellent, irritant and toxic actions (Roberts *et al.* 2000). However, in the 1970s the use of DDT was prohibited, since DDT was observed to remain in the environment and possible toxic effects in humans were discovered (Roberts *et al.* 2000; Rogan and Chen 2005). Furthermore, studies conducted by Bornman and colleagues provided compelling evidence of the association of DDT with urogenital birth defects in newborn boys. This is because the environmental exposure of DDT (including its' metabolites) was observed to accumulate in adipose tissue, serum and breast milk of pregnant women. Other studies including *in utero* DDT exposure of animals, were linked to the formation of ovarian tissue, decreased penis size in alligators and other male reproductive disorders such as lowered semen quality in human males (Bornman *et al.* 2009).

Recently, the use of *Bacillus thuringiensis israelensis* for the control of the larval stage showed promise and resistance against this larvicide is not probable since it is composed of 4 endotoxins that cause mosquito death. The control of the adult stages comprises of entomopathogenic fungi (e.g. metarhizium anisopliae), insect-pathogenic viruses (species specific, non-toxic to humans and easily distributable), push–pull systems based on chemical ecology (odour-baited targets attracts mosquitoes), use of genetically modified mosquitoes and sterile insect techniques. Genetically modified mosquitoes carry effector genes that upon expression render the mosquito useless to *Plasmodium* (Takken and Knols 2009). Furthermore, this disease can be limited by reducing mosquito transmission of the *Plasmodium* parasite to the human host. The genome sequence of the vector for the parasite, *Anopheles gambiae* was reported in by Holt *et a*l. (2002). The identification of odor receptors in conjunction with studies using the orthologous *Drosophila melanogaster* model allowed for the discovery of a repellent *N,N*-diethyl-meta-toluamide, an inhibitor of insect odor receptors (Ditzen *et al.* 2008).

The ultimate purpose would be to prevent malaria transmission by replacing the natural mosquito population. The control of this epidemic can be established by combining several tools that complement each other based on the vector specie, the environment, phenology and epidemiology (Takken and Knols 2009).

1.4.2 Vaccines

Vaccine discovery aims at treating several stages of the parasites' life cycle. Many different antigens were identified at the pre-erythrocytic stages (sporozoite invasion of hepatocytes), the hepatic stages and merozoite invasion of erythrocytes, the asexual stages as well as the sexual stages of development. The most widely studied vaccine for the treatment of malaria is RTS,S (recombinant hybrid molecule expressed in yeast), which incorporates the partial sequence of circumsporozoite protein (CSP, a protein found on the sporozoite surface), a central tandem repeat, and carboxyl-terminal regions that are fused to the N-terminus of the S-antigen of hepatitis B virus (HBs Ag) in a particle that includes the unfused S-antigen. It is administered in combination with immunostimulants, monophosphoryl lipid A (MPL) and *Quillaja saponaria* fraction 21 (QS21). The effect that RTS,S had on *P. falciparum* parasitemia was sterile protection or the delay in the onset of parasitemia. This vaccine has reached phase III trials after 40 years of ongoing research (Waitumbi *et al.* 2009). Furthermore, the vaccine was 50% safe and effective for infants (Abdulla *et al.* 2008).

A replication defective viral vaccine vector that expressed the malaria antigen merozoite surface protein-1 (MSP-1) is another promising vaccine candidate (Draper *et al.* 2008). During invasion, this large polypeptide undergoes cleavage to a 33 kDa (MSP-133) and 19 kDa (MSP-119) fragments (Holder *et al.* 1988). These viral vaccines are employed to induce a high T-cell response to target intracellular pathogens (Draper *et al.* 2009). Other antigens that are targeted in the hepatic stage of the parasite life-cycle include the liver-stage antigen 1,3 (LSA-1,3), the exported antigen (EXP-1) and those aimed at eliminating the asexual blood stages including the ring-infected erythrocyte surface antigen (RESA) and PfEMP1 on the trophozoites and schizonts, respectively (Richie and Saul 2002; Greenwood *et al.* 2005). Furthermore, the sexual developmental stages in the mosquito (macrogametes and ookinetes) possess *Pf*s 25, 28, 48 and *Pf* 45 antigens, respectively. Current vaccine candidates are based on a few proteins that were initially described before the *P. falciparum* genome project was investigated. Of these proteins, there are about 5,400 protein-coding genes that are expressed in an intricately, stage-specific manner (Holder 2009). As such, the need for new developments in novel vaccine candidates is crucial. Holder (2009) queried whether a vaccine could mature the immune system of a naïve individual to one protected from disease. This led to the proposition of a transmission-blocking vaccine aimed at the sexual stages solely or

sexual- and pre-erythrocytic stages. This shows promise since there is a lack of naturally acquired immunity to these stages. Therefore, combinations of vaccines that are aimed at different life stages of the parasite would require investigation for eradicating this problem (Holder 2009).

1.4.3 Diagnostics

The diagnosis of malaria is based on the identification of the malaria parasite morphologically or the presence of antigens (or its' derivatives) in the patients blood. Without the proper diagnosis of malaria, most antimalarial treatment gets administered to those who do not have the disease but have fever symptoms. Other reasons for requiring improved diagnostics include financial implications (economic considerations of misdiagnosis or over diagnosis); preservation of drug supplies as well as drug efficacy and to examine trends in the incidence of malaria (Amexo *et al.* 2004) to name a few. Furthermore, the efficacy of diagnosis is dependent on the different morphological forms (the different blood-stages) of malaria species; drug resistance, sequestering of the parasite in hidden tissues and falsely low parasitemia. As such, there is a necessity for improved diagnostics but also characterized and effective diagnostics that take into account the morphological variation of the parasite. These considerations are essential for clinical diagnosis in the use of chemoprophylaxis and treatment of this disease (Tangpukdee *et al.* 2009).

Different modes of diagnosis are available. Clinical diagnosis is a judgement made on the basis of symptoms and physical findings (fever, headache, diarrhoea, vomiting, chills, dizziness, weakness, abdominal pain, anorexia and myalgia) of patients made by medical doctors. This approach is non-specific, variable and expensive (Looareesuwan *et al.* 1999). The non-specificity of clinical diagnosis suggests that symptoms of other diseases overlapping those of malaria impairs diagnostics, which compromises the quality of patient health care and unnecessary use of antimalarial drugs (Tangpukdee *et al.* 2009). Parasitological techniques used for diagnosis includes either staining the *Plasmodium* parasite thick or thin blood films with Giemsa or staining the nuclei of the parasite with a DNA-specific dye (acridine orange). However, both are disadvantageous due to low sensitivity and high cost, respectively. Other techniques have been developed, since the WHO requested for novel, rapid, sensitive, specific and inexpensive diagnosis. Polymerase chain reaction, loop-mediated isothermal

amplification, flow cytometry assay, microarray, mass spectrophotometry. Lateral flow techniques and automated blood cell counters have been developed as techniques to better diagnosis of malaria. However, each technique has constraints and is impractical for routine clinical diagnosis (Moody 2002; Tangpukdee *et al.* 2009).

1.4.4 Malaria control via chemotherapy

Chemotherapy is described as the act of destroying rapidly dividing cells by chemical methods. Antimalarial drugs are either administered for the prevention of malaria i.e. prophylactic drugs and those administered to infected individuals i.e. therapeutic drugs (Dhanawat *et al.* 2009). At present antimalarial treatment is based on three types of drugs that include: quinolines, antifolates, artemisinin and derivatives of these. Antifolates and quinoline-derived compounds are the most frequently used but the intravenous usage are restricted by resistant strains worldwide (Martinelli *et al.* 2008a).

The majority of antimalarial agents have been found to act on the IDC of *Plasmodium* (Figure 1-4) (Frederich *et al.* 2002). However, antimalarial drugs can be further categorized based on the subcellular location in the parasite targeted by the specific drug. These drugs and their specific targets include the quinolines to target the lysosomal food vacuole, atovaquone to target the mitochondrion, doxycycline to target the apicoplast and antifolates to target the cytosol (Ridley 2002).

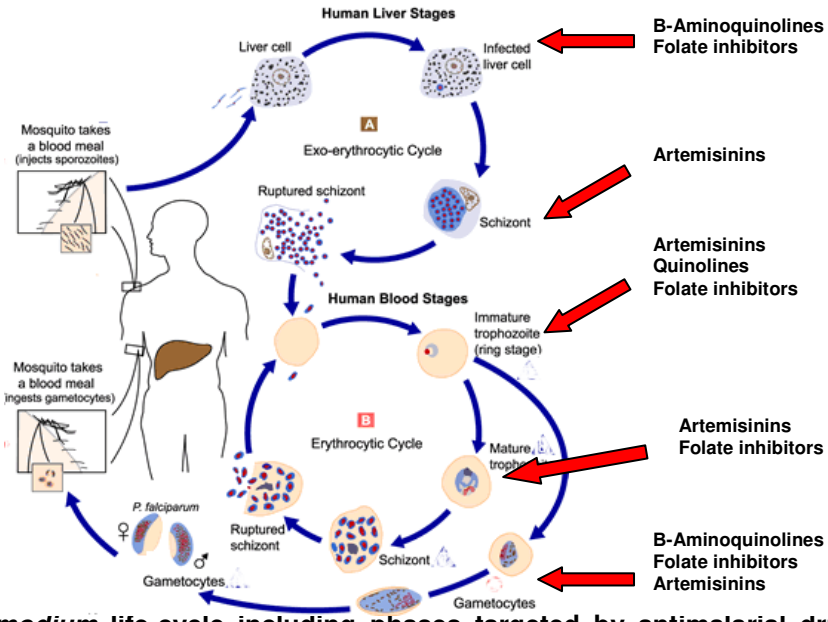


Figure 1-4*:* The *Plasmodium* life-cycle including phases targeted by antimalarial drugs (adapted from pathmicro.med.sc.edu/parasitology/blood-proto).

It is important to remember that antimalarial drugs used on large populations, enhance the adaptation of the parasite against their action forcing the need for alternate drug classes (Mordueller 2010). Various approaches aim to combat malaria and include the optimization of currently available drug therapies by formulating analogues (e.g. a trioxalane derivative, which relates to artemisinin), utilizing natural compounds, discovering resistance-reversal agents, using synergistic or additive drug combinations and manipulating drugs previously used to treat other diseases. In order to develop inhibitors, which act specifically on parasite proteins several biochemical pathways are exploited. Furthermore, a relationship between drug targets and the functions of distinct organelles in the parasite does exist (Sahu *et al.* 2008)

1.4.4.1 Drug combinations

With the progress of antimalarial drug resistance came the desperate need to develop novel techniques, which could be used to combat this deadly disease. As such, the use of combination therapies of antimalarial drugs has superseded monotherapy drug use (Martinelli *et al.* 2008a). Combination therapy makes up the basis of drug synergy and is defined as the combinations of antimalarials that act by combining the effect of different antimalarial drugs on various metabolic pathways (Gregson and Plowe 2005). Table 1-1 refers to the various antimalarial drug combinations available, including the advantages and disadvantages of their use (Martinelli *et al.* 2008a). Various advantages exist for synergistic drug use, which include increased efficacy, a reduction in selection for resistance development and the overall half-life is enhanced (Choi *et al.* 2008).

Artemisinin-based combination therapies (ACT) were established as the major therapeutic treatment for malaria. However, combinations of artemisinin with current antifolates and quinolines make treatment dependent on artemisinin in regions where resistance against these drugs is widespread (resistance against these antimalarial drugs will be focused upon later). It may not be too far in the near future when resistance against artemisinin becomes prevalent (Martinelli *et al.* 2008a). Besides resistance to pyrimethamine-sulfadoxine very little is known about the resistance to other antifolate drug combinations (Gregson and Plowe 2005).

Table 1-1: Currently used antimalarial drug combinations (Martinelli *et al.* 2008a; Wells and Poll 2010)

Drug Combination	Advantages	Disadvantages
Sulfadoxine+ Pyrimethamine (SP)	Synergistic activity, inexpensive, one-dose cure and can be used during pregnancy	Wide-spread resistance, medium-long half lives, moderate pharmacokinetic mismatch
Chlorproguanil+ Dapsone	Synergistic activity, good pharmacokinetic match and short half-lives	Cross-resistance with SP
SP+ Amodiaquine	Effective against certain chloroquine-resistant strains of *P. falciparum.* Similar pharmacokinetic profiles	Spreading of resistance to both components and medium-long half-lives
SP+Artesunate	Efficacious, with similar pharmacokinetic profiles	Efficacy dependent on level of SP resistance
Proguanil/ Dapsone+ Artesunate	Efficacious, similar pharmacokinetic profiles and short half-lives	Cross-resistance with SP may limit its application
Artesunate+ Mefloquine	Efficacious against all malaria species. Long half-life (21 days) for mefloquine	Pharmacokinetic mismatch, resistance to mefloquine spreading and high cost
Artesunate+ Amodiaquine	Efficacious	Pharmacokinetic mismatch, and spreading of resistance to amodiaquine
Artesunate+ Fosmidomycin	Efficacious, excellent pharmacokinetic match and short half-lives	Not suitable for prophylaxis
Artesunate+ Clindamycin	Efficacious, similar pharmacokinetic profiles and short half-lives	Not suitable for prophylaxis, treatment failure observed
Proguanil+ Atovaquone	Synergistic activity, efficacious, active against all *Plasmodium* species and Inhibits pre-erythrocytic parasite development in the human liver and oocyst development in the mosquito	Moderate pharmacokinetic mismatch, resistance easily acquired and high cost

The half-life of inhibitors may affect drug efficacy and several hypotheses were made. Firstly, *in vitro* studies have revealed that higher concentrations of the drug result in a quicker route to drug resistance regardless of the time required for exposure (Gregson and Plowe 2005). Secondly, shorter periods of time for the presence of the drug in the bloodstream (clearance) results in lowered resistance. An example of such a combination includes chloroproguanil-dapsone, which is less susceptible to resistance (Gregson and Plowe 2005). Nevertheless, the WHO in 1966 stated that the combined use of drugs with different modes of action might prevent the development of resistance. However, several combinations of drugs failed as antimalarial therapies. Drugs with complementary activities show promise. An example includes Chloroquine and Primaquine for schizont and gametocyte killing activity, dihydrofolate reductase/dihydroopterate synthase (DHFR/DHPS) inhibitors, dihydroorotate inhibitors and shikimic acid pathway inhibitors. This is an example of the suppression of many

steps along a metabolic pathway leading to the same product (Gregson and Plowe 2005).

The medicine for malaria venture (MMV) proposed an innovative era of drug discovery in 2008 not only for the treatment of malaria but for eradication too (Table 1-2).

Table 1-2: Currently developed commercial antimalarial drugs and targets. (MMV report, 2008, (Martinelli *et al.* 2008a; Wells and Poll 2010).

Commercial name	**Combination**	**Stage**	**To treat**	**Advantages/ disadvantages**
Coartem ®	Artemether-lumefantrine	Approved: WHO prequalified, Comprehensive safety data, Pediatric formulation with a sweet taste, First-line therapy in most of Anglophone sub-Saharan Africa	Uncomplicated *P. falciparum* or unidentified malaria	Easy and cost-effective to manufacture some cases of resistant strains, half-life shorter than other ACT's
Pyramax ®	4-aminoquinoline pyronaridine (artesunate)	Phase III: Pediatric formulation still in progress- for launch in 2012	Uncomplicated *P. falciparum* and *P. vivax* (blood stages)	Cost-effective, long half-life of pyronaridine
Eurartesim ®	Dihydroartemisinin-piperaquine (a 4-aminoquinoline)	Phase III: On the WHO treatment guidelines, but no product approved by WHO prequalification., Pediatric formulation still in progress- for 2012 launch	Uncomplicated *P. falciparum*	Piperaquine has an extremely long half-life, long post-treatment prophylactic effect
Dacart™	Chlorproguanil-dapsone-artesunate	Phase III — Discontinued	Uncomplicated *P. falciparum*	Loss of haematocrit
Intravenous artesunate	-	Phase II	Severe malaria	-
Artemisone	N10-subsituted artemisinin derivative	Phase II	-	-
Tafenoquine	8-aminoquinoline	Phase I	Dormant liver stages of *P. vivax*	-
GSK 369796	N-tertiary butyl Isoquine	Phase I	-	-
GSK 932121A	4-pyridone class	-	-	Blocking the electron transport chain in the parasite mitochondrion
Mefloquine	Diastereoisomers of mefloquine	No plans for pediatric formulation	Uncomplicated *P. falciparum* malaria and chloroquine resistant *P. vivax* (blood stage)	Long half-life. High cost

Mirincamycin lincomycin (similar to clindamycin), BCX 4945 (purine nucleoside phosphorylase inhibitors, immucillins) and OZ 439 (synthetic peroxides) also form part of

commercial drugs but are not currently developed and are still under trial for use as antimalarial drugs (MMV report, 2008).

As such, it is imperative that novel drugs/targets are developed or that the available antimalarial drugs are optimized due to antimalarial drug resistance (Foley and Tilley 1998). However, understanding the mechanism of resistance of *Plasmodium* to antimalarial drug targets is important for designing other drugs suitable for these targets. Assessing the mutational rates of genes are essential for testing new antimalarial inhibitor effects or combinations, mathematical modeling to determine the time course of the increase in drug resistance and to anticipate the spread of drug resistance within a population (Hyde 2008). Thereafter, antimalarial drugs must meet certain criteria in order to be successful as treatment. They should decrease mortality and complication rate, have a good oral bioavailability, be stable at room temperature, be effective with minimal amounts of dosing and within the shortest period of time and most importantly be cost-effective (Plouffe *et al.* 2008).

1.4.4.2 Quinoline

Quinoline-containing drugs and artemisinin-based drugs are categorized as blood schizonticides since they act on the intra-erythrocytic stages of parasites and their major target is predicted to be the food vacuole (FV). Quinoline derivatives are classified as type-1 (4-aminoquinolines; Chloroquine) and type-2 drugs (arylamino alcohols; Quinine and Quinidine, Mefloquine, Halofantrine). Type-1 drugs have weak basic properties, are diprotonated and hydrophilic at a pH of 7, which is in contrast to the type-2 drugs that are much weaker bases and are hydrophobic at a pH of 7. The two groups differ in the interaction with their target and have an inverse relationship with reference to dose sensitivities (Olliaro 2001).

1.4.4.2.1 Chloroquine

Chloroquine (CQ) is an antimalarial agent that is active only against the blood stages of the *Plasmodium* life cycle when the parasite is continuously degrading Hb (Peters 1970). CQ is a weak diprotic base that is a lysosomotropic (an agent that selectively enters the lysosomes of certain cell types). At a neutral pH, in its unprotonated form CQ is capable of diffusing through the FV (pH ~5) membrane within the parasite infected erythrocytes where it gets trapped once it becomes protonated leading to the generation of a

concentration gradient (Martinelli *et al.* 2008a). During the trophozoite life-stage, the parasite is capable of digesting Hb in the FV. CQ however, prevents haem polymerisation, a process crucial for the removal of cytotoxic haem. Furthermore, since the quantity of haem left in the lysosome does not seem sufficient for killing of the parasite, it is believed that the action of CQ is due to the formation of a haem-CQ complex. This complex diffuses out to the cytosol where the complex is degraded by reduced glutathione (Martinelli *et al.* 2008a).

In an attempt to spread the use of CQ as a prophylactic, CQ was added to table salts in certain areas of the world. It was thereafter believed that due to the salt programme, drug resistant strains of the parasite emerged in the late 1980s (Dhanawat *et al.* 2009). The CQ resistant (CQR) strain results due to lowered CQ uptake in the FV. Based on transfection studies, *P. falciparum* CQR transporter (*Pf*CRT) is associated with a mutation in K76T. In these studies using *Xenopus laevis* (*X. Laevis*) oocytes expressing CRT, a loss in positive charge occurred as a result of the mutation. This in turn allows the drug to traverse the drug-metabolite transporter back into the cytoplasm (Fidock *et al.* 2008). *Pf*CRT is characterized as an integral membrane protein with transmembrane domains, situated on the FV membrane of the intra-erythrocytic malaria parasite where CQ has its effect. As such, *Pf*CRT mediates the transport of protonated CQ through the FV membrane conferring resistance (Martinelli *et al.* 2008a).. Further investigations revealed a dependence on the strain whether the mutant *pfcrt* conferred resistance or tolerance. This implied other factors contributed to resistance (Fidock *et al.* 2008).

Recently the transport activity of *Pf*CRT on CQR and CQ sensitive (CQS) strains were investigated at the surface of *X. laevis* oocytes. The K76T mutation is accompanied by a number of downstream mutations, making this a polymorphic gene (a total of 15 variant residues are known to exist). These mutations not only aid to compensate for the loss of function associated with K76T but may also be linked to conferring resistance to other related antimalarial drugs (Bray *et al.* 2005). One such mutation is S163R, the addition of either this mutation or K76T to *Pf*CRT^{CQR} resulted in loss of CQ transport activity. However, in contrast the addition of the T76K mutation to *Pf*CRT^{CQS} resulted in no significant increase in CQ uptake. This investigation confirmed that K76T is required but insufficient for CQ transport via *Pf*CRT and that other mutations act synergistically to confer CQR. Furthermore, the success of an expression system for *Pf*CRT has potential

to facilitate the discovery of novel drugs similar to CQ to bypass the resistance (Martin *et al.* 2009).

1.4.4.2.2 Quinine methanols: Mefloquine, Quinine and Halofantrine

These antimalarial drugs are monoprotic bases, which are therefore presumed to accumulate in the acidic parasitophorous vacuole (with lower efficiency than CQ). Mefloquine (MQ) and Quinine (QN) have lipophilic properties allowing them to interact with peptides on the surface of *P. falciparum*-infected erythrocytes, indicating these peptides either function in drug-uptake or signify direct targets of action (Martinelli *et al.* 2008a). Similarly to CQ these drugs act in the intra-erythrocytic *Plasmodium* life stages. MQ causes morphological alterations to the FV due to the degranulation of hemozoin rather than clumping as observed with CQ (Olliaro 2001). Recently, it has been proposed that MQ inhibits endocytosis of nutrients into the parasite FV (Martinelli *et al.* 2008a). However, the use of MQ as a therapeutic against malaria is limited by high cost and the appearance of detrimental neuropsychiatric effects. It is believed that both MQ and Halofantrine (HF) share a similar mode of action. HF is used for the treatment of CQ resistant and multi-drug resistant, uncomplicated malaria (Dhanawat *et al.* 2009).

An amplification of the *P. falciparum* multidrug-resistant phenotype (*pfmdr1*) and overexpression of its product (pfgh1) has been associated with resistance towards MQ, QN and HF (Cowman *et al.* 1994). Evidence suggests that single nucleotide polymorphisms within *pfmdr*1 gene may control the parasites responsiveness to QN, MQ and HF. Transfection studies have revealed that point mutations in the nucleotide sequence of *pfmdr*1, which encode the polymorphisms cause the parasite to accumulate greater levels of MQ and HF. Furthermore, an amino acid mutation that lies in the transmembrane domain has been proposed to alter the substrate specificity of the pump (Martinelli *et al.* 2008a).

1.4.4.3 Artemisinin derivatives

Artemisinin (AS) was isolated in the early 1970's from *Artemisia annua*. Various analogues of AS were synthesized to improve its solubility and pharmacokinetics. Of the analogues the most commonly used in antimalarial drug therapy, include artesunate, dihydroartemisinin, artether and artemether. AS and its analogues are sesquiterpene trioxane lactones consisting of a peroxide bridge that conveys the antimalarial activity

(Martinelli *et al.* 2008a). The use of AS-based combination therapies were developed since AS alone is fast acting however, by combining it with an antimalarial drug that has slower activity the efficacy is improved. Examples of such combinations include artemether-lumefantrine, artesunate-amodiaquine, artesunate-mefloquine, and dihydroartemisinin- piperaquine (Laufer 2009).

The mode of action of AS is uncertain. However, the formation of free carbon-centered radicals has been linked to AS activation. It is proposed that cleavage of the endoperoxide bridge by free iron leads to the formation of these radicals resulting in oxidative stress and a decrease in reduced glutathione levels. Additionally, AS has been shown to interfere with the mitochondrial electron transport and inhibit the respiratory chain in *P. falciparum* (Martinelli *et al.* 2008a). Certain proteins are hypothesized to be targets of AS. These include cysteine proteases, electron transport chain proteins, translationally controlled tumor protein and PfATP6 (SERCA-type Ca^{2+}ATPase) (Fidock *et al.* 2008). Increased tolerance to AS has been associated with a point mutation in PfATP6 but additional mutations in other genes have been proposed to cause AS resistance. Recent studies done on *in vivo* rodent malaria models have stated that AS resistance was possible and the analysis of the genes involved include a counterpart of *pfcrt*, *pfmdr1* and others (Martinelli *et al.* 2008a).

1.4.4.4 Atovaquone (naphtoquinone)

Atovaquone is a structural analogue of ubiquinone in the electron transport chain believed to disrupt electron transfer (cytochrome bc_1) during the oxidation of dihydroorotate to orotate and the loss of mitochondrial membrane potential (Dhanawat *et al.* 2009). This reaction connects pyrimidine *de novo* synthesis and the mitochondrial electron transport chain leading to the disruption of protein transport (Hyde 2002). This hydroxynaphtoquinone used in combination with proguanil interferes with the mitochondrial electron transfer (Olliaro 2001). Atovaquone resistance *in vitro* has been associated with mutations that arise in the cytochrome complex localized in the apicoplast genome (Dhanawat *et al.* 2009). These mutations occur at the binding site resulting in a reduced drug efficacy (Hyde 2002). *In vitro* studies indicated the first point mutation originated at position 133 (M133I) and as the concentration of AS increased, additional mutations at positions 283 and 284 arise (Witkowski *et al.* 2009).

1.4.4.5 Antifolates

Antifolates are known to act on the asexual and gametocytic life-cycle stages of the malaria parasite. This occurs as a result of inhibiting enzymes of the folate pathway and therefore hampering the production of DNA and the amino acids serine and methionine. They can be categorized into two subclasses based on the inhibition of DHPS (class I antifolates, sulfones and sulfonamides) and DHFR (class II antifolates, pyrimethamine, biguanides and quinazolines). Type I antifolates inhibit the formation of dihydropteroate by mimicking pABA and thereby competing for the active site of DHPS. Type II antifolates act on DHFR therefore preventing the reduction of dihydrofolate (DHF) to tetrahydrofolate (THF) (Olliaro 2001).

Increased resistance to antifolate drugs such as pyrimethamine and sulfa drugs (both normally used in combination) has been observed in *P. falciparum*. Despite the reduced effectiveness of the drug, sulfadoxine-pyrimethamine (SP) it is still used instead of combinations of artemisinin or its derivatives known as ACT due to lowered costs. Africa is known to harbour the most lethal SP resistant strain. The genotypes are the triply mutant versions of *dhfr*, Ser108Asn, Asn51Ile and Cys59Arg and the doubly mutant version of *dhps*, Ala37Gly and Lys540Glu. However, in the 1980s a quadruply mutant version of *dhfr* arose with an additional mutation at Ile64Leu (Hyde 2008). X-ray crystallographic studies of the doubly mutant (Ser108Asn Cys59Arg) *P. falciparum* DHFR-TS conjugated to pyrimethamine revealed the side chain of Asn108 sterically hinders the *para*-Cl of pyrimethamine, thus reducing its affinity (Japrung *et al.* 2007). DHFR was purified from highly resistant and moderately resistant parasite strains. The parasite strains had 500- and 15-fold less affinity, respectively for pyrimethamine and sulfadoxine than the sensitive strains (Gregson and Plowe 2005). Furthermore, the quadruply mutant DHFR-TS bound to WR99210 (an analog of cycloguanil) revealed that the flexibility of the side chain prevents steric hindrance with Asn108 (Japrung *et al.* 2007). These studies provide an explanation of sequence-specific mutational accumulation and frequently occurring alleles based on the combined effects of drug resistance and enzyme kinetics. Since most folate inhibitors were pterine derivatives the length of the pteridine ring fits in the active site of the enzyme between residues 108 and 54 (Gregson and Plowe 2005). The question that remains is whether the use of antifolates will be abolished due to this quadruply mutant *dhfr* form. However, it was established by *in vitro* kinetic studies that this mutant form of *dhfr* is a highly

compromised enzyme and that parasites dependent on this enzyme, will experience a decrease in fitness (Hyde 2008).

1.5 Advances in antimalarial targets:

Yuthavong (2002) described various approaches to overcoming antifolate drug resistance however, this can be applicable to all antimalarial drugs to which resistance has emerged. Two approaches to overcoming problems associated with antimalarial chemotherapy include the development of novel antimalarial drugs, which can be used effectively against the resistant *Plasmodium* strains and/or to utilize other novel antimalarial drug targets. The former approach is reliant on the assumption that the mutation responsible for resistance against a specific inhibitor, is caused by decreased binding strength. However, the binding strength can be increased by structural modifications of the existing drugs. For this the knowledge of the wild-type and mutant drug targets (e.g. DHFRs) are essential. The other approach relies on identifying another target located within the same pathway of the drug target to which resistance exists or other pathways (Yuthavong 2002).

1.5.1 New permeability pathways

It is well known that the intra-erythrocytic parasite modifies the host erythrocyte transport pathways. Polyspecific, anion-selective new permeability pathway (NPP) is one such modification believed to increase the permeability of the erythrocytes to a variety of low-molecular weight, charged molecules. This allows the influx of essential nutrients and the efflux of waste products (lactic acid). This pathway can be exploited by inhibiting transport to deprive the parasite of its essential nutrients or developing cytotoxic drugs, which enter the parasite particularly through the permeability pathway (Sahu *et al.* 2008).

1.5.2 Mitochondrion

During the asexual blood-stage the parasite mitochondrion is immature and lacks cristae of the inner mitochondrial membrane (Figure 1.5). The *Plasmodium* mitochondrion possesses an active electron transport chain, enzymes required for coenzyme Q, iron-sulphur cluster biosynthesis and functions in pyrimidine biosynthesis making it essential for parasite survival and is therefore effective as an antimalarial chemotherapeutic target (van Dooren *et al.* 2006). The *Plasmodium* ubiquinone (functioning in the respiratory chain) is different to the mammalian homologue. Atovaquone was developed to target

the parasite mitochondrion since it is effective against ubiquinol-cytochrome oxidoreductase (complex lll), allowing it to suppresses the electron transport- chain and collapse the membrane potential (Sahu *et al.* 2008). Recently, the respiratory chain type II NADH dehydrogenase (complex I) was recombinantly expressed to evaluate substrate and inhibitor specificity (Dong *et al.* 2009).

1.5.3 Apicoplast

A chloroplast-like organelle of apicomplexan parasites known as the apicoplast (Figure 1.5) formed as a result of endosymbiosis. Parasite-specific pathways in the apicoplast make excellent targets since this organelle is specific to the parasite (absent in the host). Additionally, this organelle maintains parasite growth by functioning in the metabolism of fatty acid, heme and amino acids (Sahu *et al.* 2008).

It is known the *Plasmodium* parasite has a high glycolytic flux, which allows the parasite to undergo rapid cell division resulting in increased parasitemia within days. Furthermore, it is known that the *Plasmodium* infected RBC's are capable of consuming 75 times more glucose than uninfected RBC's (Sherman 1979). During glycolysis, the conversion of dihydroxyacetone-phosphate or glyceraldehyde-3-phosphate results in the formation of the toxic by-product, methylglyoxal. Methylglyoxal causes the formation of modified nucleic acids and inactivated proteins. However, the ubiquitous glyoxalase system in *Plasmodium* removes it. This system consists of the isomerase glyoxalase I and two isozymes of the thioester hydrolase glyoxalase II (cGloII and tGloII) that form part of the glyoxalase system. tGloII was identified in the apicoplast pointing to alternative metabolic pathways. However, no reports of inhibition studies have been conducted on tGloII (Urscher *et al.* 2010).

1.5.4 Food vacuole

The cleavage of host proteins by the parasites is essential during invasion of the host erythrocytes (Figure 1-5). The malaria parasite degrades 65-75% of host Hb, which releases free amino acids required for parasite survival. These amino acids are used for protein synthesis and maintaining the osmotic pressure within the RBC. The only amino acid that is not found in human nor synthesized by the parasite is isoleucine. The uptake of isoleucine is mediated by a process involving an ATP-independent transporter (exchanging intracellular leucine for isoleucine) and the ATP-dependent storage or

accumulation of isoleucine. The neutral aminopeptidases (catalyzes the release of neutral amino acids from the N-terminal), M1 alanyl aminopeptidase (PfM1AAP) and M17 leucine aminopeptidase (PfM17LAP) are suggested to play a crucial role in this process. As such, studies currently are underway to define the active site, which could aid in identifying specific inhibitors (Adams *et al.* 2009).

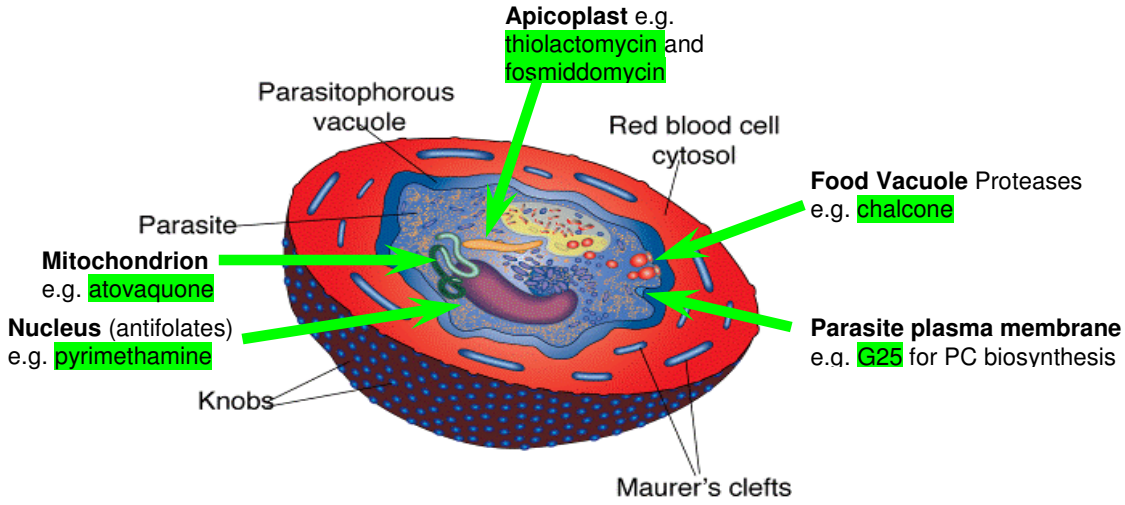


Figure 1-5: A representation of the structure of a parasite-infected red blood cell and novel antimalarial drug targets in *Plasmodium* parasite (adapted from www.latrobe.edu.au/biochemistry/lab/maier/index).

1.5.5 Microtubulins

In vivo and *in vitro Plasmodium* development (particularly asexual intraerythrocytic stage), growth and proliferation was severely compromised by inhibitors of tubulins and microtubulins (Bell 1998). These inhibitors do not possess the selective toxicity required to combat the malaria parasite and were previously used as anticancer drugs (Sahu *et al.* 2008). An antimitotic herbicide, trifluralin was assessed in comparison to vinblastine to assess its activity and uptake. The uptake of trifluralin was immediate, non-saturable and readily reversible. However, due to its hydrophobic characteristics accumulation occurs in the parasite membrane, decreasing the levels in the soluble fraction and restricting access to the target. Vinblastine on the contrary accumulates in the soluble fraction, uptake was saturable and mostly irreversible allowing access to the target. Therefore, a more hydrophilic derivative of trifuralin should be synthesized (Naughton *et al.* 2008).

1.5.6 Oxidative stress

Plasmodium contains antioxidant enzymes required to prevent oxidative stress. Superoxide dismutase, glutathione peroxidase, thioredoxin reductase and catalase are essential for action against damages like nucleic acid modifications, lipid peroxidation and oxidation of thiol containing proteins and reactive oxygen species (ROS) produced by the host immune system, erythrocytes or possibly by the parasite itself. Therefore oxidative stress can be an exceptional target for antiparasitic chemotherapy (Sahu *et al.* 2008; Banerjee *et al.* 2009). Studies revealed that the *P. falciparum* thioredoxin reductase bears dissimilarity to the host enzyme. Inhibition of this enzyme will most likely affect the parasite causing enhanced oxidative stress, ineffective DNA synthesis and cell division and redox regulatory processes. Since the crystal structure of the enzyme is unavailable, homology modeling was used to determine the three-dimensional structure using a thioredoxin reductase type 2 of mouse as a template. This would aid in rational drug design of this enzyme as a drug target (Banerjee *et al.* 2009). Additionally, gluthathione S-transferase is present in large quantities in *Plasmodium* and differs from the human host making it a good antimalarial target. It allows for the conjugation of gluthathione with hydrophobic compounds, which are eliminated (Sahu *et al.* 2008)

1.5.7 Polyamine biosynthesis

Polyamines include the diamines putrescine, cadaverine, spermidine and spermine. These polycationic metabolites influence apoptosis, embryonic development, cell differentiation and proliferation. Polyamine metabolism is considered as a drug target, which can be targeted by inhibiting intracellular polyamine biosynthesis or by the displacement of the active polyamine store (inside cells) with polyamine analogues that are non-functional (Muller *et al.* 2008). Within the malaria parasite, polyamines make up 14% of its metabolome and the human host constitutes low levels thereof (Olszewski *et al.* 2009). The bifunctional enzyme ornithine decarboxylase- S-adenosylmethionine decarboxylase (AdoMetDC- ODC) functions in the formation of putrescine and spermidine, respectively. The *P. falciparum* ODC is a pyridoxal 5-phosphate (vitB6) dependent enzyme. Studies conducted by Muller *et al* (2009), involved assessing the inhibitory effect of novel pyridoxal-adducts on *P. falciparum* ODC. The results indicated that the cyclic pyridoxyl-tryptophan derivative enters the malaria parasite, presumably through an amino acid transporter and displaces pyridoxal 5-phosphate from the ODC

active site resulting in growth arrest. Further investigations will be required to reveal whether other *P. falciparum* enzymes are also targeted by these inhibitors (Muller *et al.* 2009).

1.6 Nucleic acid metabolism

The presence of DNA in the malaria parasite was verified in the 1940's using Feuglen staining however, further metabolic studies were initiated only five years after its discovery. It was found that *P. lophurae* (avian malaria specie) is capable of synthesizing DNA on its own and is not dependent on the host cell nucleus or erythrocytes since the latter are anucleate and require low levels of nucleotides (Sherman 1979). Studies of the resultant effects that antimalarial drugs, pyrimethamine and proguanil had upon DNA synthesis revealed that by inhibiting DHFR, THF levels would decrease therefore inhibiting thymidine and DNA synthesis of *P. gallinaceum* (Schellen.K and Coatney 1961). It is evident from the high proliferation rate of dividing cells and pathogens that a continuous and sufficient supply of nucleotides is imperative. For the replication process certain nucleotides classified as purine deoxyribonucleotide 5'-triphosphates (dATP and dGTP) and pyrimidine deoxyribonucleotide 5'-triphosphates (dCTP and dTTP) are required (Sherman 2009). As crucial as they already are for nucleic biosynthesis, they are also essential as a source of chemical energy (ATP, GTP), precursors for other molecules (GTP converts to folates), secondary messenger (cAMP), signals for gene and metabolic regulation and enzyme cofactors (NAD^+). Due to their involvement in various cell function and replication processes their biosynthetic pathways are appealing targets for antimalarial drug discovery (Christopherson *et al.* 1998).

In general these nucleotide bases and their precursors can be obtained via *de novo* synthesis, from direct acquisition, salvage or interconversion pathways. In the case of *Plasmodium* it synthesizes purines via salvage routes and pyrimidines via *de novo* pathways. In contrast mammals synthesize purines *de novo* and pyrimidines via salvage routes or *de novo* routes (Sahu *et al.* 2008). Additionally, the host homologues may be sufficiently unrelated allowing inhibitors to be designed specifically for the parasite. Inhibitors aimed at suppressing these pathways have been successful in previous attempts with the apicomplexan parasites (Hyde 2007).

Figure 1-6 illustrates the interconnectivity of folate metabolism, purine salvage and pyrimidine biosynthesis in *P. falciparum*. Since the parasite is incapable of purine synthesis it is reliant on permeases to facilitate the passage of nucleotides and nucleobases. GMP can generally be obtained by the salvage of guanine using hypoxanthine/ guanine phosphoribosyltransferase (HXGPRT) (main route since it lacks guanosine kinase) or via conversion of other nucleosides catalysed by GMP synthetase (E.C 6.3.5.2). Furthermore, the salvage of purines in *Plasmodium* is localized to HXGPRT. Adenosine and inosine converts to hypoxanthine and guanosine to guanine, followed by the ribophosphorylation of IMP and hence conversion to AMP and GMP catalysed by HXGPRT. A multiple step catalysed by inosine 5'-monophosphate dehydrogenase (E.C. 1.1.1.205) catalyses the conversion of AMP to GMP (Hyde 2007).

The precursor of pyrimidine nucleotides, uridine 5'-monophosphate (UMP) forms upon decarboxylation of orotidine 5'-monophophate (orotic acid and a ribose phosphate moiety). A double phosphorylation reaction using UMP as the substrate produces uridine 5'-triphosphate (UTP). Cytosine derivatives can then be formed upon the conversion of UTP to CTP (mediated by CTP synthase; E.C 6.3.4.2). Prior to the phosphorylation reactions, the synthesis of thymine nucleotides occurs during the methylation (mediated by thymidylate synthase; E.C 2.1.1.45) of dUMP to form dTMP in the folate pathway (Hyde 2007).

As a summary, pyrimidine biosynthesis is required for the synthesis of dCTP, dUTP and dTTP nucleotides and folate metabolism for the synthesis of folate derivatives (THF). The connection between these two pathways lies within the thymidylate cycle. Purine biosynthesis supplies ATP and GTP nucleotides, this is coupled to folate metabolism since the first substrate in folate metabolism is GTP. Various other metabolic pathways, which are also related to the folate pathway, include the pentose phosphate pathway, the shikimate pathway, glutamate metabolism, methionine and polyamine metabolism.

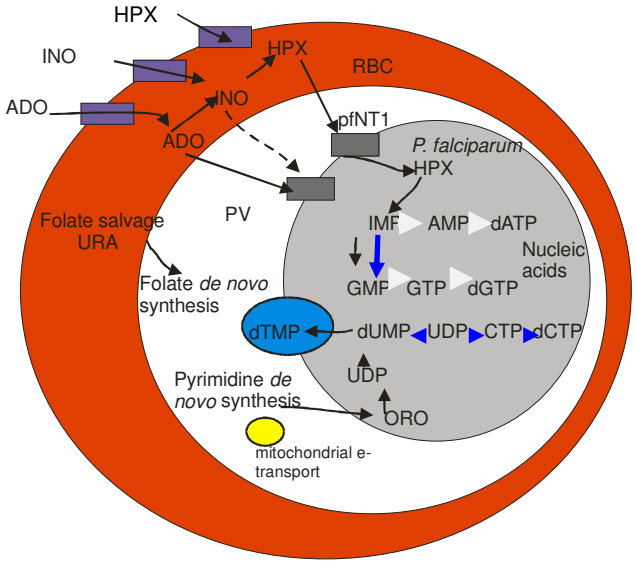


Figure 1-6: The interconnectivity of purine salvage, pyrimidine synthesis and folate metabolism in *P. falciparum*- infected erythrocytes. Hypoxanthine (HPX), adenosine (ADO), inosine (INO), uracil (URA), parasitophrous vacuole (adapted from (Olliaro and Yuthavong 1999; El Bissati *et al.* 2006).

1.6.1 Thymidine synthesis via the folate pathway

Bioinformatic studies used to determine the 'druggability' of various protein classes (23 target types) referred to 90% of proteins involved in vitamin (B_1, B_6 and B_9) processing as 'druggable' since they scored the highest in the specific category (Hajduk *et al.* 2005). The malaria parasite in contrast to its host is capable of synthesizing folates (vitamin B_9) *de novo* whereas the host must rely on dietary folate uptake (Sherman 1979). As such, antimalarial compounds can be designed to target the *Plasmodium* parasite with high specificity with no risk of harming the human host. Folate biosynthesis consists of many potential drug targets that could broaden therapeutic intervention (Muller *et al.* 2009). By establishing differences of folate metabolism in the host and malaria parasite a better understanding can be obtained in order to target other enzymes in this pathway (Table 1-3).

Table 1-3: The comparison of folate metabolism in *Plasmodium* and humans (adapted from (Nzila *et al.* 2005; Nzila *et al.* 2005; Dittrich *et al.* 2008; Hyde *et al.* 2008; Sheng *et al.* 2008; Pribat *et al.* 2009). DHNA, dihydroneopterin aldolase; PPPK, hydroxymethyldihydropterin pyrophosphokinase; DHPS, dihydropteroate synthase; DHFS, dihydrofolate synthase; PTPS, pyruvoyltetrahydropterin synthase; SR, sepiapterin reductase

Differences	**Folate metabolism in Humans**	***Plasmodium***
Folate supplies	Reliant on dietary folate	*De novo* synthesis & salvage
Folate derivatives	Nine variations exist: • folic acid • dihydrofolate (DHF) • THF • 5,10-methenyltetrahydrofolate (5,10-CH^+-THF) • 5,10-methylenetetrahydrofolate (5,10-CH_2-THF) • 5-methyltetrahydrofolate (5- CH_3-THF) • 5-formyltetrahydrofolate (5-CHO-THF) • 10-formyltetrahydrofolate (10-CHO-THF) • 5-formimnotetrahydrofolate (5-NH=CH-THF)	Folate enzymes active thus folate derivates present: • Dihydropteroate (DHP) • DHF • THF • 5,10-CH_2-THF • 5- CH_3-THF • 10-CHO-THF (formylmethionine-tRNA present as substrate) absent/still be discovered: 5-NH=CH-THF. 5-CHO-THF & 5,10-CH^+-THF (reliant on exogenous purine supply)
Major folate	80-90%: 5-methyltetrahydrofolate 10%:THF	5-methyltetrahydrofolate
folate enzymes	DHFR, TS and FPGS (no DHFS) are monofunctional proteins PTPS and SR required to produce 5,6,7,8-tetrahydrobiopterin from 7,8-dihydroneopterin triphosphate (due to cys residue in the PTPS active site)	DHPS-PPPK, DHFS-FPGS and DHFR-TS are bifunctional proteins believed to share the same active site domain PTPS produces 6-pyruvoyl-5,6,7,8-tetrahydrobiopterin from 7,8-dihydroneopterin triphosphate (due to glu residue in the PTPS active site)
Methionine shortage	Results hyperhomocystemia	methionine obtained from host plasma
	Similarities	
Functions	• Deoxythymidine-5'monophosphate synthesis (5,10-CH_2-THF) • Formylation of tRNA (10-CHO-THF) • Glycine and serine metabolism (THF & 5,10-CH_2-THF) • Histidine and glutamate metabolism (5-NH=CH-THF) • Methionine synthesis (5- CH_3-THF) from homocysteine • Folate storage: regulation (5-CHO-THF)	

The purpose of folate metabolism is to supply reduced folates for the conversion of dUMP (deoxyuridine 5'-phosphate) to (deoxythymidine 5'-phosphate) dTMP (referred to as the thymidylate cycle). As depicted in Figure 1-7 tetrahydrofolate (THF) in its 5,10-methylene derivative form undergoes oxidation to form dihydrofolate (DHF). The one-carbon unit released upon oxidation is then transferred to dUMP for the conversion to dTMP and thus the formation of thymidine used for DNA replication (Figure 1-8b) (Hyde

2007). Furthermore, the derivatives of folates function as essential cofactors for DNA replication, for the formylation of methionine (hence for the activation of protein synthesis), glycine and methionine synthesis as well as for the metabolism of serine, histidine and glutamic acid (Nzila *et al.* 2005).

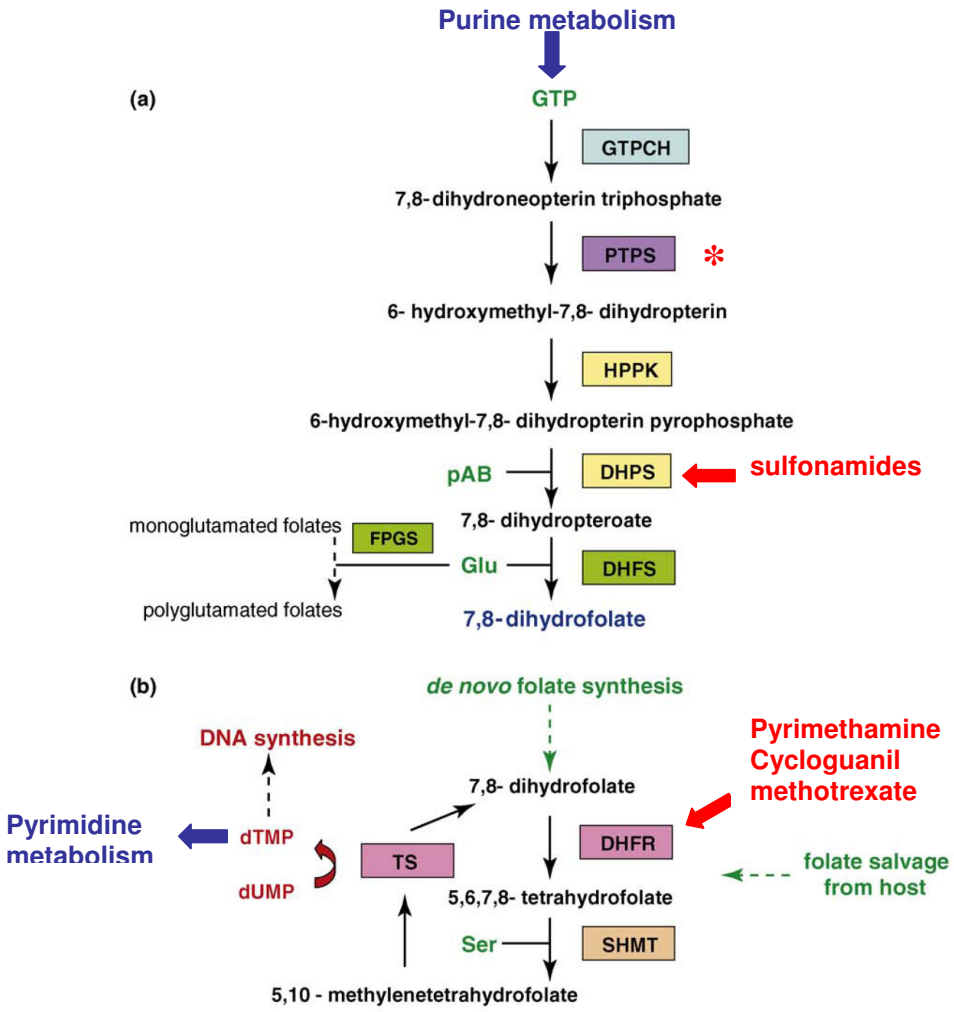


Figure 1-7: a) The synthesis of reduced folates via *de novo* and b) salvage pathways (thymidylate cycle). Indicated in red are the known antimalarial drugs targeting DHPS and DHFR. Enzyme abbreviations: DHFS, dihydrofolate synthase; DHPS, dihydropteroate synthase; FPGS, folylpolyglutamate synthase; GTPCH, GTP cyclohydrolase I; HPPK, hydroxymethyldihydropterin pyrophosphokinase; PTPS, pyruvoyltetrahydropterin synthase III. Boxes of the same color indicate enzyme activities encoded by a single bifunctional gene and dashed arrows signify multistep processes (Hyde *et al.* 2008; Muller *et al.* 2008; Pribat *et al.* 2009).

1.6.1.1 Folate salvage

P. falciparum is capable of salvaging folates from exogenous stores and synthesizing it via *de novo* routes. However, since not all *P. falciparum* strains are capable of using

exogenous folates, the salvage pathway is therefore believed to supply minute amounts of folates. The overall high polarity of folate derivatives indicate that salvage mediated transport is used. Furthermore, at physiological pH folates are dianionic molecules that must be imported against an inwardly negative potential within the *Plasmodium* parasite. Folate uptake by *P. falciparum* is under strict control and is highly dependent on temperature (optimal 37°C), pH (pH 6.9 for folate import) and the availability of any sugar molecules capable of undergoing glycolysis (to maintain an electrochemical gradient across the parasite plasma membrane) (Wang *et al.* 2007).

Primarily the mechanism of proton export within *P. falciparum* is hypothesized to be via V-type H^+ ATPase situated on the plasma membrane. Studies by Wang *et al*. (2007) revealed that an inhibitor of this ATPase type (concanamycin A) repressed folate uptake in low concentration. Other investigations uncovered the existence of genes encoding subunits A and B of the ATPase and the expression of the enzyme throughout the *Plasmodium* erythrocytic life-cycle. Additionally, the proton gradient generated via V-ATPase contributes to an inward negative membrane potential. It is suggested that folate transport is via a H^+/symporter (Wang *et al.* 2007).

It was initially hypothesized that with the import of exogenous folate derivatives within the *Plasmodium* parasite, the affectivity of the antifolate drugs against the parasite lowered. This was demonstrated in studies when *in vitro* cultures were supplemented with folates resulting in a 1000-fold reduction in the inhibitory effect of sulfadoxine. These results confirmed that *dhps*-based sulfa drugs are the most affected by an exogenous supply of folates (Gregson and Plowe 2005). Although, it should be noted that suppression of folate biosynthesis can not be counterweighed completely by this salvage mechanism and therefore, the *de novo* route (located to a separate compartment) is likely to be vital for parasite growth.

It is known that folate derivates i.e. folic acid, folinic acid and THF increase the substrate (DHF) and product (THF) concentrations of the *Plasmodium* DHFR (Tahar *et al.* 2001). Therefore, one would expect that the activity of DHFR inhibitors to decrease in the presence of these folate derivatives. The antimalarial activity of the anticancer drug, methotrexate (MTX) was established as an inhibitor of DHFR using transfection studies conducted by (Fidock *et al.* 1998). Additionally, MTX was proposed to be imported via

folate receptors (Wang *et al.* 2007). However, experiments by (Nduati *et al.* 2008) revealed that the antimalarial activity of certain anticancer drugs (e.g. MTX) did not decrease in the presence of these folate derivatives (Nduati *et al.* 2008). *In vitro* studies on the *Plasmodium* parasite reveal that the salvage of folates is essential for growth. However, the importance of folate salvage *in vivo* remains unknown relative to folate biosynthesis (Wang *et al.* 2007).

1.6.1.2 Antifolate targets worth probing

Although *de novo* folate synthesis can be bypassed by folate salvage, the use of commonly used sulfa drugs (inhibiting DHPS) exposes the importance of this pathway as an antimalarial drug target (Hyde 2007). Furthermore, if the synthesis of THF is eliminated, thymidine starvation would occur and the parasite would not survive. This route can be explored: by inhibition of *de novo* synthesis, folate salvage and recycling of THF (Hyde 2007). Thus it is justifiable to investigate the importance of other potential targets within folate metabolism of the *P. falciparum*.

1.6.1.2.1 6-pyruvoyltetrahydrobiopterin synthase (PTPS)

All of the genes encoding enzymes in the folate pathway have been identified in the *P. falciparum* genome except for dihydroneopterin aldolase (DHNA). Neither the gene of DHNA nor the enzyme could be identified in *Plasmodium* when radioloabeled GTP was used as the substrate (Gardner *et al*. 2002). Recently Hyde *et al*. (2008) suggested that DHNA could be bypassed by an orthologue of 6-pyruvoyltetrahydrobiopterin synthase (PTPS). In some species PTPS uses DHNTP (being similar to DHN) and in others it catalyses the second reaction in the pathway of tetrahydrobiopterin, a cofactor for aromatic amino acid hydroxylase, glycerol ether monoxygenase and nitric oxide synthase (Dittrich *et al*. 2008) (Figure 1-7a). In comparison to orthologues in other organisms the *Plasmodium* PTPS is conserved and is comprised of residues (e.g. triad of histidine amino acids) essential for its function (Hyde *et al.* 2008). A fundamental difference exists between the human and *Plasmodium* PTPS. The human counterpart has a Cys residue at the active site nucleophile, whereas *Plasmodium* possesses a Glu residue. There are additional structural variations, making PTPS a target for antimalarial drug intervention (Muller *et al.* 2009).

1.6.1.2.2 Hydroxymethylpterin pyrophosphokinase (HPPK, E.C. 2.7.6.3)

Attempts at deciphering a possible crystal structures for the bifunctional *P. falciparum* HPPK-DHPS was conducted using homology models and other templates from other species. These templates included the crystal structures of *Saccharomyces cerevisiae* (*S. cerevisiae*) (PPPK–DHPS), *Mycobacterium tuberculosis* (DHPS), *Bacillus anthracis* (DHPS), and *Escherichia coli* (PPPK). The *P. falciparum* HPPK-DHPS structure was derived using the structure of the *S. cerevisiae* counterpart. It was observed that the active site of HPPK and DHPS was highly conserved among *P. falciparum* and certain species. Using this model alternate and effective inhibitors can be designed based on the derived structure of HPPK-DHPS (de Beer *et al.* 2006). Furthermore, since this enzyme is bifunctional, novel inhibitors can be developed to block both HPPK and DHPS activities (Hyde 2007).

Speculations can be made about the bifunctional nature of HPPK-DHPS as well as other folate enzymes. This is because substrates can be channeled/ tunneled from one active site to the next. Crystallographic studies using DHFR-TS from *Leishmania* revealed the existence of an electrostatic highway that would channel DHF between both active sites. These results have not been confirmed in *Plasmodium* although it is important nonetheless, when designing inhibitors, since exogenously supplied antifolate drugs might not affect substrates that are channeled. Besides DHPS, inhibitors have been designed for HPPK to mimic the substrates ATP and 6-hydroxymethyl-7,8-dihydropterin in *E. coli* (Bermingham and Derrick 2002).

1.6.1.2.3 Dihydrofolate synthase (DHFS, E.C. 6.3.2.12)

DHFS mediates the final step in the *de novo* pathway, which is the conversion of dihydropterate to dihydrofolate upon addition of a L-glutamate residue to dihydropteroate. The gene that codes for DHFS has been characterized in *P. falciparum* and is observed to form a bifunctional protein with FPGS (Nzila *et al.* 2005). Activities of the enzyme closely related to DHFS is folylpolyglutamate synthase (FPGS, EC 6.3.2.17), which catalyses the attachment of L-glutamate residues to the γ-tail of polyglutamate derivates additionally requiring Mg^{2+} and ATP. Monoglutamated folate derivatives are not as effective as substrates for folate-dependent enzymes, having longer retention times (increased concentrations within the cell) and displays lower enzymatic patterns in contrast to the polyglutamated forms (Coward and McGuire 2008).

The need for polyglutmation has been demonstrated in organisms that carry a mutated FPGS. One such example was studies using chinese hamster ovary cells; these cells required addition of the end-products of folate biosynthesis and displayed lower levels of intracellular folates. Polyglutamation also functions to inhibit folate leakage through the cell membranes by increasing the overall negative charge of the molecule. DHFS activity has been displayed in humans and antimalarials designed to target the bifunctional enzyme will have a synergistic effect on both the *de novo* pathway and polyglutamation, this poses as a good candidate for drug discovery (Nzila *et al.* 2005).

The characterization of *P. falciparum* DHFS-FPGS requires its expression followed by detailed investigation of structure, kinetics and inhibition studies. A cDNA fragment encoding the *P. falciparum dhfs-fpgs* gene was cloned from the K1 *P. falciparum* isolate into the expression vector. The expressed enzyme was observed to be monomeric like most of the homologues from other organisms. Three substrates exist for DHFS-FPGS, the pterin moiety, DHP for DHFS; mono-/poly-glutamated THF for FPGS and glutamate. Experiments revealed that the fully reduced THF was favored for *Plasmodium* FPGS in contrast to the human FPGS that is capable of utilizing both DHF and/or folic acid for polyglutamation at comparable efficiency. Since the *Plasmodium* FPGS behavior differed in comparison to humans, the binding of the anticancer drug, methotrexate was evaluated as a substrate for *Plasmodium* FPGS. Methotrexate accumulated within the parasite to greater levels in comparison to folinic acid but could not be processed to the polyglutamated forms. Inhibition studies were followed using this drug, however it had no affect on the activity levels of FPGS and DHFS but decreased growth of *Plasmodium* cultures. This effect on parasite growth was due to the tight binding of methotrexate to DHFR, despite the mutations DHFR carry that confer pyrimethamine resistance. As such, novel folate analogues have to be developed that can exploit the fact that DHFS has no human homologue and that the *Plasmodium* FPGS utilizes alternate substrates in comparison to its human counterpart (Wang *et al.* 2010).

1.6.1.2.4 Serine hydroxymethyltransferase (SHMT, E.C. 2.1.2.1)

Another enzyme featuring in the thymidylate cycle that has potential as antimalarial drug target is serine hydroxymethyltransferase (SHMT). It catalyses the transfer of methylene (from Ser) to THF forming 5,10-methylene THF. The *shmt* gene has been identified in *P. falciparum* and other parasites (Hyde 2007). Since the activity of SHMT increases

dramatically during the S-phase of the parasite growth cycle, inhibition of this enzyme will result in the impairment of cell growth (Nzila *et al.* 2005). The *Plasmodium* SHMT is well conserved, although three residues within the active site differ in comparison to that of humans. Furthermore, SHMT in most eukaryotes are either dimers or tetramers (formed from the connection of two dimers). His residues, which provide this connection for tetramer formation are however, absent in *P. falciparum*. To clarify this, studies to determine the oligomeric state of SHMT in *P. falciparum* were conducted. Data revealed SHMT existed as a dimer since the molecular weight was calculated to be 81.5kDa (slightly lower than the expected molecular weight for a dimer, which is 100kDa) (Pang *et al.* 2009).

The mung bean SHMT is enzymatically active without the requirement for pyridoxal 5'-phosphate (PLP), however in general almost every SHMT requires this cofactor. Results confirmed the need for PLP for enzyme activity. Even though antifolate inhibitors could be designed to target the PLP binding site, many other enzymes (e.g. enzymes of polyamine metabolism: S-Adenosylmethionine decarboxylase and lysine decarboxylase) also require PLP. Therefore, inhibitors have to be designed very specifically. Inhibition studies were followed using antifolate inhibitors for DHFR and TS given the structural similarity of the substrates (THF to DHF and 5'-methylene THF), however inhibition was not significant enough (IC_{50} of 150μM). Since the *P. falciparum* SHMT was expressed using a heterologous expression host, biochemically characterized and its kinetic properties determined; specific inhibitors can be developed against *P. falciparum* SHMT as an antimalarial drug target provided further studies will be conducted to assess enzyme behaviour (Maenpuen *et al.* 2009; Pang *et al.* 2009).

1.6.2 Guanosine triphosphate cyclohydrolase (GTPCHI, E.C. 3.5.4.16)

1.6.2.1 Characteristics of the active proteins GTPCHI in other species

GTPCHI catalyses the synthesis of dihydroneopterin triosphosphate (pterin ring) and formic acid (Figure 1-8) (Thony and Blau 1997). Initially, GTPCHI was identified in bacteria and displays high conservation across many phyla. GTPCHI was previously neglected despite its importance in metabolic pathways due to difficulties experienced in

acquiring soluble, active protein using heterologous expression systems (Muller *et al.* 2008)

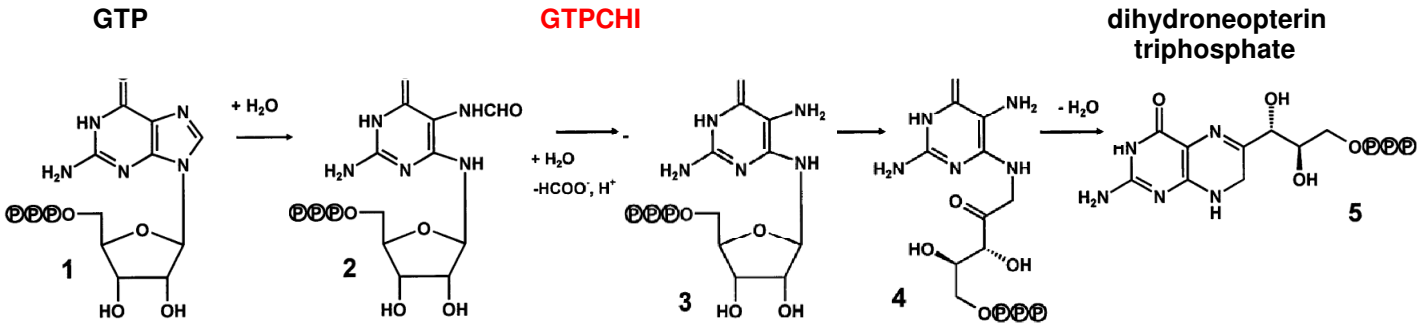


Figure 1-8: A hypothetical reaction mechanism of the first reaction of folate biosynthesis in *P. falciparum* catalyzed by GTPCHI (adapted from(Burg and Brown 1968).

Gene sequence analysis revealed that *gtpch*I is a single copy gene when genomic DNA was digested with restriction enzymes and analysed using a southern blot. The chromosomal location was established to correspond to chromosome 12 when the chromosomes of parasite strains (3D7 and K1) were hybridised with PCR products to CHEF blots (Lee *et al*. 2001). The human *gch*I is also encoded by a single gene, but its loci is mapped to chromosomes 14q21 (Thony and Blau 1997). Investigation revealed that the full-length of *P. falciparum* GTPCHI protein consisted of 389 amino acid residues with a molecular mass of 45.9 kDa and an isoelectric point of 9.89 (Krungkrai *et al*. 1995 and www.PlasmoDB.co.za). The partially purified *P. falciparum* enzyme via size exclusion chromatography proposed a value of *ca* 300000, indicating it is a multi-subunit complex (as a decamer) (Krungkrai *et al.* 1985; Lee *et al.* 2001). The crystal structure of *E. coli* GTPCHI was determined at 3.0Å and the human homologue at 3.1Å. The molecular weight of the *E. coli* protein is 24.7 kDa and the mammalian protein is 28.0 kDa, smaller than the *P. falciparum* protein. The *E. coli* GTPCHI (having 70% amino acid homology to mouse and human) is a homodecamer consisting of 2 pentameric toroids (doughnut-shaped structure) assembled face-to-face (Nar *et al.* 1995). In *E. coli* each domain of this protein consists of 221 amino acids, which fold into a α-β structure and a primarily helical N-terminus. The N-terminal antiparrallel helix pair (α_2/α_3) is observed to be isolated from the rest of the molecule and is positioned on top of α_4 and α_5. The C-terminal region forms part of the monomer (comprising 95-217 residues) and is composed of a central four-stranded antiparrallel β_1-β_4 structure that is flanked on either side by α_4, α_5 and α_6-helices (Thony *et al.* 2000).

The *P. falciparum* GTPCHI has an elongated N-terminal residue (~150 amino acids) in comparison to other organisms that is composed particularly of Lys and Asn residues (Trp, Ala and Pro are absent). This domain is hypothesized to function in protein-protein interactions for regulatory activity (Witter *et al.* 1996) and is found on the surface of the decameric holoenzyme. The C-terminal domain (~120 amino acids) is however, well conserved revealing a 39% identity to other homologues. A superfamily of tunnelling fold (T-fold) forms part of the C-terminal domain and functions in pterin binding. This superfamily consists of five known T-folds found in five various enzymes with diverse functions. This includes dihydroneopterin-triphosphate epimerase (DHNTPE), DHNA, GTPCHI, PTPS, and uricase (UO, uroate/urate oxidase). These enzymes share a fold-related binding site with a glutamate or glutamine residue anchoring the substrate and many conserved interactions. The substrate is located in a deep and narrow pocket at the interface between monomers (Thony *et al.* 2000). It is known that the human host lacks the ability to synthesize folates but does have GTPCH activity, which is used for reduced biopterin formation (cofactor required for aromatic amino acid hydroxylation and nitric oxide synthesis). A 37% identity and 62% similarity of *P. falciparum* to its human counterpart was observed, indicating a sufficient enough divergence to its host. *P.falciparum* GTPCHI poses as a candidate target for antimalarial drug discovery (Lee *et al.* 2001).

1.6.2.2 Copy number variation and antifolate drug resistance

Interest in GTPCHI has been rekindled since the *Plasmodium gtpch*I gene was amplified 11-fold in certain strains that carry resistance due to mutations in *dhfr* and *dhps* (Kidgell *et al.* 2006). Copy number polymorphism (CNP) is a characteristic that is enhanced for particular gene classes, indicating evidence for positive selection. However, this could also point to isolating selection against CNP in gene classes. A CNP on chromosome 5 underlies a multidrug resistance phenotype and upon manipulation of the CNP, the response to drugs is altered. DHFR (chromosome 4) and DHPS (chromosome 8) are downstream enzymes relative to GTPCHI (Nair *et al.* 2008). By upregulating the flux (i.e. increasing the gene dosage) through the pathway the lowered efficacy of *dhfr* and/or *dhps* bearing resistance mutations could be compensated for (Kidgell *et al.*,2006). Amplification of GTPCHI was believed to contribute to an increase in flux through the pathway and might therefore contribute to antifolate resistance. Knock-out studies were

conducted using transfection systems and these results indicated the suppression of DHPS activity. Inhibition studies still need to be conducted on the enzyme but based on the results obtained for knock-out experiments of this gene, it will be useful to assess GTPCHI as an antifolate target (Nzila *et al.* 2005). Studies by Nair *et al*. (2008) revealed compelling proof that the *P. falciparum gtpch*I CNP formed as a result of adaptive evolution associated with antifolate drug treatment. Additionally, the mutation in *dhfr* correlates to the *gtpch*I CNP, which shows a functional linkage between genes on different chromosomes of the same pathway. CNP may be a common evolutionary response to strong drug selection since they are involved in drug resistance in other protozoa, bacteria and cancers. At present scientists are identifying genetic determinants of drug resistance and virulence (Nair *et al.* 2008).

This CNP in *gtpch*I suggests it could bridge a gap to malaria susceptibility. By effectively blocking this target in *P. falciparum* the current problem associated with resistance against antimalarial drugs (targeting *dhfr* and *dhps*) can be circumvented. This would entail designing effective inhibitors of *P. falciparum* GTPCHI.

Research aims

Since the *P. falciparum* GTPCHI has not been previously expressed sufficiently in soluble form the objective of this study was to express the GTPCHI from the native, synthetically devised (codon optimized or harmonized) gene sequence in sufficient quantity. Specific aims include:

- **Chapter 2** concentrates on the construction of a codon harmonized *P. falciparum gtpch*I gene, the various sequence analyses preceding the synthesis, as well as a focus on the heterologous expression of *P. falciparum* GTPCHI from the native, codon optimized and codon harmonized *gtpch*I gene and optimization thereof for the production of soluble purified his tag proteins
- **Chapter 3** includes the attempt to further improving *P. falciparum* GTPCHI expression using the Cell-free wheat germ based expression system
- **Chapter 4** is the final summation of all the chapters, the conclusions deduced and including future perspectives

The results of chapter 2 were previously presented at the SASBMB conference (2010).

Chapter 2
Heterologous expression and optimization of native, codon optimized and harmonized *P. falciparum* GTPCHI using *E. coli*

2.1 Introduction:

2.1.1 *E. coli* as an expression host:

It is known that the production of a functional protein is linked to the cellular machinery of the expression host. However, target proteins (including those of *Plasmodium*) of interest cannot be isolated in acceptable quantities from the natural host. As such, no single expression system is suitable for universal protein expression. Therefore, this necessitates the use of heterologous hosts for the expression of the target proteins. Furthermore, the expression system could be chosen depending on whether the protein to be expressed is a membrane protein or is required in soluble form (Birkholtz *et al.* 2008a). *Escherichia coli (E. coli)* remains the favoured choice for the expression of recombinant proteins since it provides the simplicity of use, lowered cost, speed and high level protein production. Even though this expression system is preferred the expression of *Plasmodium* proteins has led to many failed attempts (Birkholtz *et al.* 2008a). However, many large-scale studies have been conducted to express these proteins but of the proteins purified only a select few were present in soluble form (Mehlin *et al.* 2006). The majority of the proteins express in an insoluble form, which are refractile aggregates of miss-folded proteins that are resistant to proteases and sequestered in inclusion bodies (Carrio and Villaverde 2002). Inclusion body formation is caused by an unbalanced equilibrium that occurs between *in vivo* protein aggregation and solubilization (Sorensen and Mortensen 2005). This is most likely due to the high A+T content of the *Plasmodium* genome and other characteristics indicated in Table 2-1.

Table 2-1: Attributes of *P. falciparum* genome and reasons for poor expression. Adapted from (Mehlin *et al.* 2006; Vedadi *et al.* 2007)

Attributes of *P. falciparum*	**Reasons for poor expression in *E. coli***
80% A+T rich genome	• Transcription into codons (AGA and AGG) rarely used by *E. coli* proteins. • Proteins have low complexity regions consisting of long hydrophobic stretches (lysine and asparagines) • Long repeated stretch of asparagine and glutamine prone to form amyloid-like aggregates
Post-translational modifications	Glycosylation and phosphorylation patterns are unique to the parasites
Genes contain cryptic start regions	Formation of many, truncated protein products during overexpression
Mean gene length between 2.3 kb—1 kb, characteristic of unique inserts	Longer in comparison to other homologues
pI	Increase in pI (>6) results in lower expression of soluble protein
Size	Greater the size (>56kDa), lower expression of soluble proteins
Signal and transit peptides, sequence inserts	Affect recombinant protein expression and crystallization
10% of proteins predicted to be secreted to host, N-termini contain sequence motifs	Regulating their localization rather than function but are membrane-like in their effects on expression and folding

Inclusion body formation can be minimized by controlling the temperature, the use of different promoters (to control rate of expression), the use of enriched media for growth, alternate host, insertion of a suitable tag and by the co-expression with plasmid-encoded chaperones (Jonasson *et al.* 2002). Other procedures such as the solubilization of the purified aggregates are possible using detergents like urea and guadinium hydrochloride followed by refolding of the target protein (Sorensen *et al.* 2003). Figure 2-1 depicts the strategies for the expression of *Plasmodium* proteins using various other expression systems and techniques of optimizing expression in *E. coli*.

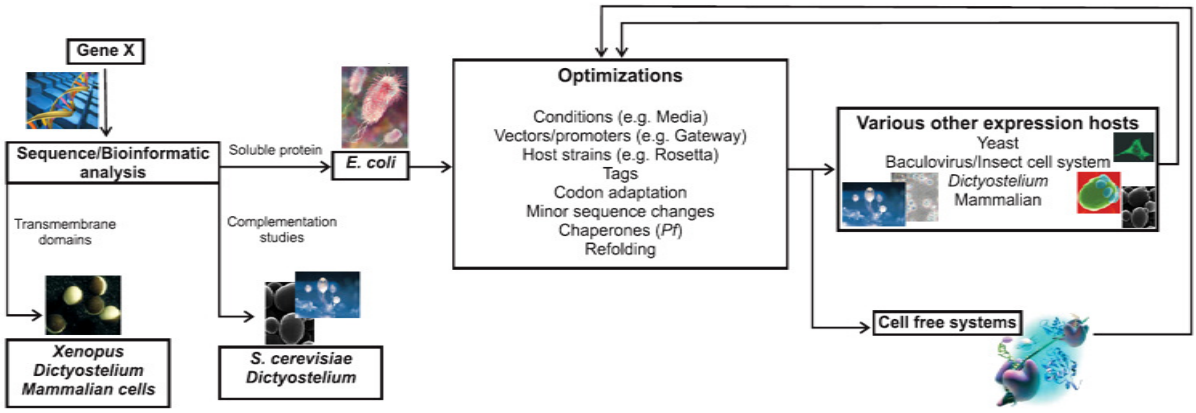


Figure 2-1 Strategies for expression of *Plasmodium* proteins (Birkholtz *et al.* 2008a)

2.1.2 Optimization of recombinant protein expression

2.1.2.1 Expression vector system

There are fundamental elements when designing studies for recombinant protein expression (Table 2-2). One of these is the choice of plasmid; this includes the antibiotic (e.g. ampicillin) resistant marker, origin of replication, transcriptional promoters (induced by sugar compounds, isopropyl-beta-D-thiogalactopyranoside or by temperature) / terminators and translational initiation (Hannig and Makrides 1998; Baneyx 1999; Jonasson *et al.* 2002). It has been suggested that the most favourable expression vector and *E. coli* host strain used for the expression of *Plasmodium* proteins is based on the combination of the pET vector systems and the BL21 (DE3) *E. coli* strain. This consensus was achieved based on experimental data of the structural genomics consortia with the expression and purification of greater than 100 000 proteins from Eubacteria, Archea and Eukarya (Sorensen and Mortensen 2005). The pET expression system boasts the presence of hybrid promoters, protease cleavage sites and multiple cloning sites. The reason for the use of the BL21 cell lines is primarily because the pET system require a host strain lysogenized by a DE3 phage segment, which encodes the T7 RNA polymerase regulated by IPTG inducible *lacUV5* promoter. It is also deficient in ompT and lon proteases, which normally hinders isolation of recombinant proteins. Furthermore, the *recA* negative strains stabilize target plasmids containing recurring sequences (Novagen BLR strain), the *trxB/gor* negative mutants improve cytoplasmic disulfide bond formation (Novagen Origami and AD494 strains) and the *lacY* mutants allow proteins to be expressed in regulated levels (Sorensen and Mortensen 2005).

The stability of mRNA, competency of transcription and the frequency of translation also control expression levels. The decay of mRNA directly influences the gene expression levels. Therefore *E. coli* strains possessing a mutation in the gene encoding RNaseE are designed for improving mRNA stability in recombinant expression systems (Invitrogen BL21 star strain) (Lopez *et al.* 1999). In conjunction with the choice of the host strain, it has been reported that fusion tags (hexahistidine, gluthathione-S-transferase, solubility 'Enhancing' Ubiquitous Tag etc.) enhance the expression of proteins in soluble form and the outcome can be predicted using bioinformatic tools. This information is useful when selecting the appropriate expression vector (Birkholtz *et al.* 2008a)

Table 2-2: The strategy of optimizing recombinant expression of *P. falciparum* proteins in *E. coli*. Adapted from Human. 2007.

Techniques	Comments	Examples	Advantage/Disadvantage	References
***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. PfMSP-119 in *E. coli* BLR(DE3) & M15pREP4	Changes in levels of tRNA might have metabolic effects on host.	Stratagene & Novagen (Mazumdar *et al*., 2009)
***Plasmid Copy number**	Determines gene dosage. Moderate to high copy number plasmids are favoured.	Genes cloned into plasmids that replicate in a relaxed manner, present at 15–60 (e.g. pMB1/ColE1- derived plasmids) or a few hundred copies/cell (e.g. pUC series of pMB1 derivatives).	Plasmid loss of very high copy number plasmids, plasmid- borne genes are toxic to the host: plasmid-encoded antibiotic-resistance markers (supplementing with antibiotics) Vectors are designed to carry gene(s) or repressors that cause cell death upon plasmid loss e.g. *lac* operator	(Jana and Deb 2005) (Baneyx. 1999)
***Co-expression plasmids**	Co-transformation with plasmid to be able to overproduce appropriate tRNAs may increase soluble expression.	Overexpression of *Pf*. DHPS by co-transformation with RIG-plasmid.	Problems such as transcription of A+T rich genes, mRNA stability and secondary structures using RIG plasmid	(Baca and Hol 2000)
***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 Baculovirus-derived polyhedrin and mammal-derived cytomegalovirus promoters.	Most of these promoters do have leaky expression. The baculovirus-based PbCSP vaccine displayed & expressed protein on the viral envelope, high antibody titers and induction of both Th1 and Th2 complete protection against sporozoites challenge.	(Jana and Deb 2005) (Yoshida *et al*., 2009)
Codon optimization §	Utilizing the optimal *E. coli* codons frequently improves protein yield.	*Pf*MSP-119 in *E. coli* Plasmid vaccines encoding *Py*CSP, *Pf*CSP and *Py*HEP17 based on native or mammalian codon	Costly and time consuming. May result in increased protein levels but not in soluble protein. Higher levels of *in vitro*	(Mazumdar *et al*., 2009) (Dobano *et al.* 2009)

		optimized gene	expression of the target antigen that translated into more potent antibody responses. non-codon-optimized DNA vaccines proved similar in their ability to induce antigen-specific T cell responses.	
Codon harmonization §	Synonymous codons from *E. coli* were selected that closely resemble the codon usage of target gene in the native host, including regions coding 'link/end' segments of proteins in native host	*Pfs*48/45 gene block transmission of the parasite from invertebrate mosquito vector to vertebrate host	CH-r*Pfs*48/45 (male gamete fertility gene) showed remarkable immunogenicity and functional activity, i.e. transmission blocking activity	(Chowdhury *et al.* 2009)
Translational attenuation§	Effectiveness and spacing of transcription terminators are needed to prevent ribosome skipping.	These characteristics, together with codon optimization, can be seen as codon harmonization where the codon bias, secondary structures & ribosomal pause sites of gene of interest are taken into account.		(Kincaid *et al.*, 2002)
mRNA stability §	The secondary structure of mRNA and stability influences the yield of protein expression.			
Translational signalling§	Efficiency of gene expression is influenced by secondary structure at translation initiation region of mRNA.			
Temperature ¤	Temperature has a pronounced effect on protein folding and stability.	Lowering of the temperature to 15- 20℃ for protein expression. syn*Pf*MSP-119 expression at 25℃.	Slows the rate of transcription, translation, refolding, thereby allowing for proper folding & decrease heat shock protease activity	(Sahdev *et al.* 2008) (Mazumdar *et al.*, 2009)
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. Bioreactor cultivations of syn*Pf*MSP-119 in a semi defined medium containing: glucose, yeast extract, various salts, trace elements, citric acid, EDTA & Thiamine–HCl or defined media without yeast extract	Results in increased soluble protein levels. Yeast extract of semi defined media enhanced the level of soluble PfMSP-119 expression by inhibiting proteolysis	 (Mazumdar *et al.*, 2009)
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 *Pf*EMP1 protein. Batch and fed-batch	Yield of PfMSP-119 obtained using the two-step feeding strategy was 1.5 times higher than the yield of PfMSP-119 obtained using	(Flick *et al.* 2004) (Mazumdar *et al.*, 2009)

		cultivations of synPfMSP-119 batch mode, cells were induced at OD_{600} of 8.0 and fed-batch mode, culture was grown in batch mode until carbon source was exhausted from medium, induction was performed at OD_{600} of 40	the single step feeding strategy	
¥ Small heat-shock proteins and Hsp70/90/60	Interact and stabilize partially folded target proteins to prevent aggregation. Aid correct folding and refolding of proteins during expression.	hsp90 and hsp70 are the most abundantly expressed. Importantly *Pf*hsp70 is considered a potential vaccine candidate and *Pf*hsp90 has been implicated as a drug target against the parasite.	Co-expression provides an alternative to complicated and costly expression systems (e.g. mammalian cultures)	(Acharya *et al.* 2007) (Tolia & Tor. 2006)

Key: *- Expression vector system, §- Codon usage, ¤- Growth conditions, ¥- Chaperone proteins

It is known that there are no universal applicable methods for the expression of *Plasmodium* proteins. However, unique qualities of *Plasmodium* genes and proteins are known to affect the expression of soluble proteins. As such, this chapter will focus on manipulating codon usage of the native *P. faciparum gtpch*I gene by means of codon harmonization. Additionally, experimental conditions such as the usage of *Plasmodium* chaperone proteins, expression strains and growth conditions will be altered to improve the expression of soluble *P. falciparum* GTPCHI.

2.1.2.2 Codon harmonization

Previous studies by Kincaid *et al.* (2004) proposed the use of codon harmonization or translational attenuation for identifying regions of mRNA that translate slowly. The process of codon harmonization entails substituting codons to replicate the codon frequency preference of the target gene in *P. falciparum* rather than the codon frequency preference in the expression host. This will then fix the positional frequency of the low/intermediate and high frequency codons in the non-natural expression host allowing translational processes to duplicate that of the host (*P. falciparum*). It is known that during the process of translation, when ribosome's come in contact with low frequency codons, translational pausing occurs. This is crucial for the correct folding of proteins since, halting at these pause sites allows protein expression to slow down at appropriate intervals (Angov *et al.* 2003; Angov and Lyon 2006). Therefore, the natural translational processes such as pausing and protein folding in the natural host (*P. falciparum*) should occur similarly in the heterologous host (Thanaraj and Argos 1996; Hillier *et al.* 2005).

Figure 2-2 explains the codon harmonization process in comparison to codon optimization. Codon optimization is altering the codons of the target genes to the preferred codons of the expression host (Angov *et al.* 2008).

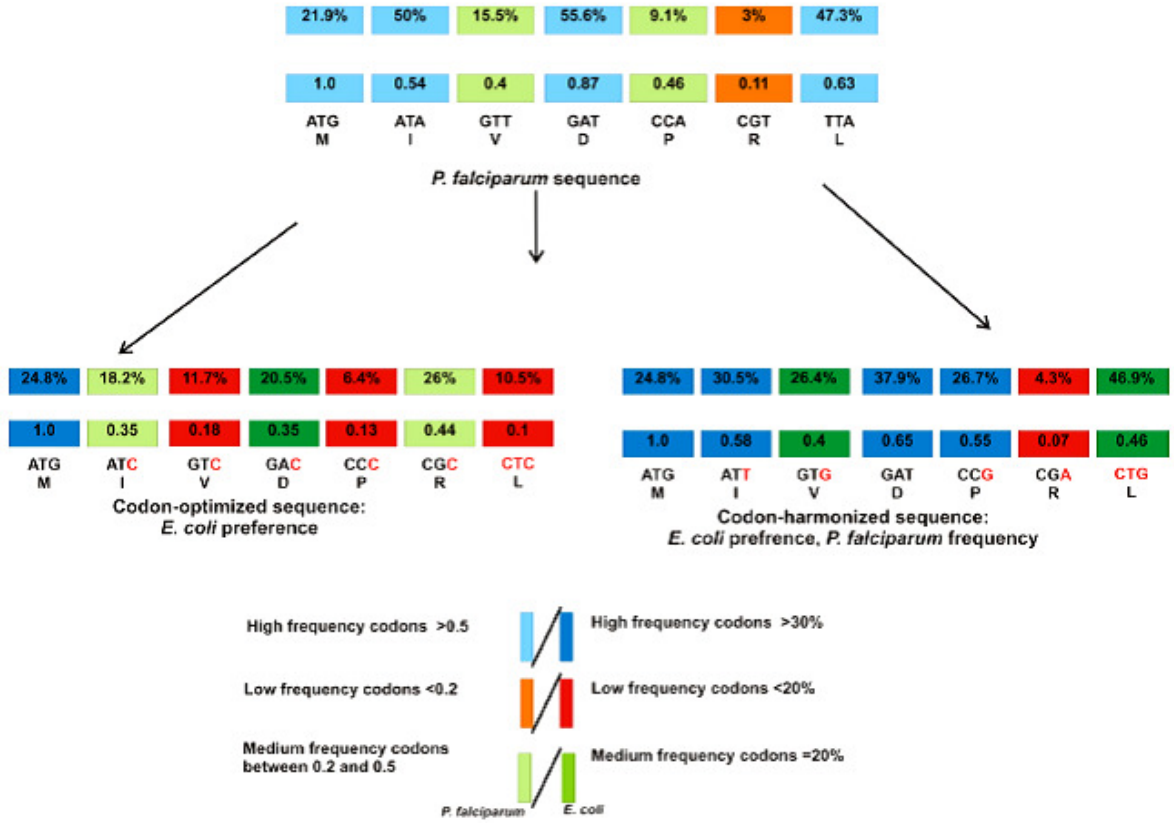


Figure 2-2: The codon harmonization algorithm, which synonymously replaces the codons in *E. coli* to the codon usage frequency similar to the native *(P. falciparum)* expression host (de Ridder. 2005).

Angov *et al*. (2008) conceptualized the procedure for codon harmonization. The principle technique relies on identifying regions of the gene that might translate slowly and recoding of other regions in the gene (Angov *et al.* 2008). Investigations by (Thanaraj and Argos 1996) dealt with identifying regions of mRNA that translated slowly on a set of *E. coli* proteins (three-dimensional structure was available). The frequency of codon usage in the mRNA of highly expressed *E. coli* genes was used to determine the speed of codon translation. Results indicated that in regions that translate slower, abundant codons and rare codons that translate at a high rate were avoided. Furthermore, the translationally slow regions may potentially form hairpin secondary loops that could stall the ribosome movement. To predict transcript regions, which translate slower, residues that occur frequently in unstructured regions (disorganized structure containing biased

codons) of known *E. coli* proteins were indicated based on results by Thanaraj & Argos. (1996). Based on these results, for codon harmonization regions within the target gene sequence, which contain mostly frequently occurring amino acids: Tyr, His, Trp, Ile, Leu, Val, Ser, Thr, Pro, and Cys are assigned (Angov *et al.* 2008). The availability of tRNA determines the speed of codon translation and codon and amino acid usage is dependent on tRNA levels of highly expressed genes (Thanaraj and Argos 1996). As such, the frequency at which each codon is used within the natural host (*P. falciparum*) as a fraction of all isoacceptors capable of encoding the amino acid is assigned. A reference level is then chosen that defines low usage codon frequency in the natural host (*E. coli*). Those amino acids that have codon usage frequencies below the selected reference level are marked. To optimize the frequency of rare codons, a lower or higher reference frequency is selected. In terms of the pause sites, the codons occurring at these regions in the natural host is replaced with codons in the non-natural host to mimic its usage frequency. This process is followed for the recoding putative link/end segments (Angov *et al.* 2008). Link/end segments refer to the boundaries of polypeptide chain segments that have been found to contain clusters of rare codons (Thanaraj and Argos 1996b). The remaining gene is recoded simply by choosing synonymous codons from the heterologous expression host having codon usage frequencies that are similar (either equal to or nearest frequency) to the usage frequencies for the target gene in the natural expression host (Gustafsson *et al.* 2004).

An example of improving the expression of a pre-fertilization antigen, *P. falciparum* 48/45 was documented using codon harmonization. High yields of soluble protein, in the correct conformation were obtained. Furthermore, the recombinant *P. falciparum* 48/45 elicited potent malaria transmission blocking antibodies in non-human primates making it an effective candidate as a transmission blocking vaccine that specifically targets the sexual developmental stages of the *Plasmodium* parasite in the mosquito vector (Chowdhury *et al.* 2009)

2.1.2.3 Co- expression: Molecular chaperone proteins

The cause of difficulties in the expression of most proteins (including those of *Plasmodium*) is due to incorrect folding in the non-natural host. Missfolding of proteins is expected due to stress conditions (e.g. heat, codon bias, mRNA instability, etc.) when proteins are overexpressed resulting in the formation of inclusion bodies. The nature

behind inclusion body formation is unknown, however it could be due to hydrophobic interactions between exposed patches on unfolded chains or by precise clustering of folding intermediates (Sorensen and Mortensen 2005). Furthermore, recombinant protein expression induces a “metabolic burden”, which is the amount of raw materials and energy used by the host for expression and maintenance of foreign protein (Bentley and Kompala 1990). Both factors (inclusion body formation and metabolic burden) are linked to the failure of recombinant expression of soluble proteins (Sorensen and Mortensen 2005)

As mentioned, by controlling various parameters of the host system (including the choice of vector), gene sequence and growth conditions (temperature, media composition, etc.) protein aggregation can be reduced. Another aspect is the co-expression of the target protein with chaperones encoded by a separate plasmid (Sorensen and Mortensen 2005). Molecular chaperones are proteins that assist other polypeptides to fold and function correctly or reach a specific cellular location by interacting with solvent-exposed hydrophobic patches in partially folded proteins. This prevents self-association and aggregation of proteins. Other features include protein assembly and transport, regulation of signal transduction, differentiation and development in higher eukaryotes (Acharya *et al.* 2007; Botha *et al.* 2007).

Heat shock proteins (hsps) are well conserved and several function as molecular chaperones. *P. falciparum* heat shock proteins (*Pf*hsp) may possess a double function. This includes cytoprotection of the parasite and remodeling of the host cell within the erythrocyte cytoplasm (Pesce *et al.* 2008). Another *Pf*hsp known to play a role as a co-chaperone of *Pf*hsp70’s is *Pf*hsp40, since it delivers the substrate to the *Pf*hsp70’s to modulate its activity. The parasite genome encodes at least 43 *Pf*hsp40s indicating a greater function of these proteins for regulating both the host and parasite *Pf*hsp70’s (Cheetham *et al.* 1994; Kelley 1998; Suh *et al.* 1998; Young *et al.* 2003)

P. falciparum encodes at least 6 *Pf*hsp*70* genes. The expression of certain *Pf*hsp70 is enhanced upon heat stress (Flaherty *et al.* 1990; Wang *et al.* 1993). *Pf*hsp70-I is a cytosolic chaperone in the erythrocytic stages and is involved in the multi-chaperone machinery with *Pf*hsp 90 (Acharya *et al.* 2007). Previous studies revealed that an in increase in protein yield and solubility occurred for proteins co-transformed with

molecular chaperones from the same species cloned into different plasmids (de Marco and De Marco 2004).

2.1.2.4 Auto-induction

The pET expression system necessitates a host strain lysogenized by a DE3 phage fragment, which encodes the T7 RNA polymerase under the control of an inducible promoter. The disadvantage of using an inducible T7 expression system in the over-expression of proteins in *E. coli* host strains is due to an overactive T7 RNA polymerase. As such, this results in considerable expression of the target protein even in the absence of inducer. If the target protein is toxic to the heterologous host, the presence of the vector may cause the accumulation of mutations or failed expression of target protein. This situation can be improved by utilization of auto-induction media to express proteins that are regulated by the *lacUV5* promoter without the supplementation of an inducer, such as IPTG. The type of auto-induction media used is dependent on the metabolism, which the heterologous host uses to control the use of the energy and carbon supplies. Expression studies conducted using glucose resulted in the inhibition of lactose uptake thus suppressing allolactose formation (an inducer of the *lacUV5* promoter) (Peti and Page 2007). However, Studier (2005) revealed that the presence of both glucose and lactose bacterial cultures are capable of improved growth since glucose supplies are utilized initially and depleted followed by induction by allolactose. However, a factor that hinders expression is a decrease in pH that occurs when glucose is used as a carbon source thus glycerol is an effective substitute for supporting growth (Studier 2005). Glycerol is also known to accumulate an acidic environment but in contrast to glucose, glycerol does not prevent the inducing activity promoted in complex media. The pH can also be maintained by a low phosphate content of media, allowing sufficient buffering capacity since higher phosphate concentrations reduce cell growth. Auto-induction media is suitable, proficient and inexpensive for protein production at almost any level (Studier 2005).

This chapter emphasized an array of strategies available for the optimization of heterologous expression of *P. falciparum* proteins in *E. coli*. However, specifically for the expression of soluble *P. falciparum* GTPCHI the use of codon harmonization, different host strains, altered media composition and co-expression with molecular chaperones

will be investigated. Furthermore, comparative expression studies of GTPCHI encoded by the native, codon optimized and codon harmonized genes will be compared.

2.2 Materials and Methods

2.2.1 Introduction of *P. falciparum gtpchI*

The native *P. falciparum gtpchI* gene was amplified from genomic DNA (containing no introns) obtained from *Pf*3D7 strains isolated by de Ridder (2006). The codon optimized *P. falciparum gtpchI* gene was cloned into the pET46-Ek-LIC plasmid (obtained from Prof Hyde, Manchester, UK). Table 2-3 gives an indication of the vector used for sub-cloning, the insert gene as well as the annotation of the genes and proteins used in the text.

Table 2-3: Vector-construct and the gene and protein annotations utilized.

Vector	Insert origin	Annotation in text	
		Gene	Protein
pET22b	Native *P. falciparum* GTPCHI	n*gtpchI*	nGTPCHI
	Codon optimized *P. falciparum* GTPCHI	o*gtpchI*	oGTPCHI
	Codon harmonized *P. falciparum* GTPCHI	h*gtpchI*	hGTPCHI

2.2.2 Primer design

Primers were specifically designed for cloning the n*gtpchI* and sub-cloning the o*gtpchI* and h*gtpchI* genes into a pET22b expression vector. The forward primers (nF, oF and hF) were constructed to include a 5' end *NdeI* (CAT↓ATG). The ATG codon from the included restriction endonuclease site was utilized as the start codon. The reverse primers (nR, oR and hR) contained 3' *BamHI* (G↓GATCC) restriction enzyme sites for directional cloning into pET22b vector. Furthermore, for complete read-through of the pET22b containing the *gtpchI* genes and the C-terminal His tag, additional bases (TG) were incorporated in the 5' end and the stop codons of each gene were altered for expression of a C-terminal His_6-tag using the reverse primers (refer to Table 2.6). Primers were analyzed for specificity, internal stability and miss-priming using the Oligo 6 primer analysis software program (Molecular Biology Insights, Inc. version 6.7.1.0). Whitehead Scientific Ltd. (South Africa) synthesized the primers that were dissolved in 10 mM Tris-HCl (pH 8.6) buffer to a final 100 µM concentration (according to the manufacturer's instructions) and rotated at 4 ℃ overnight to ensure a complete solution and stored at –20 ℃ in nuclease free tubes.

2.2.3 Nucleic acid concentration determinations

In order to prepare working solutions of 5 pmol/µl for the primers, the absorbance of each primer was first determined spectrophotometrically at 260nm using the GeneQuant Pro UV/Vis Spectrophotometer (Amersham Biosciences, Sweden). Thereafter, the molar concentrations were calculated using the manufacturer's value for oligonucleotide as follows:

Molar Concentration = $\frac{A_{260nm} \text{ x [ng/µl] x dilution factor}}{\text{Molecular mass (ng/nmole)}}$

Similarly, using a 1/10 dilution (with $dddH_20$) of PCR products and plasmids, DNA concentrations were determined.

2.2.4 PCR gene amplification

2.2.4.1 n*gtpch*I

The n*gtpchI* gene was amplified from genomic DNA previously isolated from the *Pf*3D7 strain by de Ridder (University of Pretoria, South Africa) using the designed primers in section 2.2.2. All the PCR reactions were performed in a Gene-AMP PCR System 9700 (Perkin Elmer, USA) in 0.2 ml thin-walled PCR tubes. The reactions contained 0.3 pmol/µl of nF and nR, 0.02 U/µl KAPA2G fast polymerase (KAPA Biosystem, USA), Buffer A (proprietary mix), 0.2 mM dNTP mix, 0.4 mM $MgCl_2$ and 8 ng/µl isolated genomic DNA in a final volume of 25 µl. The KAPA 2G fast polymerase allows for high processivity and speed. Additionally, it possesses 5'-3' polymerase and 3'-5'exonuclease activity and was capable of generating 3' A-overhangs, making A/T cloning possible. The PCR cycling profile composed of a 2 min denaturing step at 95 ℃, 5 cycles of denaturation at 95 ℃ for 30 sec, annealing at 41 ℃ for 1 min and extension at 68 ℃ for 2 min followed by 25 cycles of denaturation at 95 ℃ for 30 sec, annealing at 50 ℃ for 30 sec and extension at 68 ℃ for 5 min. A final extension of 68 ℃ for 7 min was conducted and PCR products were stored at 4 ℃. In order to visualize the amplifications, PCR products were diluted with loading dye (0.025% (w/v) bromophenol blue, 30% (v/v) glycerol in a 1:6 ratio. These samples were analyzed by loading onto a 1.5% (w/v) agarose gel (Saarchem-Merk, South Africa) containing 0.5 ng/ml ethidium bromide, in 1X TAE buffer consisting of 0.04M Tris-acetate and 1 mM EDTA, adjusted to pH 8 and electrophoresis at 4-10V/cm in a HE33 Mini Horizontal Agarose submarine unit (Hoefer Scientific Systems, USA). A 1kb DNA ladder (Fermentas Life sciences,

Lithuania) was used as a standard marker to assess the molecular range of the separated bands. The gel was visualized at 312nm on a Spectroline TC-312A UV transilluminator (Spectronics Corporation, USA) since the ethidium bromide intercalates DNA and emits fluorescence at a wavelength of 300nm (Karcher. 1995).

2.2.4.2 o*gtpch*I

The o*gtpchI* gene was amplified from the pET46-Ek-LIC vector (Prof Hyde and Dr Sims, University of Manchester, UK). For a comparative analysis of protein expression, the gene was sub-cloned using the primers mentioned in section 2.2.2. The o*gtpchI* gene was isolated using the High pure plasmid isolation kit (refer to section 2.2.8.2). DNA concentration was assessed by spectrophotometry at A_{260nm} (section 2.2.3) The PCR reactions consisted of 2 ng/µl plasmid DNA, 0.4 pmol/µl each of oR and oF, 0.0125 U/µl of Takara *ExTaq* DNA polymerase (TAKARA Bio Inc, Japan), Takara *ExTaq* DNA polymerase buffer (0.2 mM Mg^{2+}, Proprietary mix) and 0.02nmol/µl of each dNTP in a final volume of 25 µl. The use of *ExTaq* DNA polymerase for amplification is due to the combination of *Taq* polymerase that infers thermostability and *Pfu* polymerase that possess 3' to 5' exonuclease activity (proof reading activity) for the amplification of longer target genes. The DNA polymerase also adds adenine nucleotides to the 3' ends of the insert, which improves the efficiency of ligation of the insert into the plasmid therefore allowing for A/T cloning since recircularization of the vector (has T overhangs) is prevented and compatible A-overhangs are supplied for the insert. The PCR profile consisted of 2 min denaturing step at 95 ℃, 5 cycles of denaturation at 95 ℃ for 30 sec, annealing at 41 ℃ for 1 min and extension at 72 ℃ for 2 min followed by 25 cycles of denaturation at 95 ℃ for 30 sec, annealing at 51 ℃ for 1 min and extension at 72 ℃ for 2 min. A final extension of 72 ℃ for 5 min was conducted and PCR products were stored at 4 ℃.

2.2.5 Cloning

2.2.5.1 The pGEM® T-Easy vector system

The n*gtpch*I and o*gtpch*I inserts contained 3' A-overhangs, which were generated by the polymerases. This facilitated A/T cloning into the pGEM®-T Easy vector (refer to Appendix B for vector map) according to the manufacturers protocol (Promega, USA). The purified A-tailed *gtpch*I PCR products and pGEM®-T Easy vector were combined in a molar ratio of 3:1. The amount (ng) of insert required was determined by the equation:

$$\text{ng of insert} = \left[\frac{\text{ng of vector x kb size of insert}}{\text{kb size of vector}}\right] \times \frac{3}{1}$$

In a final volume of 10 µl, 0.3 Weiss units/µl of T4 DNA ligase and 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 (5 ng/µl) and the A-tailed PCR product (1.25 ng/µl n*gtpch*I or 1.2 ng/µl o*gtpch*I). This reaction was incubated at 16 ℃ for 16 hrs. The ligation reaction was inactivated at 65 ℃ for 20 min (see section 2.2.6.1 and 2.2.7.1 for the heat shock method) or the ligated plasmids precipitated using ethanol (see section 2.2.6.2 and 2.2.7.2 for the electroporation method). Both the heat shock and the electroporation methods were assessed.

2.2.6 Preparation of competent cells

2.2.6.1 Calcium chloride heat shock cells

E. coli cells in general have to be made competent for foreign DNA uptake since these cells are not naturally transformable. In this case the divalent cation, calcium is capable of altering the permeability of *E. coli* cell envelope to form channels allowing DNA import. The *E. coli* cell line DH5α [genotype F-F80 d*lac*z DM15 D(*lac*ZYA –argF) U169 *deo*R *rec*A1 *end*A1 *hsd*R17 (rK+, mK+) *pho*A *sup*E44 l *thi*-1 *gyr*A96 *rel*A1] (Gibco BRL, Life Technologies, USA) was inoculated into 10 ml of Luria Bertania (LB)-Broth (1% Tryptone, 0.5% Yeast Extract, 1% NaCl (w/v) pH 7) and grown overnight at 37 ℃ with shaking at ~150-200 rpm. 5 ml of the *E. coli* cell culture was diluted 10-fold in LB-broth and grown at 37℃ until OD_{600nm} = 0.4 (4x 10^7 cells/ml). The cells were placed on ice for 10-20 min, to cool them down and thereafter centrifuged at 3000xg in a Beckman Avanti J-25 centrifuge with fixed angle rotor for 30 min at 4 ℃ to pellet the cells. The cell pellet was resuspended in 2.5 mM ice cold sterile $CaCl_2$ since the negative charges of incoming DNA and polyanions of the bacteria's outer surface repel each other and the $CaCl_2$ aids to neutralize the interaction. The centrifugation step was repeated and the cell pellet resuspended in a combination of 0.25 mM ice cold sterile $CaCl_2$ and 15% (v/v) cold glycerol. The competent cells were incubated on ice for 1 hr and 90 µl of aliquots were then snap frozen in liquid nitrogen.

2.2.6.2 Electrocompetent cells

This procedure involves subjecting *E. coli* cells to lowered temperatures making the cells competent for foreign DNA uptake. DH5α *E. coli* cell lines [genotype F-F80 d*lac*z DM15

D(*lac*ZYA –argF) U169 *deo*R *rec*A1 *end*A1 *hsd*R17 (rK+, mK+) *pho*A *sup*E44 l *thi*-1 *gyr*A96 *rel*A1] (Gibco BRL, Life Technologies, USA) was utilized for routine cloning. The first step involved the inoculation of 10 µl of *E. coli* in 10 ml of LB-broth followed by growth overnight (16 hrs) at 37 ℃ with shaking at ~150-200 rpm. The culture was thereafter diluted 100-fold into pre-warmed LB-broth. Cells were grown at 37 ℃ until OD_{600nm}= 0.5-0.6 (6 x 10^{16} cells/ml), cooled down on ice for 20 min and pelleted by centrifugation at 3000xg for 30 min at 4 ℃ (Beckman J-6 centrifuge). The cell pellet was then resuspended in 10 ml ice- cold H_2O. The proceeding steps involved the repeated washing of the pelleted cells with 240 ml ice- cold H_2O and centrifugation at 3000xg for 30 min at 4 ℃. After the supernatant was removed in the final wash step the cell pellet was resuspended in 10% (v/v) glycerol and incubated on ice for 30-60 min. Cells were centrifuged as before. The cell pellet was resuspended in 10% (v/v) glycerol. Competent cells were flash-frozen in liquid nitrogen and stored at –70 ℃ as 100µl aliquots.

2.2.7 Transformation

2.2.7.1 Heat shock method

The ligated mixtures containing the n- and o-*gtpch*I-pGEM®-T Easy plasmids and competent ($CaCl_2$) DH5α cells were incubated on ice for 30 min to stabilize the *E. coli* lipid membrane and permit greater interaction between calcium ions and the polyanions of the cell. This is followed by treating cells to 42 ℃ for 90 sec using a heating block (Thermoresearch, UK), which alters the fluidity of the cell membrane allowing DNA uptake. Cells were immediately cooled on ice for 2 min and revived (from temperature shock) by the addition of 900 µl of warm (~37–40 ℃) LB-glucose (0.0018 mM) to the cells. The cells were then grown with shaking (~150-200 rpm) for 1 hr at 37 ℃. For primary selection of recombinant clones in pGEM®-T Easy vector, bacterial cells were then 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, Switzerland) and for secondary selection, 40 µg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactoside (Xgal) (Promega, USA) and 0.2 mM of IPTG were added. Secondary selection involves blue-white screening to identify the recombinant clones when using the pGEM®-T Easy vector. This method proved to be effective since it involved the insertional inactivation of the *lac*Z gene in the presence of the insert and therefore the formation of white colonies. The *lac*Z gene as such, encodes for the enzyme β-galactosidase. In the absence of the insert, β-

galactosidase uses X-gal as a synthetic product and IPTG as an inducer resulting in the formation of blue colonies.

2.2.7.2 Electroporation method

To precipitate the plasmid DNA, 1/10 of 0.3 M sodium acetate (pH 5.3) and 3 volumes 100% ethanol (Merck, Germany) was added and incubated at –70 ℃ for a minimum of 2 hrs. The low pH of the sodium acetate buffer and the ethanol allows for the plasmid DNA to precipitate from the unwanted cell debris. The pellet containing the precipitated plasmid was thereafter washed with 70% ethanol and dried *in vacuo* to remove any trace ethanol, which would hinder the transformation of the plasmid into the heat shock- or electro-competent prepared bacterial cells. The dried plasmid was dissolved in 10 µl of dddH_2O, half of which was transformed into thawed (on ice) electrocompetent-competent DH5α *E. coli* cells (refer to section 2.2.6.2). Transformation of competent cells was performed as follows: the electrocompetent *E.coli* DH5α cells were thawed on ice, to which 1.1 ng/µl of precipitated plasmid DNA was added and transferred to cold electroporation curvettes (Biorad, USA). Electroporation was carried out at 2000V for 5ms. To revive the bacterial cells 1000 µl of warm (~37 –40 ℃) LB-glucose (0.0018 mM) was added to the cells. The cells were then grown with shaking (~150-200 rpm) for 1 hr at 37℃. Cells were plated out for primary and secondary selection as in 2.2.7.1. Recombinant clones were selected on plates containing agar supplemented with 100 µg/ml ampicillin (Roche, Switzerland).

2.2.8 Identification of recombinant clones

2.2.8.1 Colony PCR n*gtpchl* and o*gtpchl*

Recombinant (white) clones were selected from the LB-broth agar plates and grown overnight (16 hrs maximum) at 37 ℃ with shaking (~150-200 rpm) in 10 ml LB-broth supplemented with 50 µg/ml ampicillin (Roche, Switzerland). Glycerol stocks (using 50% glycerol) were made of the selected clones and stored at -70 ℃. From the bacterial culture, 1 µl was utilized for a PCR reaction. PCR reactions were performed and analyzed (gel electrophoresis) as in section 2.2.4.1 for the n*gtpchl* gene and section 2.2.4.2 for the o*gtpchl* gene. Recombinant clones containing the insert were inoculated into 1x LB-broth pH 7.5 supplemented with 50 µg/ml ampicillin (Roche, Switzerland) and grown overnight at 37℃ with shaking (~150-200 rpm). The n*gtpchl*- and o*gtpchl*-

pGEM®-T Easy plasmids were isolated using the Zyppy™ Plasmid Miniprep kit (Zymoresearch ™, U.S.A).

2.2.8.2 Plasmid isolation

The Zyppy™ Plasmid Miniprep kit was utilized according to the manufactures instructions in order to obtain high yields of pure plasmid DNA. Of the overnight culture (refer to section section 2.2.8.1), 3 ml was pelleted at 14100xg at 22 ℃ for 30 sec in a Microlitre Centrifuge (Hermle, Germany). The supernatant was discarded and the pellet resuspended in 600 µl dddH_2O. To the resuspended pellet, a 7x lysis buffer (lysozyme, EDTA and sodium hydroxide) was added to shear the bacterial cell wall since the buffer contained lysozyme. These was followed by the immediate addition of the neutralization buffer (chaotropic reagents, RNAse A) and complete mixing by inverting the tube to ensure the cells do not lyse excessively to denature the plasmid DNA. The unwanted DNA is denatured by the high pH of the buffer, the plasmid DNA however stays in the aqueous phase. The genomic DNA and contaminating proteins form insoluble globules with the salts and EDTA, which are present in the pellet when the suspension was centrifuged at 14100xg for 5 min. The pellets were discarded and the supernatant containing the plasmid DNA was placed into a Zymo-Spin™ IIN column. The DNA is bound to silica in the presence of the chaotropic salt, guanidine-HCl. Chaotropic salt solutions and ethanol-containing buffers (Endo-wash buffer and Zyppy™ wash buffer) are then used in several wash steps to remove contaminating bacterial components. Plasmid DNA was eluted in 40 µl Zyppy™ elution buffer (10 mM Tris-HCl, pH 8.5 and 0.1 mM EDTA) and stored at –20 ℃. Thereafter, DNA concentration was assessed as mentioned in section 2.2.3.

2.2.8.3 Screening of recombinant clones

Restriction enzyme digestions were utilized for screening the recombinant clones from colony PCR that were tested as “positive” (2.2.8.1) of n*gtpch*I- and o*gtpch*I-pGEM®-T Easy plasmids. Digestion reactions were each set up in a total volume of 10 µl and consisted of 5 ng/µl plasmid DNA, 1 U/µl *EcoR*I and a combination of 1 U/µl *Nde*I and *Bam*HI (Promega, USA). Using Buffer H (pH 7.5, 90 mM Tris-HCl, 10 mM $MgCl_2$, and 50 mM NaCl) (Promega, USA) for *EcoR*I and Buffer D (pH 7.9, 6 mM Tris-HCl, 6 mM $MgCl_2$ and 150 mM KCl) (Promega, USA) for *Nde*I and *Bam*HI. The reactions were incubated at 37 ℃ for a minimum of 3 hrs and stopped by heat inactivation at 65 ℃ for 20 min or

by loading onto a 1.5% (w/v) agarose gel (Merck-Saarchem, South Africa) containing 0.5 ng/ml ethidium bromide.

2.2.9 Sub-cloning into the pET22b vector system

2.2.9.1 Restriction enzyme digestion

Plasmids (n*gtpch*I- and o*gtpch*I-pGEM®-T Easy) that were screened positive in section 2.2.8.3 were used for large-scale restriction enzyme digestions for the release of n*gtpch*I- and o*gtpch*I inserts from the pGEM®-T Easy vector to ligate with pET22b expression vector (Novagen, Germany. Refer to Appendix C for vector map). Large-scale digestions were performed on pET22b vector. Reactions consisted of the restriction enzymes *Nde*I and *BamH*I (Promega, USA). The digestion reactions were set up in a total volume of 40 µl, consisting of 0.075 µg/µl plasmid DNA, 0.25 U/µl *Nde*I and *Bam*HI and Buffer H (pH 7.9, 6 mM Tris-HCl, 6 mM $MgCl_2$, and 150 mM NaCl) (Promega, USA). Digestion was carried out at 37 ℃ for 3 hrs and inactivated by either incubation at 65 ℃ for 20 min or loading onto a 1.5% (w/v) agarose gel (Merck-Saarchem, South Africa) containing 0.5 ng/ml ethidium bromide. The gel was run and DNA was separated as described in section 2.2.4.1. The separated bands corresponding to the 1181 bp n*gtpch*I & o*gtpch*I inserts and 5405 bp pET22b expression vector were recovered from the agarose gels using a high pure DNA clean up protocol.

2.2.9.2 Cloning into pET22b

The SV gel clean/ PCR Product purification kit (Promega, USA) was utilized according to manufacturers instructions for high pure DNA clean up. The excised DNA fragment (100 mg) was dissolved in 3 volumes of binding buffer (4.5 M guanidine-thiocyanate and 0.5 M potassium acetate pH 5) at 55 ℃. This chaotropic salt has a high positive charge, ensuring a strong enough electrostatic interaction to occur so that the wash buffer that is used eliminates only contaminating primers, salts, and unwanted nucleotides that will affect downstream interactions. The solution containing the completely dissolved agarose gel slices were separated on a High Pure filter column and washed twice using a wash buffer (10 mM potassium acetate pH 5, 80% ethanol, 16.7µM EDTA pH8) with centrifugation (Microlitre Centrifuge Hermle Labortechnik, Germany) at 14100xg for 1 min and 5 min, respectively. The purification column contains silica, which a llowed the DNA to bind to it when guanidine thiocyanate is present. The bound DNA was then simply eluted using nuclease free H_2O (50 µl) and collected by centrifugation at 14100xg

for 1 min. Thereafter, DNA concentration was assessed as mentioned in section 2.2.3. The purified pET22b (5405 bp) expression vector and n*gtpchI* and o*gtpchI* inserts (1181 bp) were then ligated in a 3:1 molar ratio of vector to insert. The other components of the ligation reaction included 0.15 Weiss units/µl of T4 DNA ligase (Promega, USA) and T4 DNA ligase buffer (30 mM Tris-HCl pH 7.8, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP) to a final volume of 20 µl. This reaction was incubated overnight at 16℃. This ligation mixture was cloned into heat-shock competent *E. coli* DH5α cells (see section 2.2.6.1 and 2.2.7.1). Furthermore, the ligated reaction was precipitated and transformed into electrocompetent competent *E.coli* DH5α cells (see sections 2.2.6.2 and 2.2.7.2).

2.2.9.3 Screening of recombinant clones

Colony PCR (section 2.2.4.1 and 2.2.4.2), followed by restriction enzyme digestions of isolated plasmid DNA (section 2.2.8.2) were performed to screen DNA. Digestion reactions were each set up in a total volume of 10 µl and consisted of ~50 ng/µl plasmid DNA, 1 U/µl *EcoR*V (o*gtpch*I-pET22b), a combination of 1 U/µl *Xba*I and *Xho*I (n*gtpch*I-pET22b) and a combination of 1 U/µl *Nde*I and *Bam*HI (Promega, USA) (n*gtpch*I- and o*gtpch*I-pET22b) using Buffer D (pH 7.9, 6 mM Tris-HCl, 6 mM $MgCl_2$ and 150 mM KCl) (Promega, USA) for *Nde*I and *Bam*HI. The reactions were incubated at 37 ℃ for a minimum of 3 hrs and stopped by heat inactivation at 65 ℃ for 20 min or by analyzing onto a 1.5% (w/v) agarose gel (Merck-Saarchem, South Africa) containing 0.5 ng/ml ethidium bromide.

2.2.10 Codon harmonization of *P. falciparum gtpch*I

The *P. falciparum* n*gtpch*I gene sequence (1170bp) was downloaded from PlasmoDB (PFL1155w). Codon harmonization of the n*gtpchI* gene was achieved using a pseudo code that forms part of the codon harmonization algorithm (/http://www.sami.org.za/equalize///) for the conversion of the coding sequence (as a text box file) to a harmonized sequence. The initial process involves importing the sequence from the web-form textbox to a string variable. The sequence is thereafter separated into codons and saved in an array. An array consists of the codons assigned as elements. For e.g. the n*gtpch*I sequence atgtataaatatacg is an array with five elements (referred to as numbers enclosed within brackets), ATG= Array(0), TAT= Array(1), etc. Using the selected source (in this case *P. falciparum*) codon preference table the codon preference for the array element is determined (e.g. for Array(1) TAT the codon

preference in *P. falciparum* is 0.89). The target (*E. coli*) codon preference of all codons that encode the same amino acid as those of the *P. falciparum* is then determined from the selected *E. coli* codon preference table (e.g. for Array(1) the codon preference for TAT is 0.25). The codon from *E. coli* that has the codon preference mathematically similar to that of the *P. falciparum* codon is selected to form part of the h*gtpch*I sequence (e.g. TAC in *E. coli* has a codon preference of 0.75 that encodes for the same amino acid, Tyr). Codon preference table as well as pause site predictions were obtained from www.kazusa.or.jp/codon. The pause sites were predicted using an algorithm with adjustable parameters allowing rigidity of prediction. Individual codons were scored by normalizing all codons encoding AA to 1 and the cut-off of pause site propensity set as an adjustable parameter.

Prior to the prediction of RNA structure, the harmonized *Pf gtpch*I (h*gtpch*I) sequence had to undergo minor alterations. These minor alterations were made at http://www.sami.org.za/equalize///. This included the modification of the Shine Dalgarno sequence at position 318 CAG (GAAGAC) to CAA (GAAGAC) since this would result in the formation of truncated protein. For in-frame cloning into the expression vector the 5' *Nde*I (CATATG) and 3' *Bam*HI (GGATCC) were added to the 5' and 3' regions of the gene sequence, respectively and the stop codon altered to codon, AGA coding for Lys. Based on the analysis no restriction sites pertaining to the *Bam*HI and *Nde*I restriction sites were found within the gene sequence. The gene was thereafter analyzed in bioedit sequence alignment editor version 7.0.5.3. (Copyright © 1997-2005, T. Hall) to ensure the altered gene would clone within frame of the expression vector to allow for the expression of a tagged protein (for protein purification). The altered gene was thereafter analyzed for the presence of pause sites and their effect on RNA structure before synthesis.

2.2.10.1 RNA structure predictions and final alterations to h*gtpchI*

RNA secondary structure can be assessed experimentally, thermodynamically and comparatively. For these studies, the thermodynamic route was taken. RNA structure predictions generated with the RNA structure version 4.5 tool (Mathews, Zuker & Turner. 1996-2007) were displayed as the mountain plot representations (Hogeweg and Hesper 1984; Konings and Hogeweg 1989). For the predictions of secondary structure, accepted methods are based on dynamic programming that uses free energy

minimization with computer algorithms. The nearest neighbor model is utilized to predict the free energy change at 37 ℃. It is said that for a given RNA structure at equilibrium, there is a balance between strands folded in structure S1, and the random unstructured coil state, RC. The equation for this equilibration state is governed by K_1= [S1]/[RC]. Additionally, the free E change signifies the difference in stability of two structures. Thus, the lowest free energy state, at equilibrium typifies the most represented conformation. The free energy change calculated for each structure is dependent on the sequence identity and the most adjacent base-pairs (Mathews. 2006). These approaches ultimately allow for the computation of sub-optimally folded RNA structures that may not represent the true base-pairing relationship, since various RNA structures can be generated with similar free energies. A combinatorial approach relies on creating all probable helices, which can be formed from the sequence and thereafter combining them into optimal structures. Empirical base-pairing, stacking and loop energies are used to assign sequences that could form helices and loops with the lowest free energy (Gaspin and Westhof 1995).

These values are then simplified by setting a reference level, where the value can be either 0 or 1 i.e. true or false (if the value is above 1 than it is taken as 1 and 0 remains 0) using Microsoft Excel IF logical test equation and an average of 17 values for each codon position was taken into account.

Final alterations were made to the repeats of the AAA codon from position 22-25, which codes for a Lys-rich stretch. The nucleotide at codon position 23 was synonymously replaced by AAG altering the codon usage frequency from 0.74 to 0.26 since this was classified as a low complexity region that would form a very stable mRNA structure that would affect the translation process. The h*gtpch*I gene was thereafter synthesized by GenScript CorporationTM (U.S.A). The gene was cloned into pUC 57 using *EcoR*V and maintained in *E. coli* TOP 10 cells for transportation.

2.2.11 Sub-cloning into the pET22b vector system

2.2.11.1 Restriction enzyme digestion

The *hgtpch*I was directly sub-cloned into the pET22b expression vector instead of the pGEM®-T Easy vector. This included the release of the h*gtpch*I insert from the h*gtpch*I-pUC57 (using the restriction enzymes that were designed to flank the insert) and

thereafter sub-cloning of h*gtpch*I into the digested pET22b vector. Large-scale digestion of pET22b vector and of h*gtpch*I-pUC57 reactions consisted of the restriction enzymes *Nde*I and *BamH*I (Promega, USA) and was performed as in section 2.2.9.1. The separated bands corresponding to the 1181 bp h*gtpch*I insert and 5405 bp pET22b expression vector were recovered from the agarose gels using a high pure DNA clean up protocol (section 2.2.9.2).

2.2.11.2 Cloning into pET22b

The purified pET22b (5405 bp) expression vector and h*gtpchI* insert (1181 bp) were ligated in a 3:1 ratio using KAPA rapid ligase system (KAPA biosystems, USA) according to the manufacturers instructions. However, instead of a 5 min ligation period this was extended to 16 hrs at 16 ℃ or 40 hrs at 4 ℃. KAPA rapid ligation buffer (15 mM Tris-HCl pH 7.8, 5 mM $MgCl_2$, 5 mM DTT, 0.5 mM ATP and 5% (v/v) polyethyleneglycol) were added to the vector (5 ng/µl) and the h*gtpch*I insert (4.3 ng/µl) in a final volume of 20 µl. The ligated reaction was transformed into electrocompetent *E.coli* DH5α cells and plated out for primary selection as in section 2.2.6.2 and 2.2.7.2.

2.2.12 Identification of recombinant clones

2.2.12.1 Colony PCR h*gtpch*I

To verify selection of a positive recombinant clone, plasmids were isolated of the clones picked from the LB-broth ampicillin agar plates. Colony PCR using KAPA HiFi DNA polymerase (KAPA Biosystems, USA) was used to confirm the presence of the h*gtpchI* insert in the recombinant clones (positive for containing h*gtpch*I-pET22b). h*gtpch*I was amplified using the primers designed for the pET22b expression vector (refer to section 2.2.2). The PCR contained 0.2 pmol of hF and hR, 5x buffer with $MgCl_2$, 0.2 mM of each dNTP, 0.5 µl of culture and 0.02 U/µl of KAPA HiFi DNA polymerase in a final volume of 50 µl. The PCRs were performed in a Gene-AMP PCR System 9700 (Perkin Elmer, Coneticut, USA) in 0.2 ml thin-walled PCR tubes. The PCR cycling profile consisted of a 2 min denaturing step at 95 ℃ and 30 cycles of denaturation at 98 ℃ for 20 sec, annealing at 56 ℃ for 30 sec and extension at 72 ℃ for 2 min and a last cycle of 72 ℃ for 7 min. PCR products were stored at 4 ℃. The PCR products were electrophoresed on a 1.5% agarose gel (Merck-Saarchem, South Africa) using 1kb DNA ladders (Fermentas Life sciences, Lithuania) as a standard to assess molecular sizes of the

separated bands. Cultures of the recombinant clones were used for plasmid isolation using the high pure plasmid isolation kit (refer to section 2.2.8.2).

2.2.12.2 Screening of recombinant clones

Recombinant clones containing the plasmid DNA of h*gtpch*I-pET22b were screened by restriction enzyme digestion. *Eco*RV restriction enzyme sites within the pET22b vector and the h*gtpch*I insert were used to screen for the presence of recombinant clones. Digestion reactions were each set up in a total volume of 10 µl and consisted of 50 ng/µl plasmid DNA, 1 U/µl *Eco*RV (Fermentas) and Buffer R (pH 8.5, 10 mM Tris-HCl, 10 mM $MgCl_2$, 0.01 mg/ml BSA and 100 mM KCl) (Fermentas Life sciences, Lithuania). The reactions were incubated at 37 ℃ for a minimum of 3 hrs and stopped by heat inactivation at 80 ℃ for 20 min or by loading onto a 1.5 % (w/v) agarose gel (Merk-Saarchem, South Africa) containing 0.5ng/ml ethidium bromide.

2.2.13 Nucleotide Sequencing

The sequencing method used is based on the principles of (Sanger and Coulson 1975). It involves the incorporation of the dideoxynucleotides (ddNTPs) in addition to the normal nucleotides resulting in chain termination due the presence of the hydrogen group on the 3' carbon instead of a hydroxyl group. Turner *et al.* (2000) devised a method based on this principle, which entails incorporation of fluorescent ddNTPs to terminate the template, as the ddNTPs act as chain terminators. All four ddNTPs are conjugated to different fluorescent dyes thus allowing fluorescence to be read at different wavelengths and permitting the use of a single tube for a single sequencing reaction. Sequencing reactions were performed on recombinant plasmids of n*gtpch*I-pET22b, o*gtpch*I-pET22b and h*gtpch*I-pET22b isolated Zyppy™ Plasmid Miniprep kit (as in section 2.2.8.2). Sequences were verified by automated DNA sequencing using primers directed against the T7 promoter and T7 terminator sites on the pET22b vector (Promega, USA). Single-primer reactions were set up containing 670 ng pure plasmid DNA as template, 5x sequencing buffer (proprietary mix) (Perkin Elmer, USA), Terminator Big-Dye Ready reaction mix (proprietary mix) version 3.1 (Perkin Elmer, USA) and 5 pmol/µl T7 promoter or T7 terminator primer. The sequencing cycling profile consisted of an initial denaturation step at 94 ℃ for 2 min, 25 cycles of denaturation at 96 ℃ for 10 sec, annealing at 50 ℃ for 5 sec and extension at 60 ℃ for 4 min. The sequencing reactions were performed on a GeneAmp 9700 thermocycler (Perkin Elmer, USA).

The sequencing products were purified by means of DNA precipitation to remove all contaminants such as dNTPs, primer, and protein. Sequencing reactions were treated with 2.5 volumes of ice-cold 95% (v/v) ethanol and 1.5 M NaOAC, pH 4.6 followed by centrifugation at 16000xg for 25-30 min at 4 ℃. The pellet obtained by centrifugation was then washed with 12.5 volumes of 70% (v/v) ethanol. Centrifugation was repeated and the DNA pellet dried at 50 ℃ using a heating block in the dark to protect the light sensitive dyes. Sequencing was performed on an ABI PRISM 3100 capillary sequencer (Perkin Elmer, USA). The sequences were analyzed with BioEdit version 5.0.9 software (Hall 2001, North Carolina State University, USA) by alignment with expected *gtpch*I-pET22b sequences.

2.2.14 Protein expression:

2.2.14.1 *E. coli* cell lines

BL21 Star (DE3) (Invitrogen, USA) B strain F- *ompT hsdS (rB –mB -) gal dcm rne131*λ (DE3) 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 cause proteolysis during purification. Thus, the lack of these two proteases reduces the degradation of heterologous proteins expressed. These cell lines also contain a λDE3 lysogen that carries the gene for T7 RNA polymerase under control of the *lacUV*5 promoter. 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. The M15pREP4 (Qiagen, Germany) has a higher level of lac repressor therefore allowing for the greater induction of the lacUV5 promoter by the addition of isopropyl-β-D-thiogalactoside (IPTG). Furthermore, the strain is unable to metabolize certain sugars. However, it is compatible with plasmids containing the ColE1 origin of replication i.e. pET22b vector. The pET22b vector system (Novagen, EMD Biosciences, Germany) was utilized for protein expression by activating the *T7lacUV5* promoter with the addition of IPTG. This allowed for the production of C-terminally His_6-tagged fusion proteins thus enabling for purification using a Nickel ion affinity column (Sigma, Germany). Additionally, the overproduction of proteins (possibly toxic to the cell) is prevented since protein expression from the pET22b plasmid is tightly regulated.

Heat shock competent BL21 star (DE3) and M15pREP4 cells were prepared according to sections 2.2.6.1 and 2.2.7.1.

Table 2-4: Vector-construct, host expression lines and annotations utilized in text. * Note the expression cell lines were each used as a host strain during nGTPCHI, oGTPCHI and hGTPCHI protein expression.

Vector	* Expression cell lines	Annotation in text
pET22b	M15pREP4	m15
	BL21 star (DE3)	star
	BL21 star (DE3)-*Pf*Hsp70	star 70

2.2.14.2 Protein expression

$CaCl_2$ competent expression hosts, star and m15 were freshly transformed with plasmid containing the h*gtpchI*, o*gtpchI* and n*gtpchI* gene using a heatshock method (refer to section 2.2.7.1). Briefly, 1.1 ng/µl of plasmid DNA was added to the heatshock competent star cell line followed by consecutive cooling and heating of the cells to induce the uptake of the foreign DNA. For recombinant plasmid selection, cells containing the plasmid DNA (transformed cells) were then plated onto 2x yeast tryptone (3.2% tryptone, 2% yeast extract and 1% NaCl) (YT)-agar plates (2% w/v) supplemented with 50 µg/ml ampicillin for star cells or a combination of 50 µg/ml ampicillin and 25 µg/ml kanamycin for m15 cells since this strain is kanamycin resistant. YT-agar plates were grown at 37 °C for 16 hrs. Transformed cells were inoculated into 10 ml YT media containing in a molar ratio of 1/100, supplemented with the required antibiotics (Roche, Switzerland) and grown overnight at 37 °C with shaking (~150-200 rpm). The overnight cultures were subsequently diluted 20-fold in YT-amp (star cells) or YT-amp-kan (m15 cells) and grown at 37 °C with shaking (~150-200 rpm) until an OD_{600nm} of 0.4 was reached. Un-induced samples (2 ml each) were collected 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 then carried out at 37 °C for 4 hrs and samples were collected 1 hr, 2 hrs and 4 hrs after IPTG induction.

2.2.14.3 Cell membrane disruption and protein isolation

After 4 hrs of expression, the 500 ml bacterial cultures were harvested in centrifuge bottles by centrifugation (Beckman J-6 centrifuge) at 5000xg for 15 min at 4 °C. The

pellets were suspended and maintained in wash buffer (16.7 mM Na_2HPO_4/NaH_2PO_4, pH 8, 0.1 M NaCl), which is the buffer that was utilized to wash the Ni^{2+} ion affinity column during protein purification. A 0.1 M phenylmethylsulphonyl fluoride (PMSF) and 10 mg/ml lysozyme (Roche, Switzerland) was added to the wet pellet of m15 cells as these cells were prone to proteases. PMSF is a serine protease inhibitor that is frequently used in protein solublization in order to inhibit proteases from digesting the target protein after cell lysis. Lysozyme damages bacterial cell walls by catalyzing the hydrolysis of 1,4-β-bonds between N-acetylmuramic acid and N-acetyl-D-glucosamine residues in a peptidoglycan. This therefore prevents excess sample viscosity by releasing chromosomal DNA from the bacterial cells. The wet pellets were then aliquoted into 1 ml eppendorf tubes to promote bacterial cells shearing and placed in liquid nitrogen and thawed overnight on ice. Protease inhibitors were not added, as the star cell lines are deficient in the *lon* protease and the *ompT* outer membrane protease. Sonication relies on the application of ultrasound energy to agitate particles. It is capable of disrupting cell membranes to release cellular contents when used in conjuction with lysozyme. Sonication parameters included setting the duty to 30%, output to 3 and thereafter sonicating each 1 ml fraction for 20 sec twice on ice, with a rest period on ice in between each sonication cycle. A 1 ml aliquot was stored for total protein analysis and the rest of the aliquots were centrifuged at 16000xg for 20 min (4℃). The resulting pellet was analyzed as the insoluble fraction and the supernatant as the soluble fraction. Total fractions refer to the soluble and insoluble proteins together before centrifugation.

2.2.14.4 Co-expression: Molecular chaperone proteins

The ability of chaperones to assist protein folding was exploited in the following co-expression study to optimize expression of the GTPCHI protein. The pMRBAD vector containing the *Pf*hsp70 was obtained from Dr Stephens (University of Rhodes, South Africa) (Appendix G). Hsp70 functions in assisting the folding process of newly synthesized proteins, including the refolding of proteins that have denatured. By binding to the exposed hydrophobic regions of the nascent protein, these areas of the protein stabilize and consequently folds into its tertiary configuration (Matambo *et al.* 2004). The *Pf*hsp70 gene is situated downstream of the araB promoter as such it is induced with L-Arabinose. Induction with arabinose allows for the chaperone protein to be induced separately from the target protein, which is induced with IPTG (isopropyl galactopyranoside). The strategy for co-expression was based on the expression of

*Pf*hsp70 followed by the target protein to ensure that sufficient chaperone proteins existed before the expression of the target protein, allowing correct folding of this protein. The strategy of co-expression was based on transforming star cells first with pMRAD-hsp70, making these cells competent. Thereafter the n*gtpch*I-, o*gtpch*I- and h*gtpch*I-pET22b constructs were transformed into the heat shock competent cells.

For this, 1 ng/µl of pMRAD plasmid containing the *Pf*hsp70 was transformed using the heatshock method into $CaCl_2$ competent star (DE3) cells and incubated for 1 hr in YT-glucose. These cells were then plated out onto 2x YT agar plates (2% w/v) supplemented with 50 µg/ml kanamycin. A 1/1000 dilution of these cells into 2x YT broth containing 25 µg/ml kanamycin was thereafter grown at 37℃ for 16 hrs. A 1/100 dilution of these cells were inoculated into 2x YT media supplemented with 25 µg/ml kanamycin and made competent using the $CaCl_2$ based method (see section 2.2.6.1). Instead of using the LB-broth, 2x YT-media was utilized. 1.1 ng/µl of pET22b plasmid containing the h*gtpch*I, o*gtpch*I and n*gtpch*I gene were transformed using the heatshock method (see section 2.2.7.1) into $CaCl_2$ competent star 70 *E.coli* cells and incubated for 1 hr in 2x YT-glucose (2 mM). These cells were then plated out onto 2x YT agar plates supplemented with both 100 µg/ml ampicillin and 50 µg/ml kanamycin. A 1/100 dilution of these cells into 2x YT-media containing 50 µg/ml amp and 25 µg/ml kanamycin was thereafter grown at 37 ℃ for 16 hrs. A second dilution of 1/50 (5ml of transformed overnight culture: 250ml of 2x YT-amp-kan) and grown once again at 37 ℃ until cells reached an OD_{600nm} 0.2. Thereafter, L-arabinose was added to the cultures and incubated at 37 ℃ until OD_{600nm} 0.4 is reached, this was followed by induction of the target protein with IPTG. Proteins were thereafter isolated as indicated in section 2.2.14.3. The resulting pellet was analyzed as the insoluble fraction and the supernatant as the soluble fraction. Total fractions refer to the soluble and insoluble proteins together before centrifugation. Note the m15 cells could not be transformed with the pMRBAD vector containing the *P. falciparum* chaperone since antibiotic selection of the cell line and the pMRBAD plasmid were both based on kanamycin resistance (refer to Appendix G for the vector map of pMRBAD- *Pf*hsp70).

2.2.14.5 Auto-induction media

In bacterial hosts such as the BL21 (DE) cell lines the coding sequence for the target protein is under the control of the T7 promoter, which is identified by the specific T7 RNA

polymerase. However, a disadvantage with the use of this system is the overactivity of the T7 RNA polymerase even in the absence of the inducer. It was discovered that in the absence of the inducer bacterial cultures grown in complex media would induce the production of target protein resulting in the stressed condition of the cells. The usage of auto-inducing media for the growth of T7 expression strains was therefore initiated. The point was to develop non-inducing media, which would allow the T7 expression bacterial strains to remain viable until saturation and the auto-inducing media would account for the high-density, induced (to completion) cultures (Studier 2005).

n*gtpch*I-pET22b, o*gtpch*I-pET22b and h*gtpch*I-pET22b plasmids were transformed as in sections 2.2.7.1 into *E. coli* star 70 and m15 cells respectively. A 1/100 dilution of the inoculated cells were made into 10 ml non-inducing media (refer to Table 2.5 for media content) containing 50 µg/ml ampicillin and 25 µg/ml kanamycin (Roche, Switzerland) and grown overnight at 37℃ with shaking (~150-200 rpm). The overnight non-inducing cultures were stored as 1 ml stocks in 10 % (v/v) glycerol concentration at -70℃. The overnight cultures were subsequently diluted 50-fold in complex auto-inducing media (refer to Table 2.5) supplemented with a combination of both 50 µg/ml ampicillin and 25 µg/ml kanamycin and grown at 37℃ with shaking (~150-200 rpm) for 48 hrs. After every 16 hrs cultures were re-supplemented with the required antibiotics. Proteins were thereafter isolated as indicated in section 2.2.14.3. The resulting pellet was analyzed as the insoluble fraction and the supernatant as the soluble fraction. Total fractions refer to the soluble and insoluble proteins together before centrifugation.

Table 2-5: Composition of media used for auto induction studies (Studier. 2005 and Human. 2007).

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)

2.2.15 Protein purification

2.2.15.1 Affinity chromatography

The GTPCHI protein present in the soluble protein fractions was purified using affinity chromatography. Protein purification was conducted using immobilized metal ion affinity chromatography with a His select HC nickel affinity gel (Sigma-Aldrich, St.Louis, USA) that has specificity for proteins containing histidine. Purification is based on an electrostatic interaction between the His_6 peptide tag attached to the solubilized protein and Ni^{2+} cations attached to the resin. Approximately 2-3 mg/ml of the soluble protein fraction (the supernatant) was added to the equilibrated Ni-affinity gel (50% gel in 30% ethanol) with a bed volume of 2 cm^3. The column however was first washed using 2 column bed volumes of $dddH_2O$. This was thereafter followed by equilibration using 2 bed volumes of the Wash buffer (50 mM Na_2HPO_4/NaH_2PO_4, pH 8, 0.3 M NaCl) followed by the flow through of the sample twice through the column. Unbound proteins were washed off and the target protein with the his_6-tag was thereafter eluted using imidazole (competitively binds to Ni^{2+} resin), which constitutes the Elution buffer (50 mM sodium phosphate, pH 8, 0.3 M NaCl, 250 mM imidazole) after unbound proteins were washed off. Total fractions were collected, followed by protein concentration determination and further analysis using SDS-PAGE of nGTPCHI, oGTPCHI and hGTPCHI.

2.2.16 Protein concentration determination

Many different methods are available for protein concentration determination. However, the Bradford method was selected for simplicity sake. This method is reliant on the electrostatic and hydrophobic interactions of the sulfonic acid groups in Coomassie Brilliant Blue G-250 dye to positively charged, aromatic, basic amino acids that occurs at low pH values. After the binding of protein the absorbance maximum of Coomassie blue dye shifts from 465nm to 595nm, which was assessed spectrophotometrically (Bradford 1976). The standard bovine serum albumin (BSA) (Promega, USA) has superior protein properties and is ideal for protein concentration determination. BSA (1 mg/ml) was weighed of once and diluted with $dddH_2O$ to achieve a standard concentration range of 200, 100, 50, 25, 12.5, and 6.25 µg/ml protein allowing the extrapolation of unknown target protein concentration from the absorbance data generated at 595nm. To 50 µl of each BSA standard (prepared in triplicate) or 1/100 dilutions of protein sample, 150 µl Quick Start Bradford Dye (Bio-Rad, USA) was added in a 96 well microtitre plates (Bibby

Sterilin Ltd., UK). The absorbance was measured after 15min incubation at room temperature at 595nm on a Multiskan Ascent spectrophotometer (Thermo Labsystems, UK). Data was thereafter analyzed and the BSA standards were represented as a standard curve (Absorbance at 595nm vs. [BSA] µg/ml).

2.2.17 SDS-PAGE analysis

A 30% acrylamide: 15% bisacrylamide (w/v) (Sigma-Aldrich, USA) gel was utilized for polyacrylamide gel electrophoresis (PAGE). This solution was diluted to constitute a 4% acrylamide-methylene-(bis) acrylamide stacking gel (pH 6.8) and a 10% acrylamide-methylene-(bis)acrylamide, separating gel (pH 8.8) (Laemmli 1970). Using the concentrations determined for the target proteins 2 µg/ml of total, insoluble and soluble protein fractions and 3 µg/ml of purified protein was suspended in 4 volumes of 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 at 95 ℃ for 5 min. For the protein expression levels as a function of time study, the amount of protein used for analysis was determined by the equation devised by Dr Stephens (Rhodes University, South Africa), (absorbance at 600nm/0.5)x 150 since this allows adequate comparison of protein formed since it accounts for cell growth (This was also loaded quantitatively using the Bradford method). A 10 x 8 cm x 1mm gel consisting of the 4 % stacking gel and thereafter 10 % separating gel 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) consisting of SDS running buffer (0.1% (v/v) SDS, 0.25 M Tris, 192 mM Glycine buffer (pH 8.3).

In order to visualize the protein bands quantitatively on a SDS-PAGE, gels were stained using the colloidal method 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 Coomassie staining 0.1% (w/v) for 24 hrs. Staining in this sense refers to the reversible or irreversible binding by the proteins of a coloured organic or inorganic chemical. Protein bands were thereafter fixed with 25% methanol and 10% acetic acid and de-stained with 25% methanol to reduce background staining (Neuhoff *et al.* 1988). It was essential to quantify proteins in order to distinguish between the difference of suspected protein expression of the nGTPCHI compared to the oGTPCHI and hGTPCHI proteins. The colloidal Coomassie method is reliant on the high concentration of ammonium

sulphate assuring that the dye alters its colloidal form therefore reducing unwanted background staining. The addition of acetic acid to the alcoholic buffer (25% methanol) modifies the equilibrium from the colloidal form to dispersed dye, which is needed to assist diffusion into the gel matrix. This diffusion assists in quicker and complete staining of the gels' cross-section (Neuhoff *et al.* 1988). Gels were scanned electronically on the Versadoc 4000 image scanner (Biorad, USA) and analyzed using the Quantity OneTM Software Package (Biorad, USA), MicrosoftTM ExcelTM format. The following parameters were used in the analysis: peak density, average density, trace quantity, and relative quantity. By determining the trace quantity and the relative quantity of a band suspected to be the target protein on the SDS-PAGE gel the protein can be quantified. The trace quantity is the area under the intensity profile curve, which is defined as the average intensity across the band width (intensity) x band height (mm). The relative quantity is defined as the trace quantity of a specific band expressed as the percentage of the total quantity of every band in the particular lane (or % intensity of the lane/ % of intensity the bands).

2.2.18 Western blotting

Western blotting is based on the transfer of proteins from polyacrylamide gels via an electric field to a "blot", which is an adsorbent membrane. In combination with the electrophoretic protein transfer, probes are utilized to interact with the transferred proteins for detection. The semi-dry transfer method with flat-plate electrodes was used in this instance and is explained as the transfer of proteins over a horizontal setup, confined to the stacks of filter paper requiring limited volumes of buffer (Gravel *et al.* 1995).

An SDS-PAGE gel containing the separated proteins was equilibrated in a 10 mM 3-(cyclohexylamine)-1-propane sulfonic acid (CAPS) buffer pH>9 for 5 min. A polyvinylidene fluoride (PVDF) Immobilon-P (nitrocellulose) membrane (Millipore, USA) was cut according to the dimensions of the gel (10x15 cm) and was equilibrated in 100% methanol (Merck-Saarchem, South Africa) for 15 sec thereafter the membrane was soaked in the CAPS buffer. Since the semi-dry protein transfer method was used, 10 pieces of 3MM Chromatography paper (Schleicher and Schuell. England) cut slightly greater in dimensions to the SDS-PAGE gel were equilibrated in CAPS buffer and 5 pieces of filter paper were layered individually on the bottom of the flat plate electrodes

(ensuring no trapped air bubbles were present). The methanol-CAPS-equilibrated PVDF membrane was then lowered onto the 5 pieces of filter paper, followed by the CAPS-equilibrated SDS-PAGE gel and the rest of the equilibrated filter paper (once again ensuring no trapped air bubbles were present). The proteins were thereafter blotted at 10V, 20-25mA (Pharmacia 3000/150 powerpack) for 45 min. Once completed the layers of filter paper were discarded and an outline of the gel was made on the membrane, which now contained the transferred proteins. It is important that the membrane containing the proteins is blocked prior to antibody exposure. This is to prevent nonspecific adsorption of probe molecules on available macromolecular binding sites of the membrane. The blocking buffer used is protein-based as such it is composed of Tris buffered saline tween-20 (TBST) (25 mM Tris, pH 7.6, 0.15 M NaCl, 0.05% Tween 20) and 25 mg/ml BSA (protein component to block available binding sites). The membrane was lowered into a pre-blocked hence equilibrated plastic bag, (10 ml of blocking buffer was used to pre-block the bag and the buffer discarded before placing the PVDF membrane in it), blocking buffer was added to the PVDF membrane and incubated at 4 ℃ for 16 hrs. The blocking buffer was discarded. Proteins are thereafter detected in this case using enzyme-linked reagents. To the PVDF membrane 10 ml of TBST and a 1/5000 dilution of His Probe (Promega, USA) was added to the PVDF membrane, air bubbles were removed and the plastic bag containing the PVDF membrane was incubated at 25 ℃ for 90 min. Based on a similar principle to the isolation of protein using the Ni^{2+} affinity column, the His Probe-HRP (horseradish peroxidase) is used also to detect poly-histidine-tagged fusion proteins. Washing the membranes with 10 ml of TBST and incubation at 25 ℃ for 10 min was performed to remove excess His Probe. The wash procedure was repeated 4 times in total. Thereafter, the blotted membranes were incubated for 5 min with SuperSignal® West Pico Chemiluminescent Substrate Working solution (Pierce, USA). This working solution consisted of equal volumes of the luminol solution and stable peroxide buffer.

A dark room was compulsory for the exposure of the membranes. The SuperSignal® West Pico Chemiluminescent Substrate Working solution was removed and the membranes were exposed onto CL-xposure film at 5 min, 15 min and 3 hrs, respectively. The films were developed in 1/10 dilution of Ilford PQ universal developer (Ilford, Germany) for 1 min at room temperature, followed by a brief rinsing in $dddH_2O$ and fixed

in a 1/5 dilution of Ilford rapid fixer (Ilford, Germany) for 3 min. The exposed film was rinsed once again and air-dried in the dark.

2.3 Results

The expression of *P. falciparum* nGTPCHI, oGTPCHI and hGTPCHI using *E. coli* as an expression host was assessed. The levels of expression of nGTPCHI, oGTPCHI and hGTPCHI were compared in order to evaluate the difference that codon harmonization of the *gtpch*I gene would have in comparison to codon optimization and the native *gtpch*I gene. Furthermore, the expression conditions were altered for optimized GTPCHI expression.

2.3.1 Primers

Primers were designed to amplify the n*gtpch*I gene from genomic DNA of *Pf*3D7 strains and to amplify the o*gtpch*I gene from the o*gtpch*I-pET46-Ek-LIC vector. Primers were designed for the h*gtpch*I gene but only for screening (colony PCR) and nucleotide sequencing purposes. These are indicated in Table 2-6. The forward and reverse primers contain the *Nde*I (CATATG) and *Bam*HI (GGATCC) restriction sites, respectively and two nucleotides TG were added to the reverse primer for directional cloning into the pET22b expression vector and for the incorporation of the His_6 tag.

Table 2-6: Primers used for the sub-cloning of GTPCHI domains for expression of proteins in the pET22b His-tag expression system.

Primer	Sequence (5'-3')	Restriction enzyme sites	*Tm (°C)	Gene specific Tm (°C)
nF	CA↓TATG TAT AAA TAT ACG TCA ATA AAC	*Nde*I	54.34	46.16
nR	G↓GA TCC TGC TTA TTT AAA TTT TC	*Bam*HI	53.51	31.50
oF	CAT↓ATG TAC AAA TAC ACC TCT ATC	*Nde*I	51.73	52.1
oR	G↓GA TCC TGT TTG TTC AGG TTT TCA AC	*Bam*HI	53.51	45.5
hF	CAT↓ATG TAC AAA TAC ACG TCT ATC	*Nde*I	55.89	N/A
hR	G↓GA TCC TGT CTG TTC AGG TTT TC	*Bam*HI	60.65	N/A

- *Tm equation=69.3+0.41(%GC)-650/length (Rychlik *et al.* 1990).
- Gene-specific Tm= Tm calculated for the specific gene region of the primer, excluding restriction enzyme sites and additional bases.
- N/A i.e. not applicable since gene-specific PCR cycling was not performed
- F forward and R reverse *gtpch*I primers
- Restriction sites for cloning purposes are indicated in green and blue for the forward and reverse primers respectively

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

2.3.2 Amplication and cloning of n*gtpch*I and o*gtpch*I into pGEM®-T Easy vector

The gene specific annealing cycles (5x cycles) were used to first amplify the *gtpch*I gene followed by the restrictions sites and *gtpch*I. This proved to be effective since the expected band of ~1174 bp was observed during gel electrophoresis of the PCR products (Figure 2-3). PCR products were then excised from the 1.5 % (w/v) gel and purified.

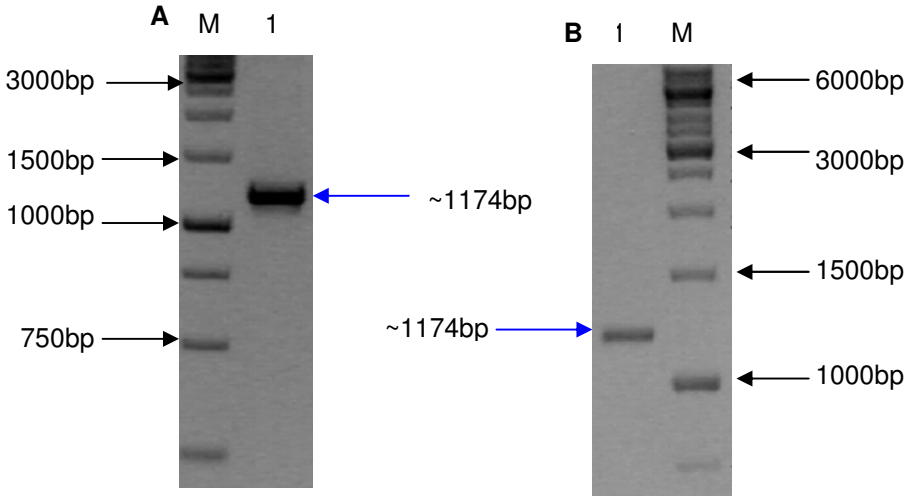


Figure 2-3: PCR amplification of the full length A) n*gtpch* and B) o*gtpch*I gene. M represents the molecular mass marker (Genruler 1kb DNA ladder, (Fermentas, Luthania). A) lane 1 represents the n*gtpch*I gene (KAPA 2G fast polymerase) and B) lane 1 represents the o*gtpch*I gene (*Ex Taq* polymerase). PCR products were loaded onto a 1.5 % (w/v) agarose gel. The blue arrow indicates the expected band (~1174 bp) for both the *gtpch*I genes.

2.3.3 Selection of positive recombinant clones of n*gtpch*I and o*gtpch*I- pGEM®-T Easy

The amplified PCR products obtained for both n*gtpch*I and o*gtpch*I (1174 bp) were ligated into pGEM®-T Easy vector. Colony PCR of the positive recombinant clones were performed (section 2.2.8.1) and the band corresponding to the *gtpch*I insert is noted in Figure 2-4 A (n*gtpch*I). This recombinant clone was then screened via restriction enzyme digestion (Figure 2-4 B). Double digestions were carried out using the *Nde*I and *BamH*I restriction enzymes (since the insert was designed to be flanked by these sites). Refer to Appendix B.1 for the plasmid vector map and restriction enzyme sites.

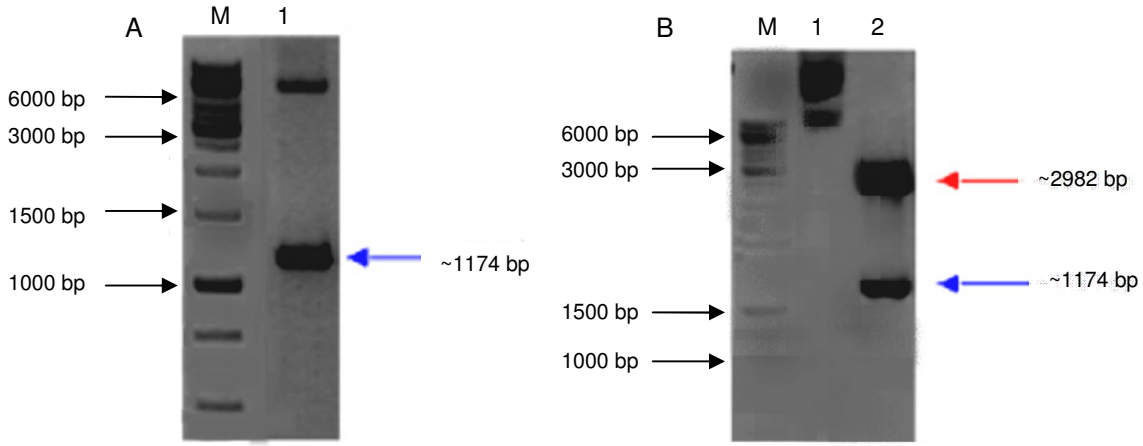


Figure 2-4: Screening of *Pf* n*gtpch*I- pGEM®-T easy vector. M represents the molar mass marker (Genruler 1kb DNA ladder, Fermentas, Luthania). A) Lanes 1 represents the amplification of the n*gtpch*I gene from a recombinant clone. B) The results obtained when n*gtpch*I-pGEM ® T easy vector was cleaved with the mentioned restriction enzymes. Lanes 1 indicates the presence of undigested plasmid DNA and lane 2 represents the digested plasmid DNA. The released n*gtpch*I insert (~1174 bp) is indicated by the blue arrow and the remaining cleaved pGEM® T easy vector by the red arrow.

Colony PCR of the positive recombinant clones for o*gtpch*I-pGEM®-T Easy vector were performed and the band corresponding to the *gtpch*I insert is noted in Figure 2-5 A (o*gtpch*I). This recombinant clone was then screened via restriction enzyme digestion (Figure 2-5 B) Both the n*gtpch*I and o*gtpch*I inserts were excised from the 1.5 % (w/v) gel and purified. Refer to Appendix B.2 for the plasmid vector map and restriction enzyme sites.

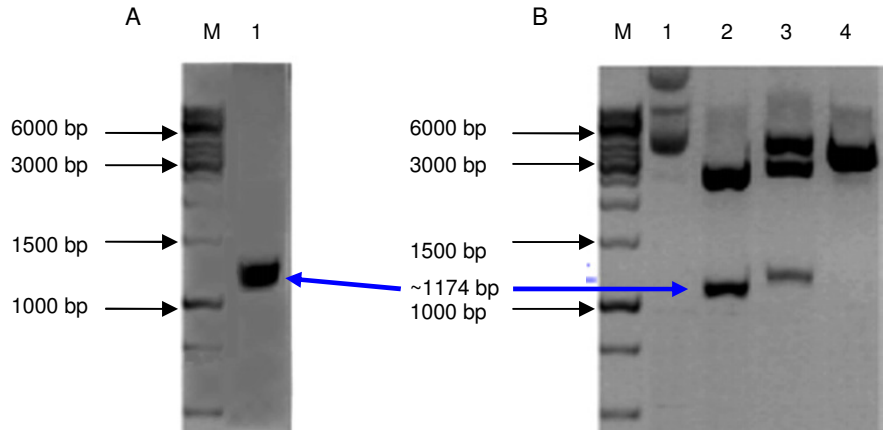


Figure 2-5: Screening of *Pf* o*gtpch*I- pGEM®-T Easy vector. M represents the molar mass marker (Genruler 1kb DNA ladder, Fermentas, Luthania). A) Lanes 1 represents the amplification of the o*gtpch*I gene. B) The results obtained when o*gtpch*I-pGEM®-T Easy vector was cleaved with a combination of *BamH*I-*Nde*I (lane 2), *EcoR*I (lane 3) and *EcoR*V (lane 4). Lanes 1 indicates the presence of undigested plasmid DNA (~4156 bp), lanes 2-4 represents the digested plasmid DNA where lane 2 represents ~1174 bp insert gene and ~ 2982 bp remaining pGEM®-T Easy vector, lane 3 represents ~1199 bp and ~ 2997 bp digested o*gtpch*I-pGEM®-T Easy vector and lane 4 represent ~ 4156 bp pGEM®-T Easy vector (since only one cleavage site is present). The blue arrow indicates the released o*gtpch*I insert (~1174 bp).

2.3.4 Isolation and restriction enzyme digestion of pET22b

Pure n*gtpch*I and o*gtpch*I inserts were used in sub-cloning of these inserts into the pET22b vector and transformed into DH5α *E. coli* cells. Firstly the pET22b vector was isolated and thereafter digested with *Nde*I and *BamH*I restriction enzymes (Figure 2-6). Refer to Appendix C.1 for the plasmid vector map and restriction enzyme sites.

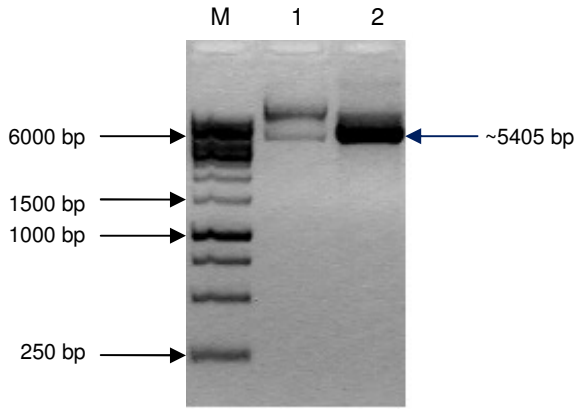


Figure 2-6: The *BamH*I and *Nde*I restriction vector map of pET22b. Restriction enzyme digestion of pET22b. M represents the molar mass marker (Genruler 1kb DNA ladder, Fermentas, Luthania). Lane 1 represents undigested pET22b and lane 2 digested pET22b. All digestions were with the restriction enzymes *Nde*I and *BamH*I (Promega, USA).

The results obtained for the restriction enzyme digestions are indicated in Figure 2-4 B) (for *ngtpch*I), Figure 2-5 B) (for *ogtpch*I) and Fig 2-6. The bands corresponding to ~1174bp (n*gtpch*I and o*gtpch*I insert) were released upon digestion of n*gtpch*I- and o*gtpch*I-pGEM®-T Easy vector including the digestion of pET22b releasing a ~5405bp digested vector. The n*gtpch*I, o*gtpch*I inserts and the pET22b vector were excised from the 1.5 % (w/v) gel and purified. The concentrations after purification from the agarose gel were 12.5 ng/ µl, n*gtpch*I 16.6 ng/ µl and pET22b 65 ng/ µl.

2.3.5 Selection of positive recombinant clones of n*gtpch*I and o*gtpch*I- pET22b

Colony screening PCR was utilized to screen for positive clones. A recombinant clone was selected for n*gtpch*I (Figure 2-7A & B) and o*gtpch*I (Figure 2-8A & B) each within the pET22b vector and screened via restriction enzyme digestion to determine the presence of the inserts. Refer to Appendix C.2 for the plasmid vector map and restriction enzyme sites.

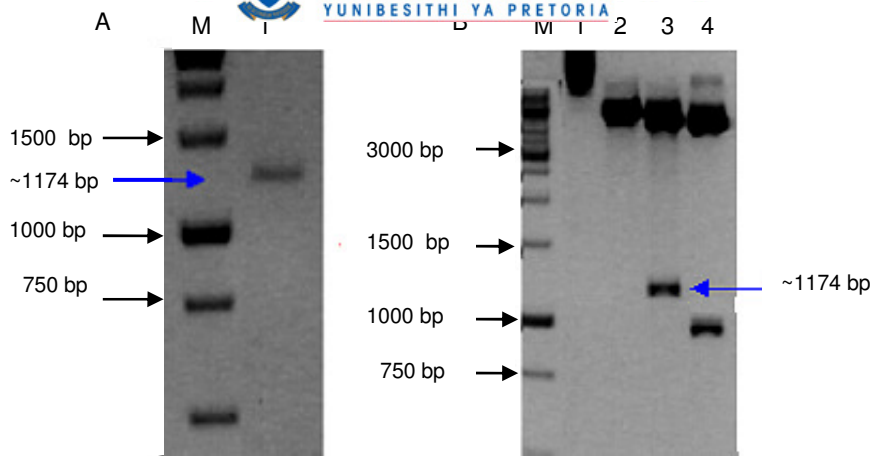


Figure 2-7: Screening of recombinant clone of n*gtpch*I- pET22b. M represents the molar mass marker (Genruler 1kb DNA ladder, Fermentas, Luthania). A) Colony PCR, lane 1: represents the amplification of the n*gtpch*I. PCR products were loaded onto a 1.5 % (w/v) agarose gel. The blue arrow indicates the expected band for *gtpch*I amplification. B) The expected bands (bp) when digested and loaded on a 1.0 % (w/v) agarose gel after gel electophoresis. Lane 1: undigested plasmid DNA, lane 2 represents ~6579 bp of pET22b vector digested with *Bsa* I, lane 3 represents ~1174 bp insert gene and ~ 5405 bp remaining pET22b vector digested with *BamH*I- *Nde*I and lane 4 represent ~ 975 bp, ~ 5325 bp and ~ 297 bp (not indicated on gel) n*gtpch*I-pET22b digested with *Xba*I- *Xho*I.

Colony PCR and screening of the positive clone for o*gtpch*I-pET22b is indicated in Figure 2-8 A & B, respectively. Refer to Appendix C.3 for the plasmid vector map and restriction enzyme sites. The bands corresponding to ~1174bp (n*gtpch*I and o*gtpch*I insert) were indicated upon digestion of n*gtpch*I- and o*gtpch*I-pET22b vector including the digestion of pET22b (5405 bp).

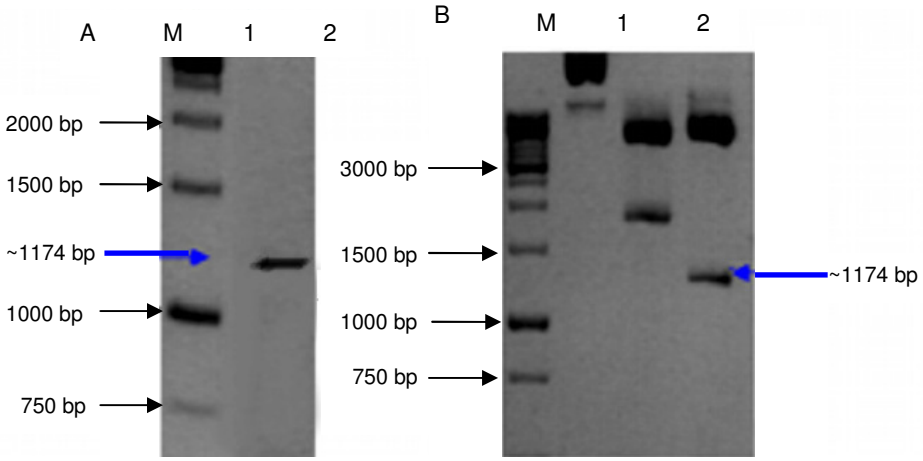


Figure 2-8: Screening of recombinant clone of o*gtpch*I- pET22b. M represents the molar mass marker (Genruler 1kb DNA ladder, Fermentas, Luthania). A) Colony PCR, lane 1: represents the amplification of the o*gtpch*I. PCR products were loaded onto a 1.5 % (w/v) agarose gel. The blue arrow indicates the expected band for *gtpch*I amplification. M represents the molar mass marker (Genruler 1kb DNA ladder, Fermentas). D) The results obtained once plasmid DNA (undigested lane 1) was digested with *EcoR* V (lane 2), *BamH*I- *Nde*I (lane 3) where lane 2 represents ~4710 bp and ~1869 bp digested o*gtpch*I-pET22b vector and lane 3 represents ~1174 bp insert gene and ~ 5405 bp remaining pET22b vector.

2.3.6 Comparison of codon frequency of the n*gtpchI* and h*gtpchI*

The codon usage frequency of an average of 10 codons for the n*gtpchI* gene of *P. falciparum* was generated from data obtained from a codon frequency generation tool conceptualized by de Ridder (University of Pretoria, South Africa) and analyzed using Microsoft Excel. Upon codon harmonization of *P. falciparum* n*gtpchI*, the codon usage frequency indicated in Figure 2-9 of the *P. falciparum* h*gtpchI* gene in *E. coli* is similar to that of wild type *P. falciparum gtpchI* in *P. falciparum* at a specific codon position. However, the codon usage frequency of the wild type *P. falciparum gtpchI* in *E. coli* varies in comparison to the h*gtpchI* and the *P. falciparum* n*gtpchI*. The harmonization tool takes into account the predicted pause sites, which are visible at distinct regions throughout the gene. It is observed at the distinct regions (e.g. codon 71) where low-medium frequency usage codons exist, predicted pause sites are found (Angov *et al.* 2003).

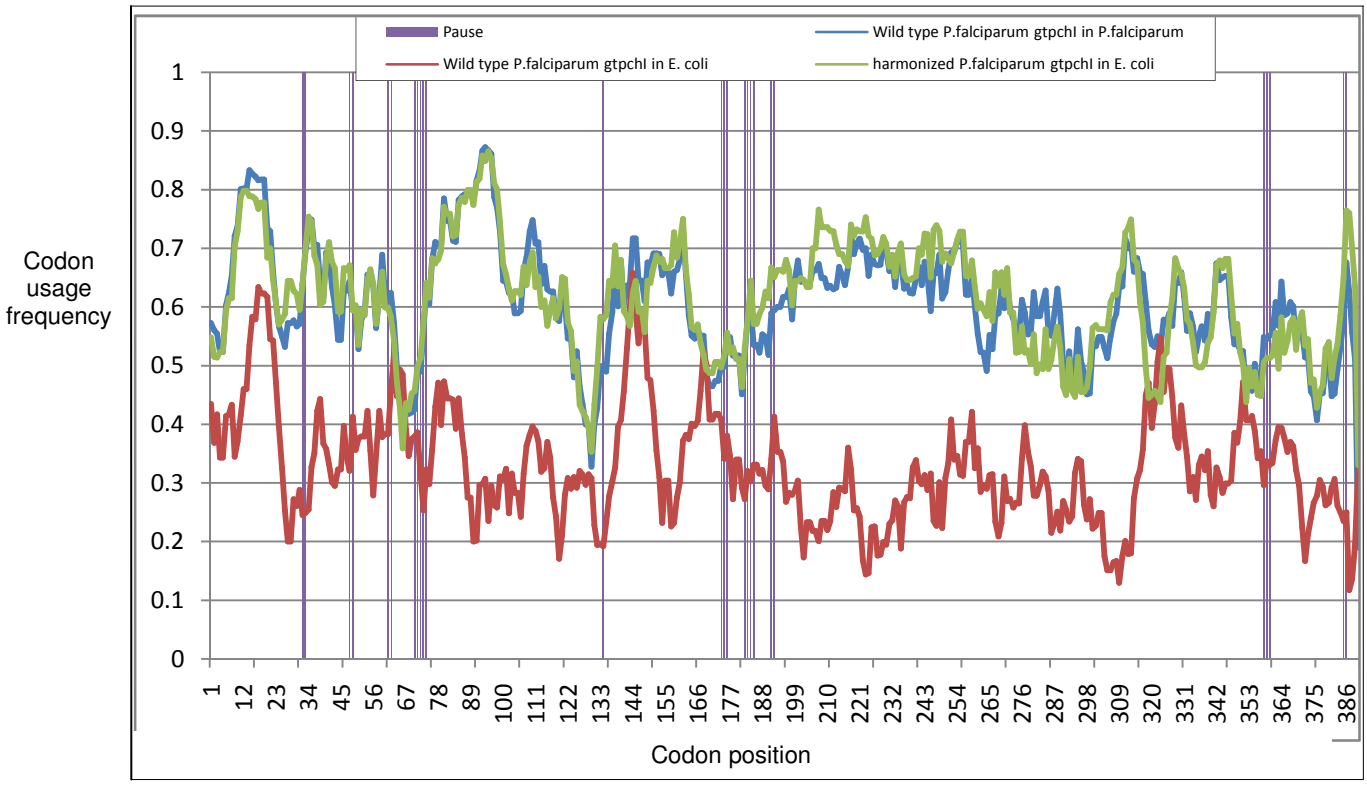


Figure 2-9: The average codon frequency of the wild type *gtpchI* gene in *P. falciparum* and *E. coli*, including the h*gtpchI* sequence in *E. coli* (10 codons). The perpendicular lines indicate the predicted pause sites. As indicated by the figure legend, the codon usage frequency of the wild type gene in the preferred *P. falciparum* host is blue; the wild type gene in the target *E. coli* host is red and the harmonized gene in *E. coli* is green.

2.3.7 RNA structure predictions of h*gtpch*I

Figure 2-10 represents the secondary structure prediction of the h*gtpch*I mRNA using the RNAstructure version 4.5 (Mathews, Zuker & Turner. 1996-2007). A mountain plot was used as a representation of the secondary structure. The y-axis is the height of the mountain at position *k* as a function of base position or codon. In Figure 2-10 the peaks represent hairpins (red arrow); plateaus signify unpaired bases (green arrow) and slopes correspond to paired bases (orange line). Seventeen RNA structures were predicted for the RNA of h*gtpchI*. This prediction was based on the ΔG for each elucidated structure. The most favorable RNA structure for h*gtpch*I had a ΔG calculated to be –210.7 kJ, other structures that were less favourable have an increased ΔG value (more + ΔG values) hence were depicted as less stable (the lower the ΔG value the greater the stability of mRNA structure). The data generated for each prediction was based on base pair binding.

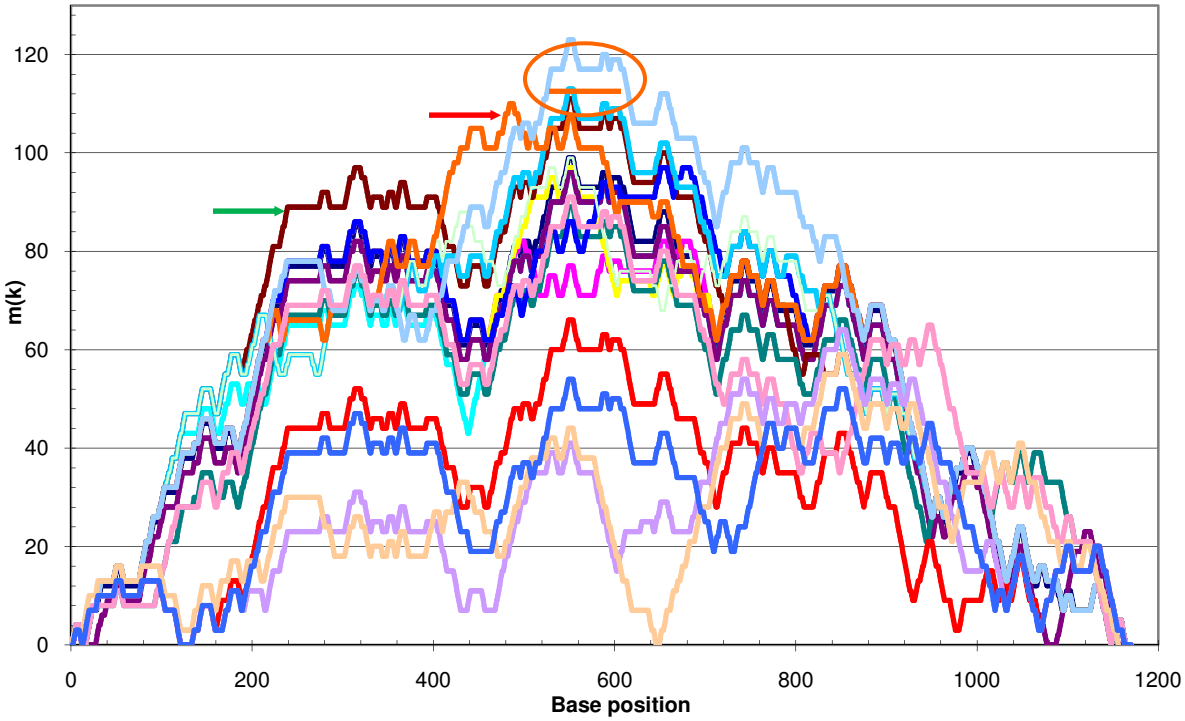


Figure 2-10: A mountain plot representation of 17 predicted secondary structures of the RNA of the h*gtpch*I gene in *P. falciparum* and the base pairing. M(k) signifies the height of the mountain a position k. The red arrow indicates a hairpin structure, the green arrow an unpaired base pair and the orange line marked with a circle is a paired base pair.

The gene was overviewed for any problematic regions that may hinder correct folding of the target protein. A final overview of the h*gtpch*I gene was performed. Analysis of the

nucleotide sequence revealed at codon position 22-25, a set of repeat of AAA codons existed, which was altered. This alteration was made to reduce intramolecular base pairing ensuring that strand slippage (due to the repeat of the AAA codon) does not occur and so the translation machinery of the heterologous host would pause to allow for the correct folding during protein expression. Refer to Appendix A for the sequence alignment of h*gtpch*I, o*gtpch*I and n*gtpch*I, including the amino acid sequence.

2.3.8 Isolation of h*gtpch*I-pUC57 and pET22b

For expression studies, the h*gtpch*I insert was sub-cloned from pUC57 into the pET22b vector. Figure 2-11 represents the restriction enzyme digestions for the release of the h*gtpch*I insert from pUC57. Refer to Appendix C.4 for the plasmid vector map and restriction enzyme sites.

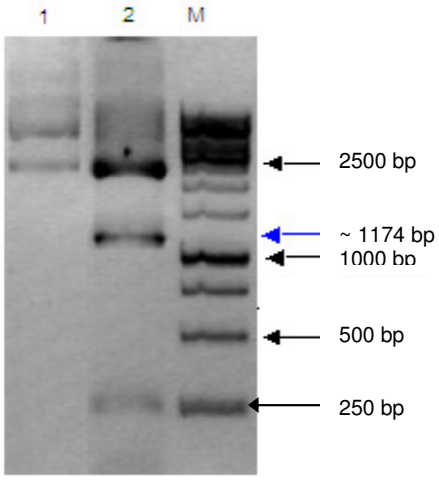


Figure 2-11: Restriction enzyme digestion of pUC57-h*gtpch*I: M represents the molar mass marker (Genruler 1kb DNA ladder, Fermentas, Luthania). Lane 1: undigested h*gtpch*I-pUC57, lane 2: restriction enzyme digestion of plasmid with *BamH*I and *Nde*I, which releases the ~ 1174 bp insert gene, a 2457 bp, 254 bp and 15 bp bands since this vector contains double *BamH*I and *Nde*I restriction sites present in the pUC57 vector.

The results obtained for the restriction enzyme digestion are indicated in Figure 2-11. The band corresponding to ~1174 bp (h*gtpch*I insert) was released upon digestion. The h*gtpch*I insert and the pET22b vector (refer to Figure 2-6) were purified. The concentrations acquired after purification from the agarose gel were, 32 ng/ µl for h*gtpch*I and pET22b 65 ng/ µl. The h*gtpch*I insert was ligated into the pET22b vector and positive recombinant clones were selected for screening.

2.3.9 Selection of positive recombinant clone of h*gtpch*I- pET22b

Colony PCR (Figure 2-12A) and restriction enzyme digestion (Figure 2-12B) were used for screening of the recombinant clone. Refer to Appendix C.5 for the plasmid vector map and restriction enzyme sites

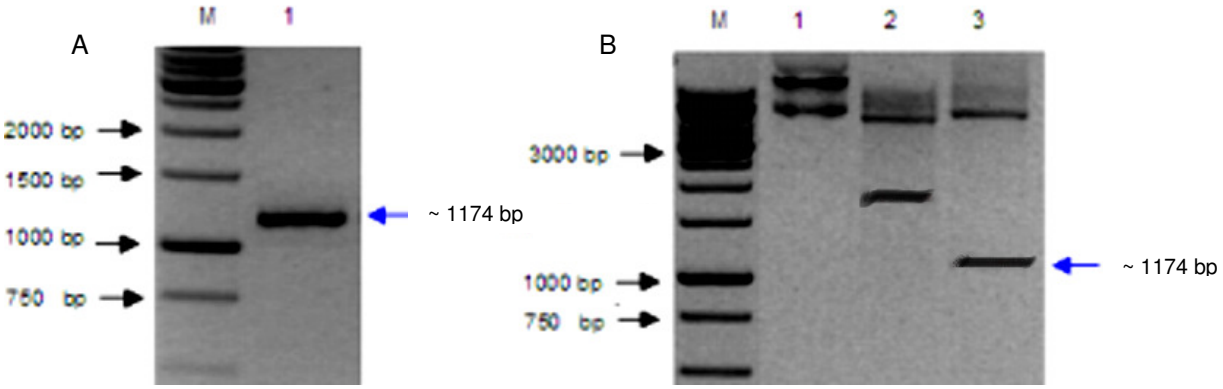


Figure 2-12: Screening of recombinant clone of h*gtpch*I- pET22b. M represents the molar mass marker (Genruler 1kb DNA ladder, Fermentas, Luthania). A) Colony PCR, lane 1: represents the amplification of the h*gtpch*I. PCR products were loaded onto a 1.5 % (w/v) agarose gel. The blue arrow indicates the expected band for *gtpch*I amplification. B) M represents the molar mass marker (Genruler 1kb DNA ladder, Fermentas). The results obtained once plasmid DNA (undigested lane 1) was digested with *EcoR* V (lane 2), *BamH*I- *Nde*I (lane 3) where lane 2 represent ~1869 bp and ~4710 bp bands of digested h*gtpch*I- pET22b and lane 3 represents the ~ 1174 bp insert gene and the ~ 5405 bp pET 22b vector.

2.3.10 Nucleotide sequences of *gtpch*I- pET22b

Nucleotide sequencing ensured that the full length of n*gtpch*I o*gtpch*I and h*gtpch*I gene sequences obtained after sub-cloning into the pET22b vector aligned with sequences from the theory (from plasmoDB for the native) without the incorporation of mutations that would alter the primary amino acid sequence and hence protein structure. Refer to Appendix D, E and F for the nucleotide sequencing results and the alignments to verify that the n*gtpch*I, o*gtpch*I and h*gtpch*I genes weresub-cloned into the pET22b expression vector, respectively.

2.3.11 Comparative protein expression

2.3.11.1 Comparative growth study

As a first line expression study n*gtpch*I-, o*gtpch*I- and h*gtpch*I-pET22b were transformed into BL21 star (DE3)-*Pf*Hsp70 (star 70), BL21 star (DE3) (star) and M15pREP4 (m15). The expressed proteins possessing the C-terminal His tag are referred to as nGTPCHI,

oGTPCHI and hGTPCHI. This study was performed to assess the efficiency of growth of the different cell lines. Figure 2-13 depicts the growth curves as a function of induction time.

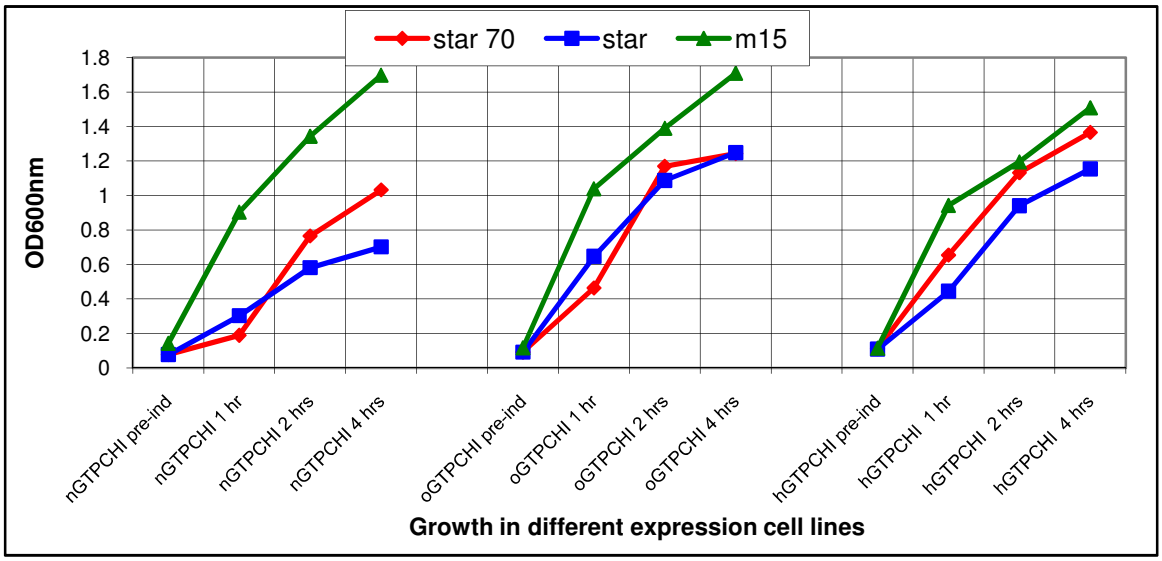


Figure 2-13: Growth curves used to monitor the nGTPCHI, oGTPCHI and hGTPCHI His tag protein expression in star 70, star and m15 cell lines as a function of induction time (pre-induction, 1 hr, 2hrs and 4 hrs post-induction).

When evaluating the variation in bacterial growth with an increase in induction time, an increase in growth of the mentioned cell lines was observed. Steeper growth curves (obtained during the expression of nGTPCHI, oGTPCHI and hGTPCHI), hence more robust growth was obtained for the m15 cell line in comparison to the other cell lines. This was followed by star 70 and thereafter star cell line. This could indicate the expression conditions (media composition, temperature etc.) favoured the expression of the m15 cell line. However, whether high growth curves could be associated with greater expression levels still requires assessment.

2.3.11.2 Protein expression levels as a function of induction time

It was assumed that the levels of protein expression increased as the time of post induction continued. To determine at what point the levels of protein expression peaked in the expression host, protein expression was monitored as a function of induction time. As such, bacterial cells containing the expressed proteins were harvested before induction, as well as 1 hr, 2 hrs and 4 hrs post-induction. Total protein extracts were loaded and analyzed with a 10% SDS-PAGE gel indicated in Figure 2-14.

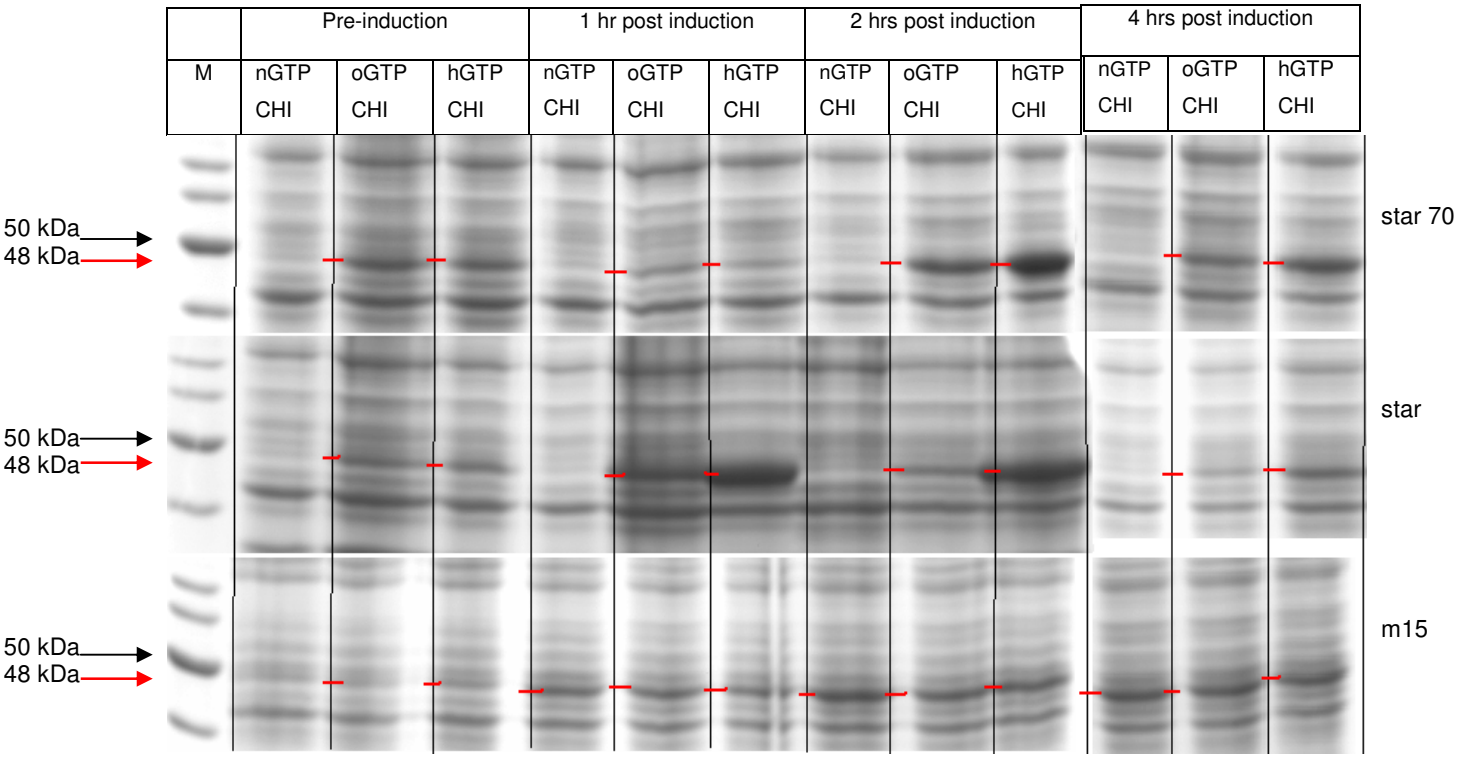


Figure 2-14: The level of expression of the nGTPCHI, oGTPCHI and hGTPCHI proteins in different expression cell lines (star 70, star and m15 *E. coli* cells) as a function of the induction time (Pre-induction – 4 hrs post-induction). The 10% SDS-PAGE gel was stained with colloidal coomassie. M: Protein Molecular weight marker (Fermentas, Luthania). The red arrow depicts the position of the expected protein band of 48kDa.

The band corresponding to the 48 kDa band on the SDS-PAGE gel was assumed to be the GTPCHI protein and the intensity of the 48 kDa band reflects the intensity of protein expression. Based on this it was evident that pre-induced extracts revealed low levels of nGTPCHI, oGTPCHI and hGTPCHI expression in each cell line assessed. Only upon induction (after 1 hr and 2 hrs) does the level of oGTPCHI and hGTPCHI expression increase significantly in the star 70 and star cell lines. Levels of nGTPCHI expression were only observed in m15 cell line and not in the other cell lines. Furthermore, based on the greatest level of nGTPCHI, oGTPCHI and hGTPCHI expression in all cell lines, 2 hrs post induction was sufficient for harvesting. The highest level of expression was obtained for hGTPCHI in star and star 70 cell line in comparison to the other constructs.

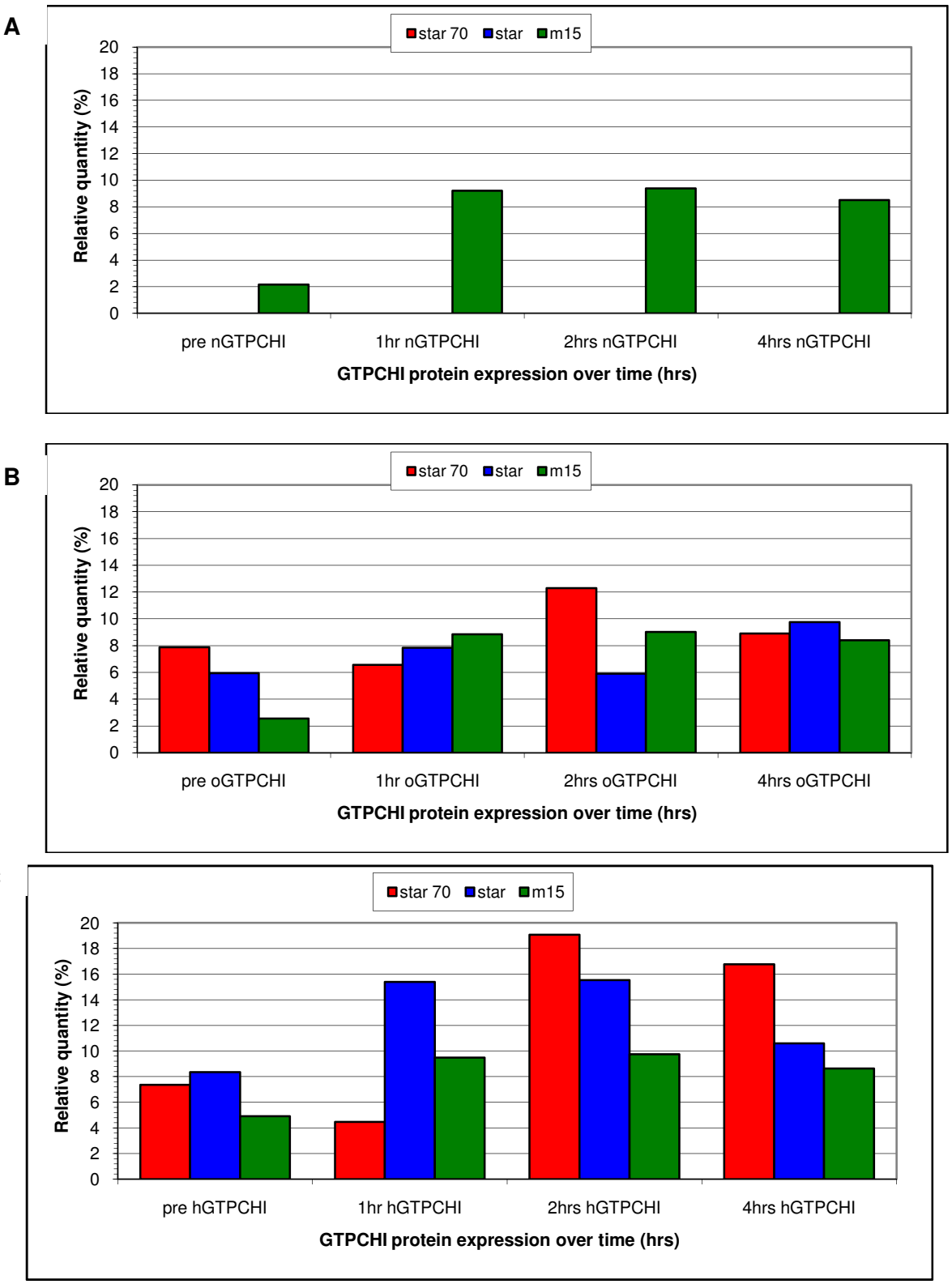


Figure 2-15: The relative quantity measured for total fractions of the A) nGTPCHI, B) oGTPCHI and C) hGTPCHI his tag proteins obtained as pre-induced, post induced (1 hr, 2 hrs and 4 hrs). Expression in star 70 cells is indicated by the red bars, star cells by the blue bars and in m15 by the green bars. Pre: pre-induction. 1hr, 2hrs and 4hrs: post-induction.

The results in Figure 2-15 of the relative quantity measured of total protein fractions corresponding to the 48 kDa band, confirmed the expression of nGTPCHI peaks after 2 hrs of induction in the m15 cell line. The expression of oGTPCHI peaked after 2 hrs of induction in the star 70 and m15 cell lines, whereas in the star cell line after 4 hrs induction. The expression levels were almost the same for oGTPCHI and nGTPCHI in all the constructs. Expression of hGTPCHI peaked after 2 hrs induction in each cell line. Furthermore, the relative quantity of hGTPCHI expression in star and star 70 cell lines was greater in comparison to oGTPCHI and nGTPCHI. The expression of nGTPCHI, oGTPCHI and hGTPCHI was similar in the m15 cell line. It should be noted that a protein band for nGTPCHI could not be detected in the star and star 70 expression cell lines using the Quantity OneTM Software Package.

2.3.11.3 Expression in BL21 (DE3) star cells

As a first line expression study, the total (tot) protein samples that include the insoluble (ins) and soluble (sol) protein were analyzed onto a 10 % SDS-PAGE gel, stained with Coomassie brilliant blue and quantified using the Versa Doc, Quantity One Program (these conditions were also utilized for m15 and star 70 cells).

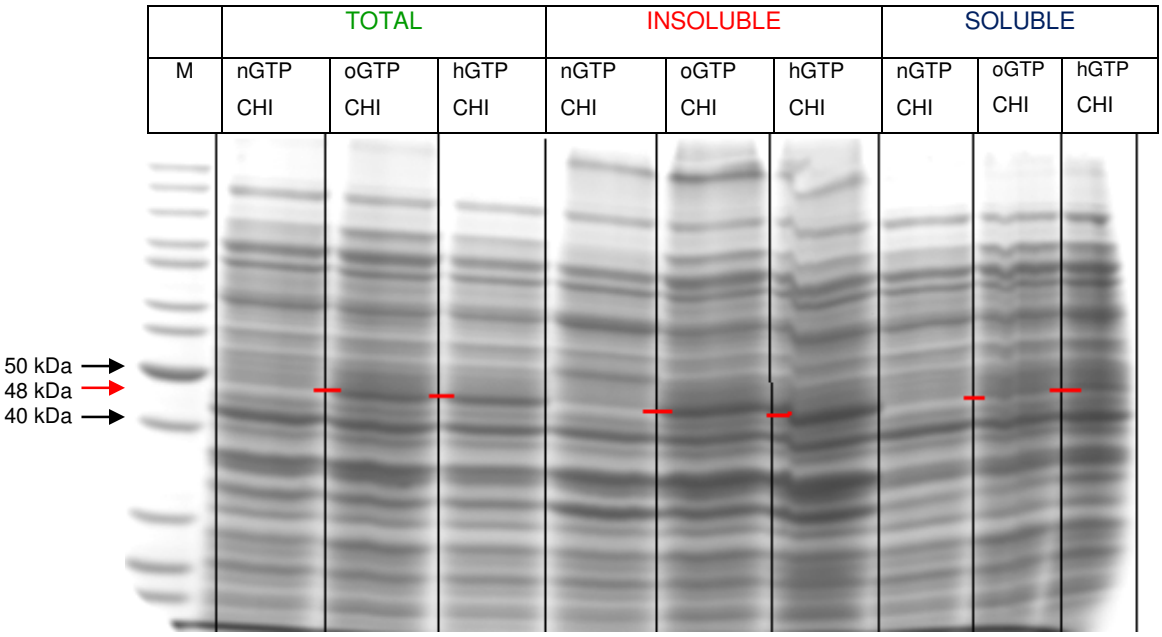
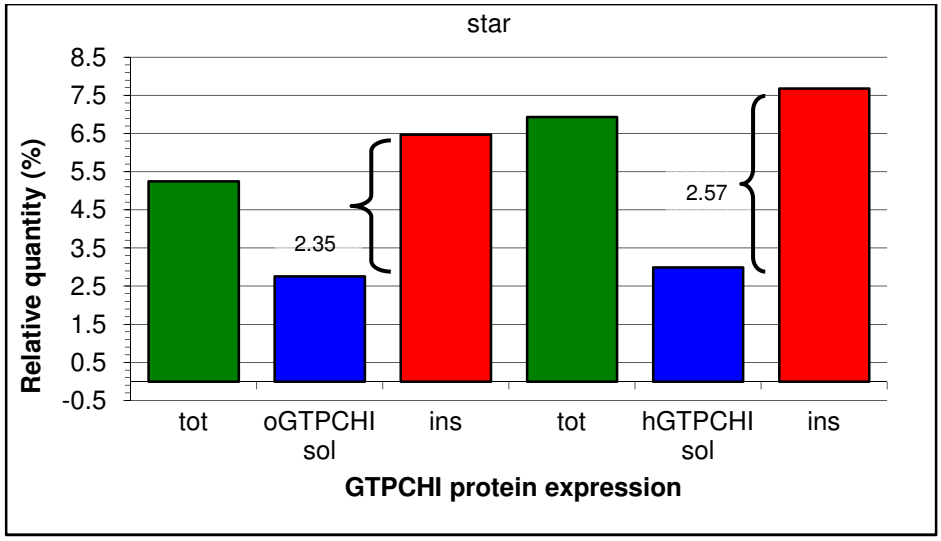


Figure 2-16: Expression of the nGTPCHI, oGTPCHI and hGTPCHI proteins in star cell line after 2 hrs of induction. M: Protein Molecular weight marker (Fermentas, Luthania), GTPCHI samples were loaded as total, insoluble and soluble on to 10 % (w/v) SDS-PAGE gel. The red arrow depicts the position of the expected protein band of 48kDa.

From Figure 2-16, it was observed that the nGTPCHI protein did not express significantly in the star cells in comparison to the oGTPCHI and hGTPCHI protein as no band was detected for nGTPCHI. It was also evident that the oGTPCHI and hGTPCHI expressed were in insoluble form when comparing it to the levels of both proteins present in the soluble fractions.



	oGTPCHI-SOL
oGTPCHI-TOT	1.90
oGTPCHI- INS	2.35
hGTPCHI-TOT	2.51
hGTPCHI-SOL	1.08
hGTPCHI- INS	2.78

Figure 2-17: The relative quantity measured for the total (tot, green bars), soluble (blue bars) and insoluble (red bars) fractions of the oGTPCHI and hGTPCHI proteins in star cells. The fold difference between the levels of soluble (sol) vs. Insoluble (ins) is indicated on the graph. Fold difference was calculated as the level of protein expression/ level of soluble oGTPCHI protein expression, where oGTPCHI expression is defined as 1.

The relative quantity of the proteins bands expected to be the GTPCHI protein were assessed to evaluate difference in expression levels (it should be noted that the western blots, to follow in section 2.3.11.6, will suggest whether the protein bands corresponding to 48 kDa on the SDS-PAGE gel were that of GTPCHI). Note the total protein will not equal the soluble and insoluble as they are both with respect to protein in that lane. Based on the analysis of relative quantity of the different protein fraction, it was evident that levels of insoluble protein was 2.35 (oGTPCHI) and 2.57 (hGTPCHI, not indicated in the table below Figure 2-17) fold, respectively greater in comparison to the levels of soluble proteins (left brace) for both oGTPCHI and hGTPCHI in the star cell line.

Furthermore, the levels of expression of the hGTPCHI was slightly greater (~2.5) than that of the oGTPCHI when comparing the tot and ins protein fractions and hGTPCHI was 1.1 fold greater for sol protein fractions. Based on this it can not be concluded that the harmonization of the *gtpch*I gene improved the expression of sol GTPCHI protein greatly. It should be noted that because the protein band corresponding to nGTPCHI could not be detected, a comparison could not be made.

2.3.11.4 M15 pREP4 cells

In an effort to enhance the expression of soluble nGTPCHI, oGTPCHI and hGTPCHI protein, the m15 cell line was utilized as an expression host. The features of the cell line include over-expression of the *lac* repressor thus upon IPTG induction expression levels would be expected to be greater in this cell line in comparison to the star cell line. The protein samples (tot, ins and sol) protein were isolated after 2 hrs of IPTG induction and analysed on a SDS-PAGE gel (Figure 2-18) as stated previously.

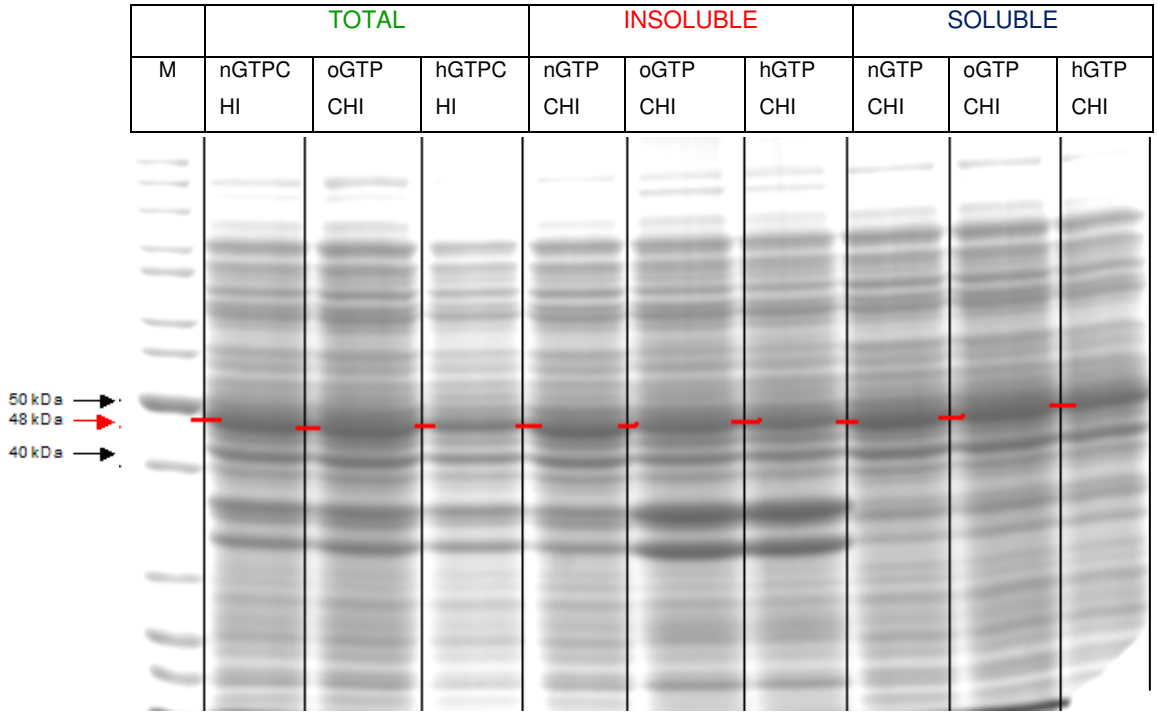
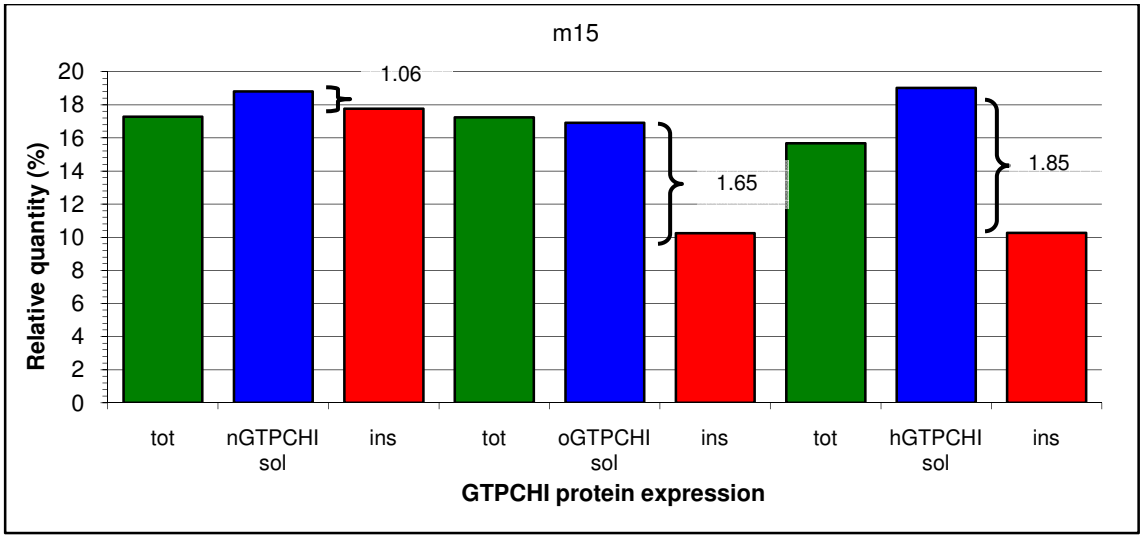


Figure 2-18: Expression of the nGTPCHI, oGTPCHI and hGTPCHI in m15 cells after 2 hrs of induction. M: Protein Molecular weight marker (Fermentas, Luthania), GTPCHI protein samples were loaded as total, insoluble and soluble on to 10 % (w/v) SDS-PAGE gel. The red arrow depicts the position of the expected protein band of 48kDa.

From Figure 2-18 it was observed that nGTPCHI, oGTPCHI and hGTPCHI expressed in the m15 cell line (due to the presence of the protein band at 48 kDa), in comparison to

expression in the star cells, where no expression was observed for the nGTPCHI protein. Based on the visual inspection, the levels of protein (nGTPCHI, oGTPCHI and hGTPCHI) present in all fractions (tot, ins and sol) were similar. Therefore, the relative quantity was determined to establish a difference in expression levels (Figure 2-19).



	oGTPCHI-INS
nGTPCHI-TOT	1.69
nGTPCHI SOL	1.83
nGTPCHI- INS	1.73
oGTPCHI-TOT	1.68
oGTPCHI SOL	1.65
hGTPCHI-TOT	1.53
hGTPCHI- SOL	1.86
hGTPCHI- INS	1.00

Figure 2-19: The relative quantity measured for total (tot, green bars), soluble (blue bars) and insoluble (red bars) of the nGTPCHI, oGTPCHI and hGTPCHI proteins in m15 cells. The fold difference between the levels of soluble (sol) vs. Insoluble (ins) is indicated on the graph. Fold difference was calculated as the level of protein expression/ level of insoluble oGTPCHI protein expression, where oGTPCHI expression is defined as 1.

Based on the analysis of relative quantity of the different protein fraction, it was evident that the level of soluble protein was 1.06, 1.65 and 1.85 fold, respectively greater in comparison to the levels of insoluble proteins for nGTPCHI, oGTPCHI and hGTPCHI proteins. Furthermore, the levels of total and soluble protein fractions were similar in the m15 cell line. However, levels of soluble nGTPCHI, oGTPCHI and hGTPCHI proteins increased when using m15 cells in contrast to the star cells. This indicated promising

evidence of greater protein expression using the m15 cells but western analysis is required to confirm this hypothesis.

2.3.11.5 Chaperone expression

The use of *Pf* molecular chaperones seemed promising since previous studies by de Marco and De Marco (2004) revealed an increase in protein yield and solubility for proteins co-transformed with molecular chaperones from the same species but cloned into a different plasmid. To look at the effect of *Pf*hsp70 co-expression the levels of soluble nGTPCHI, oGTPCHI and hGTPCHI expression, co-expression studies using *Pf*hsp70 were investigated. Once again protein samples were isolated and analysed on a SDS-PAGE gel (Figure 2-20).

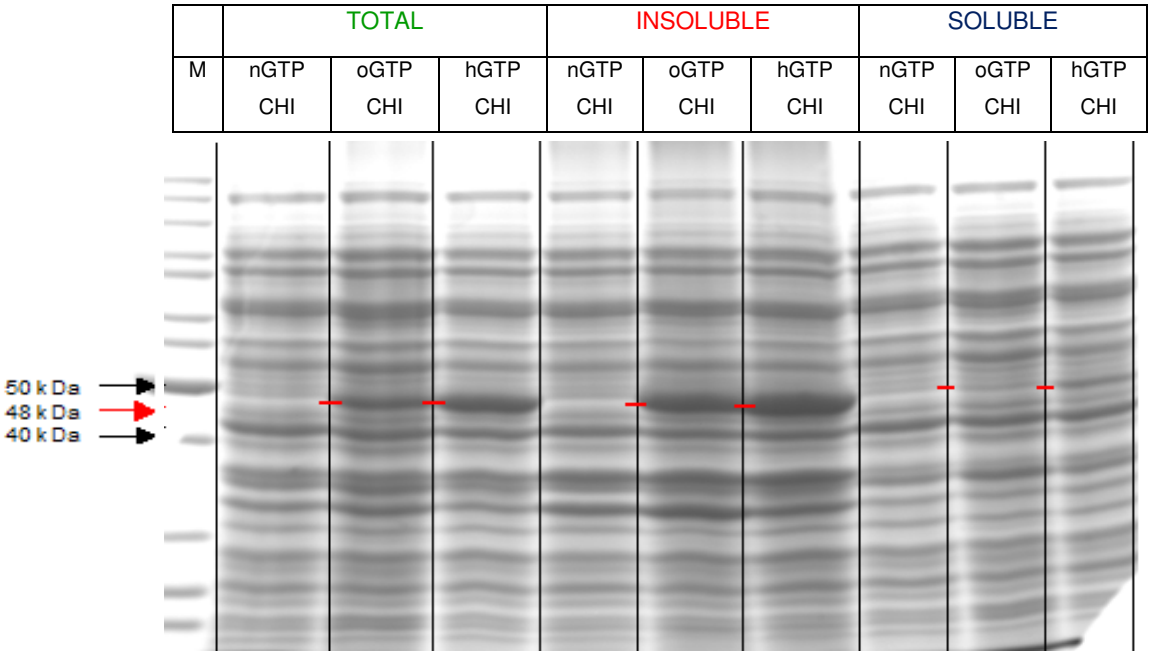
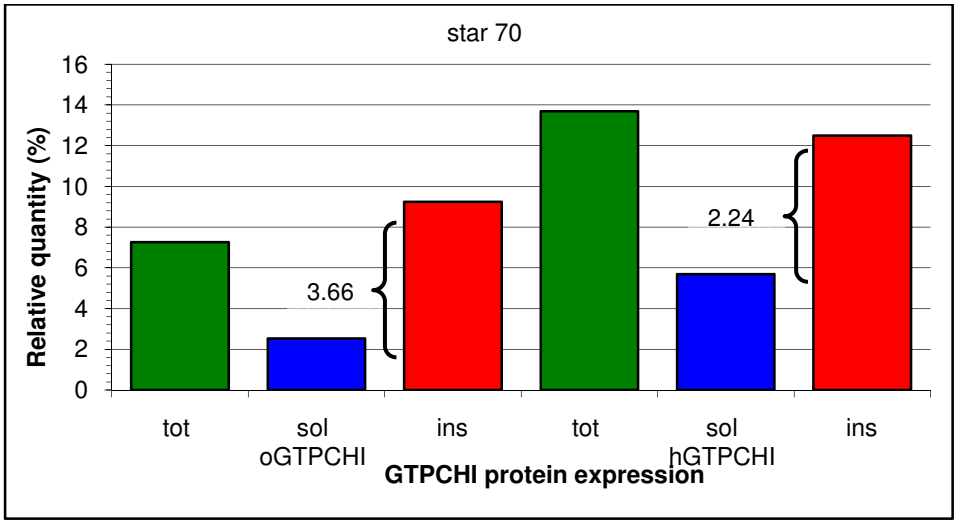


Figure 2-20: Expression of the nGTPCHI, oGTPCHI and hGTPCHI in star 70 cell line after 2 hrs induction. M: Protein Molecular weight marker (Fermentas, Luthania), Protein samples were loaded as total, insoluble and soluble. The red arrow depicts the position of the expected protein band of 48kDa.

It was evident that the majority of the proteins were expressed in insoluble form when comparing it to the levels of proteins present in the soluble fractions. Based on the protein bands, there seems to higher levels of hGTPCHI expression in comparison to oGTPCHI. However, this has to be investigated further by analysis of the difference in the relative quantity of proteins. Furthermore, a 70 kDa band corresponding to *Pf*hsp70

could possibly indicate the expression of the chaperone protein in conjunction with the target protein.



	oGTPCHI-SOL
oGTPCHI-TOT	2.87
oGTPCHI- INS	3.66
hGTPCHI-TOT	5.42
hGTPCHI-SOL	2.25
hGTPCHI- INS	4.95

Figure 2-21: The relative quantity measured for total (tot, green bars), soluble (blue bars) and insoluble (red bars) of the oGTPCHI and hGTPCHI in star 70 cells. The fold difference between the levels of soluble (sol) vs. Insoluble (ins) is indicated on the graph. Fold difference was calculated as the level of protein expression/ level of soluble oGTPCHI protein expression, where oGTPCHI expression is defined as 1.

The levels of both soluble oGTPCHI and hGTPCHI expression were lower in comparison to the expression of the soluble proteins obtained when using the m15 cell line as indicated in Figure 2-21. However, western analysis of the purified nGTPCHI, oGTPCHI and hGTPCHI would confirm the results obtained. Once again, nGTPCHI expression was not detected using star 70 cell line since the band corresponding to the 48 kDa protein was not visible. Additionally, the level of total hGTPCHI expression was greater in comparison to total oGTPCHI.

In summary, based on the results obtained for the expression of nGTPCHI, oGTPCHI and hGTPCHI it was observed that the greatest level of proteins expressed after 2 hrs of induction with IPTG. Furthermore, a comparative expression study using the various cell lines revealed that the expression of nGTPCHI was only detectable using the m15 cell

line, the levels of insoluble oGTPCHI and hGTPCHI protein were greater in comparison to both the soluble protein fractions in the star and star 70 cell lines. On the contrary, greater levels of soluble protein expression were observed using the m15 cell line. An assessment of the levels of each protein indicated that hGTPCHI expression was similar to oGTPCHI in the star and star 70 cell lines. However, all of the results presented need to be confirmed by western analysis since the target protein (GTPCHI) was assumed to have expressed based on the presence of a band on the SDS-PAGE gels that corresponded to the 48 kDa target protein.

2.3.11.6 *P. falciparum* GTPCHI His-tag protein purification

Soluble fractions pooled from the expression of the target protein using the different *E. coli* cell lines were purified using a Ni^{2+} affinity chromatography. Purified soluble protein fractions were eluted and 3 µg of protein samples were analyzed on a 10% SDS-PAGE gel and western blots as indicated in Figure 2-22.

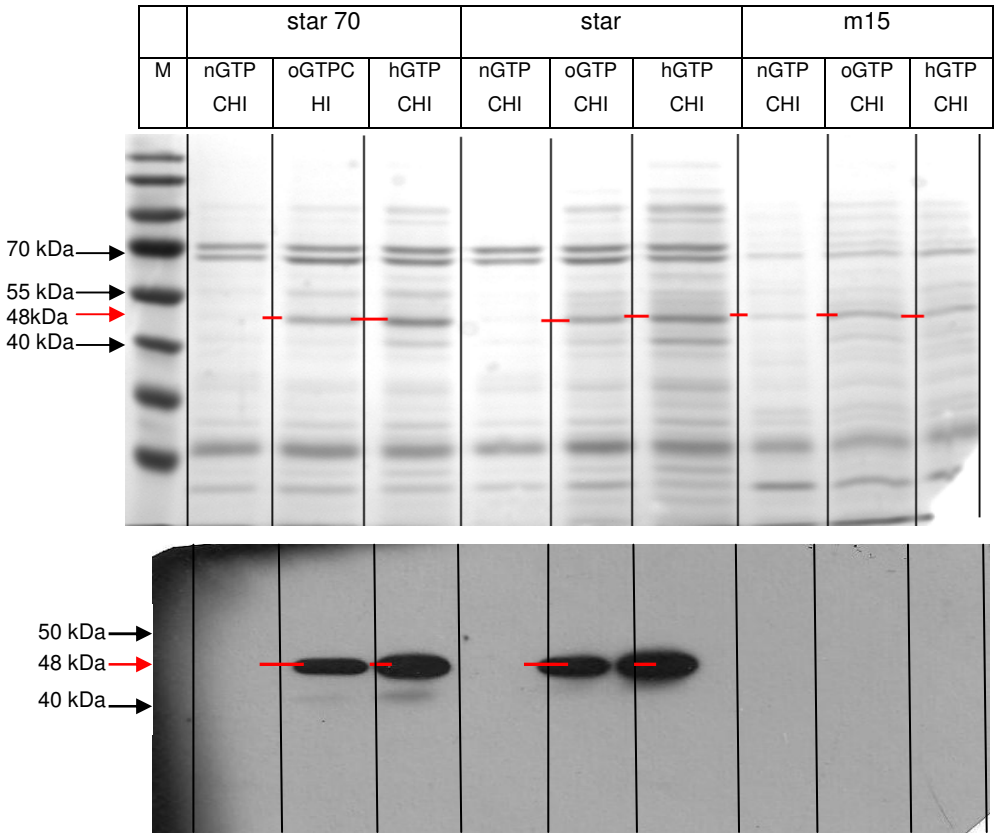
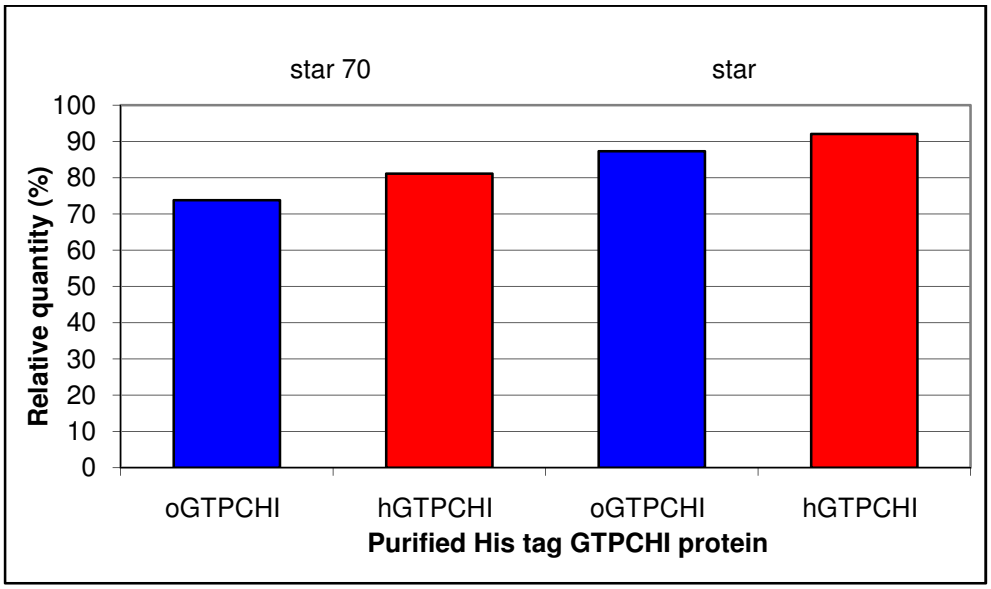


Figure 2-22: SDS-PAGE (10% (w/v)) and western blot analysis of Ni^{2+} ion affinity purification of the nGTPCHI, oGTPCHI and hGTPCHI his tag proteins from star 70, star and m15 cell lines, respectively. M: Protein Molecular weight marker (Fermentas, Luthania). The red arrow depicts the position of the expected protein band of 48kDa. His- tagged proteins were detected using His Probe-HRP (horseradish peroxidase) and detected via chemiluminescence.

His tag based purification did not result in the purification of homogenous GTPCHI proteins but the expression of GTPCHI was verified using western analysis. Furthermore, doublet of bands at 70 kDa were observed to express significantly and were purified with the target proteins but were not detected upon western analysis. Western blot analysis revealed that no nGTPCHI his tag proteins were purified using the star 70, star cell line or the m15 cell line. Levels of hGTPCHI were similar in comparison to the oGTPCHI protein when using both the star and star 70 cell lines. However, bands below both the hGTPCHI and oGTPCHI proteins were detected just above 40 kDa marker in star 70 cells based on the western blot.



	oGTPCHI-STAR 70
hGTPCHI- STAR 70	1.099304
oGTPCHI- STAR	1.183229
hGTPCHI- STAR	1.24803
	hGTPCHI- STAR 70
hGTPCHI- STAR	1.135291

Figure 2-23: The relative quantity measured for the western blot analysis of purified fractions of oGTPCHI and hGTPCHI his tag proteins. The fold difference between the levels of soluble vs. insoluble is indicated on the graph. Fold difference was calculated as the level of protein expression/ level of oGTPCHI protein expression in star 70 cells, where oGTPCHI expression is defined as 1.

The levels of expression using the star cell were slightly greater (~1.1 fold) in comparison to the star 70 cells for both oGTPCHI and hGTPCHI as indicated in Figure 2-23. Bands detected below the target protein (48 kDa) for both oGTPCHI and hGTPCHI using the star 70 cell line may have resulted in the difference noted.

2.3.12 Auto-induction media

In an effort to improve the expression of soluble proteins, growth conditions can be manipulated. As such, the media composition was altered. The use and success of auto-induction media was proposed by (Grossman *et al.* 1998; Studier 2005). Star 70 cells were used to indicate whether target protein expression could improve in conjunction with the chaperone protein when using different media compositions. Furthermore, the expression of the nGTPCHI protein using the m15 cells sparked an enquiry on whether expression of this protein in these cells could be detected when altering media composition. A major difference was that induction was not performed at specific time or growth points since the media already contained inducing components and cultures were grown for 48 hrs. The aim of this was not only to compare the expression levels upon change in media composition but also to determine whether an increase in expression levels could be attained.

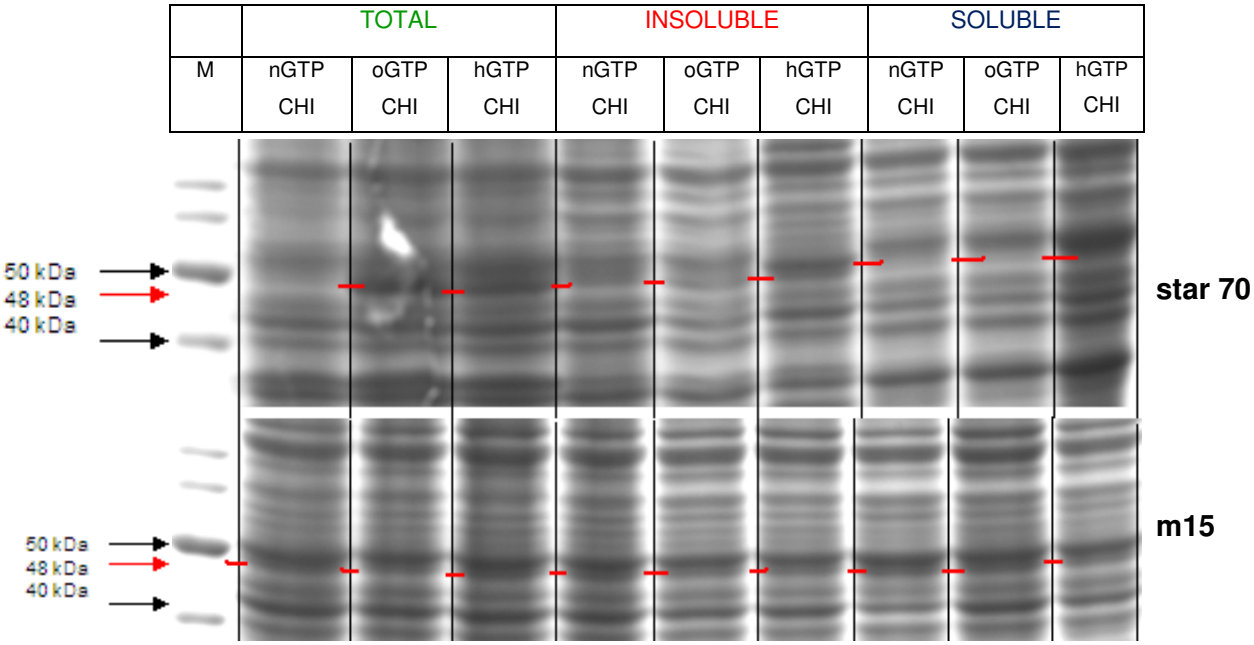


Figure 2-24: Expression of the nGTPCHI, oGTPCHI and hGTPCHI his tag proteins in A) star 70 and B) m15 cells for 48 hrs. M: Protein Molecular weight marker (Fermentas, Luthania), protein samples were loaded as total, insoluble and soluble on a 10 % (w/v) SDS-PAGE gel. The red arrow depicts the position of the expected protein band of 48kDa

Expression of nGTPCHI, oGTPCHI and hGTPCHI in auto-inducing media allowed for more prominent expression of background *E. coli* proteins as visible in Figure 2-24. The expression of nGTPCHI, oGTPCHI and hGTPCHI based on the corresponding 48 kDa

protein band was detected in the total, insoluble and soluble fraction in both cell lines. However, western analysis of purified protein fractions would confirm these findings.

Soluble fractions obtained from the expression of the nGTPCHI, oGTPCHI and hGTPCHI using auto-induction media in two different *E. coli* cell lines, were purified using a Ni^{2+} affinity column and analysed as mentioned previously.

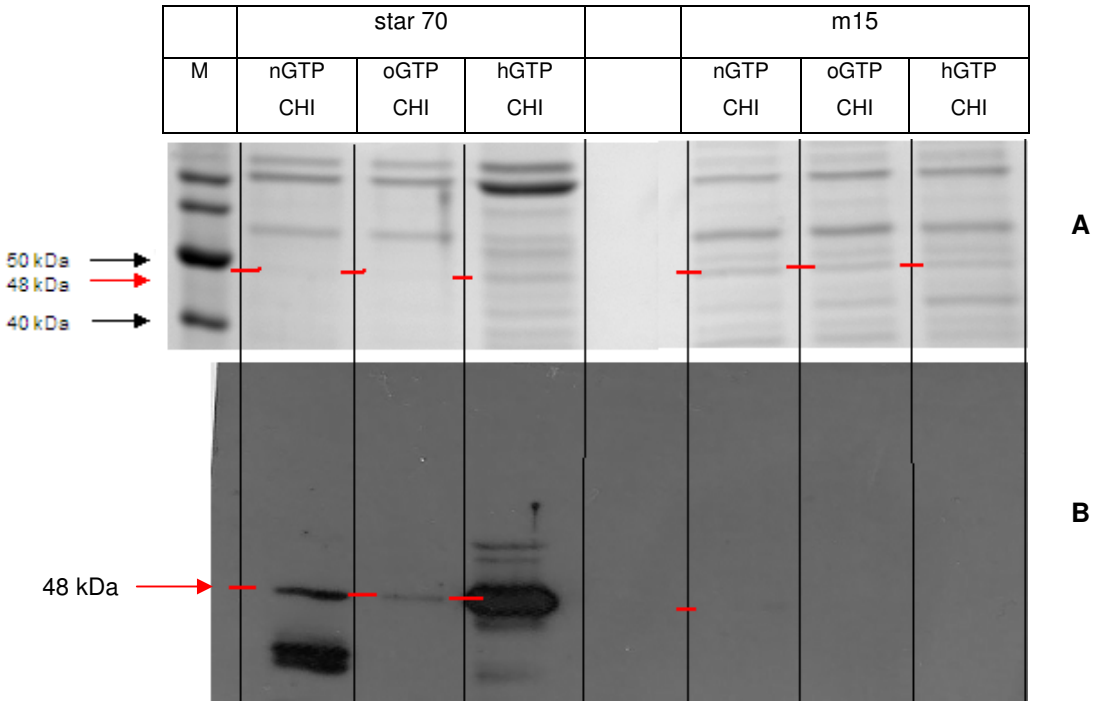
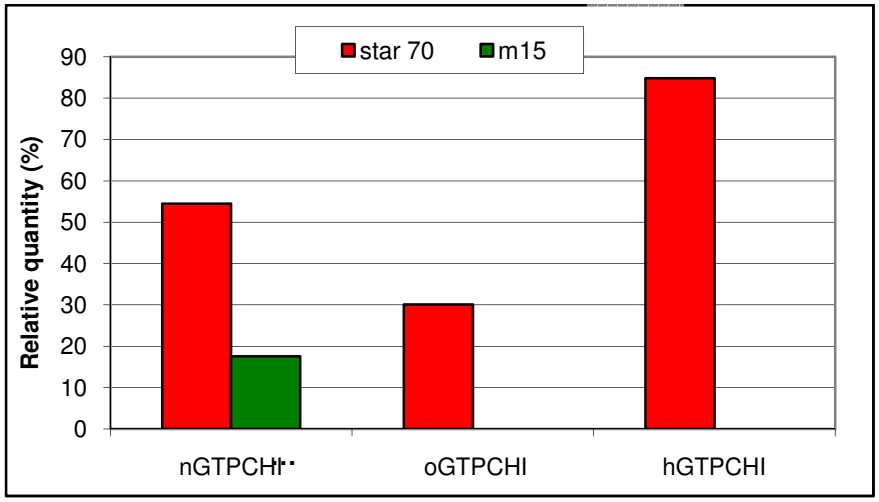


Figure 2-25: A) SDS-PAGE (10%v(w/v)) and B) western blot analysis of Ni^{2+} ion affinity purified nGTPCHI, oGTPCHI and hGTPCHI his tag proteins from star 70 and m15 cell lines, respectively. M: Protein Molecular weight marker (Fermentas, Luthania). The red arrow depicts the position of the expected protein band of 48kDa. His- tagged proteins were detected using His Probe-HRP (horseradish peroxidase) and detected via chemiluminescence.

Soluble protein fractions were not purified to homogeneity as seen in Figure 2-25 A for both the star 70 and m15 cell lines. Based on the western blot, nGTPCHI, oGTPCHI and hGTPCHI expressed using the star 70, a faint band at 48 kDa was detected for nGTPCHI in the m15 cell but no evidence of hGTPCHI and oGTPCHI expression were visible using the m15 cell line. Additional bands (between 30-20 kDa) were detected on the western blot, which do not correspond to the GTPCHI protein since the molecular masses were either smaller or greater than 48 kDa.



	oGTPCHI-m15
nGTPCHI- STAR 70	3.10422
oGTPCHI- STAR 70	1.712968
hGTPCHI- STAR 70	4.828749
	oGTPCHI- STAR 70
hGTPCHI- STAR 70	2.818937

Figure 2-26: The relative quantity measured for the western blot analysis of purified fractions of nGTPCHI, oGTPCHI and hGTPCHI his tag proteins. Expression levels in star 70 cells are indicated by the red bars and in m15 cells by the green bars. The fold difference between the levels of soluble vs. insoluble is indicated on the graph. Fold difference was calculated as the level of protein expression/ level of nGTPCHI protein expression in m15 cells, where nGTPCHI expression is defined as 1.

The analysis of the relative quantity depicted in Figure 2-26 revealed, the levels of hGTPCHI protein was greater than nGTPCHI and oGTPCHI in star 70 cells. However, using the m15 the levels of nGTPCHI protein expression was detected but both the oGTPCHI and hGTPCHI proteins could not be found.

Based on the results it was observed the change in media composition did not result in increased soluble protein expression of nGTPCHI, oGTPCHI and hGTPCHI. The levels of both purified hGTPCHI and oGTPCHI were detected on the western blot (Figure 2-25) using the auto-induction media. However, based on the relative quantity the levels of hGTPCHI and oGTPCHI using YT media (Figure 2-22), was greater. Furthermore, additional bands were detected on the western blot (Figure 2-25).

2.4 Discussion

2.4.1 Codon harmonization and RNA structure prediction

The overexpression of malaria proteins in large quantities is required for biochemical characterization. However, using *E. coli,* as a heterologous expression host remains a challenge due to various contributing factors mentioned (section 2.1.1). One such factor is codon bias, which is the preferential use of a specific codon coding for an amino acid within the organism. In general, every amino acid (except Met and Trp) can be encoded by 2 or more different codons. Furthermore, genes selectively use codons identified by the most abundant isoaccepting tRNA but avoid using codons of other tRNAs (Ikemura and Aota 1985). It is this mismatch of codons between *Plasmodium* and in particular, *E. coli* among other reasons that affects heterologous expression of malaria proteins (Birkholtz *et al.* 2008a).

On a gene level, this hindrance can be possibly overcome by codon harmonization. Codon harmonization is the synonymous replacement of codons from *E. coli* that closely match the codon usage frequency in the wild type gene by the native expression host (*P. falciparum*). The concept of using codon harmonization to improve the expression of these “difficult proteins” was developed since protein synthesis and folding in *E. coli* is co-translational and nucleotide sequence-dependent modulation of translation kinetics could possibly control nascent polypeptide folding. By changing codon usage frequencies it influences the secondary mRNA and possibly the tertiary protein structure as well as function of recombinant proteins (Angov *et al.* 2008). A codon usage frequency is assigned based on the most frequently used codons by *P. falciparum*. These codons include ATG, ATA, GTT, GAT, CCA, CGT, TTA, which translate to Met, Ile, Val. Glu, Pro, Asn and Leu, respectively (Ringe and Petsko 2003). Codon harmonization entails choosing codons to suit the codon usage frequency favoured by the host (*P. falciparum*). As such, the low usage frequency codons of the wild type gene in *P. falciparum* remain low usage frequency codons in the codon harmonized sequence and equally for high usage codons. This process guarantees that the codon frequency of low/intermediate and high usage codons at specific regions to remain similar in the heterologous host. Therefore, this allows the translational processes in the heterologous host to correspond to those of the natural host (Angov *et al.* 2008). It should be noted

that the amino acids encoded by the altered codons remain the same as in the *P. falciparum*.

Upon codon harmonization of native *gtpch*I gene sequence, the codon usage frequency of h*gtpch*I was similar to the codon usage frequency of *P. falciparum* as observed in Figure 2-9, which was required in order for the gene to be translated similarly by the *E. coli* expression host. The harmonization of the gene takes into account the predicted pause sites, which are believed to be visible at distinct regions throughout the gene. Translational pausing occurs as a result of ribosomes coming into contact with low frequency codons. These pause sites are advantageous since halting at these sites momentarily allows correct folding of proteins. In other words the translational machinery of the non-natural host is similar to the natural host (Angov and Lyon 2006).

It is evident that the function of RNA is encoded within its three-dimensional structure just as is the case with proteins (Frellsen *et al.* 2009). There are several aspects of mRNA structure, which alters the rate of translation by the ribosome and in turn the folding process in the nascent polypeptide. These include the presence of “slow” codons (used infrequently) and the presence of higher order nucleotide structures (e.g. hairpin loops) at the initiation step. In these cases ribosomes pause and possibly stack at these initiation sites (Thanaraj and Argos 1996; Takyar *et al.* 2005). Once ribosomes pass the initiation sites, they are then capable of disrupting base-pairing and translation continues. Structured elements within the mRNA coding region may cause slow elongation, but ribosomes eventually pass these regions (Paulus *et al.* 2004).

Therefore, after the final *gtpch*I sequence was analyzed RNA structures were predicted to ensure that translation would not be inhibited by mRNA structure near the translation initiation sites. Current procedures relies on energy thermodynamic calculations to determine RNA secondary structure (Frellsen *et al.* 2009). The thermodynamic method is based on estimations of the RNA sequence stability for different motifs present in the RNA. Using the RNAstructure program, computational data was generated of the predicted secondary structures. This program uses colour annotations of base pair possibilities to calculate base pairing with high confidence (Mathews 2006). Figure 2-10 is a representation of the secondary structure prediction of h*gtpch*I. A mountain plot was used as a representation of the secondary structure. Most of the structures seen in

Figure 2-10 have similar stability (ΔG), except structure 15. When comparing the RNA structure that has the lowest free energy structure to the other sub-optimal structures, the differences noted in base pairing would not affect the tertiary structure adversely. Furthermore, on average greater than 85% of known base pairs are correctly predicted in the lowest ΔG RNA structure (Mathews 2006).

The stability of the 5' end of an mRNA transcript to form stable secondary structures was also found to affect the expression levels of proteins (Griswold *et al.* 2003). Hairpin RNA structures positioned close to the 5' end, reduces the initiation of translation by obstructing ribosome loading (Kozak 1989). As such, a set of repeat of AAA at codon position 22-25 codons, was altered since it was found that by deleting or weakening the hairpin structure, (by altering codons to lower the ΔG) translation efficiency would improve (Kozak 1989).

2.4.2 Expression: Codon harmonization as opposed to codon optimization

The *gtpch*I gene in *P. falciparum*, including other malaria species has been characterized and the activity of this protein has been measured in other *Plasmodium* species (Lee *et al.* 2001). Previous attempts at expressing the nGTPCHI protein from *P. falciparum* have been unsuccessful (Prof Hyde, University of Manchester, UK). The focus of this chapter was therefore, on the expression of the GTPCHI protein using various *E. coli* cell lines, alternative media composition and molecular chaperones. A codon optimized form of the n*gtpch*I gene was constructed by Prof Hyde (University of Manchester, UK) and expression studies of the o*gtpch*I at present are still being conducted. As such, a codon harmonized *gtpch*I gene was constructed (algorithm (/http://www.sami.org.za/equalize///) and synthesized in an attempt to improve expression. These recombinantly cloned constructs were evaluated for their difference in expression levels using *E. coli* as the expression host.

Codon optimization entails substituting codons to match the codon preference in the expression host (*E. coli*). However, upon codon optimization the translational machinery differs to that of *P. falciparum (Angov et al. 2008)*. Studies aimed at the use of codon optimization to increase expression of soluble protein has been prevalent in the

expression of malaria vaccine candidates. An example includes the codon optimization of *Pf*CSP and *P. yoelli* HEP17, which resulted in enhanced expression and antibody response. However, analysis of the antigen-specific T-cell responses (protection is mediated by T- cell response of pre-erythrocytic antigens) indicated similar results to the native vaccines. These results revealed the effect of codon optimization was antigen-dependent and was ineffective for vaccines required for induction of T-cell dependent protective immunity (Dobano *et al.* 2009). The use of codon optimization has shown to be successful in the expression of some *Plasmodium* proteins. However, the synonymous replacement of codons from infrequently used to frequently used codons in regions of slow mRNA translation has been observed to be a hindrance to protein stability and this is not accounted for when using codon optimization (Angov *et al.* 2008). Furthermore, the use of rare codons in regions containing high frequency usage codons may lead to an imbalanced tRNA pool, mismatched codon usage patterns and the formation of erroneous proteins (Gustafsson *et al.* 2004; Chowdhury *et al.* 2009). This occurs because only one codon (frequently used by the host) codes for each amino acid resulting in a strongly transcribed mRNA and a gene with high codon concentrations for a group of tRNA. Since there would be no flexibility in codon usage, repetitive sequences and secondary structures would be unavoidable, inhibiting ribosome processivity (Gustafsson *et al.* 2004). The formation of erroneous proteins caused by codon miss-readings was observed in the coding sequence of the HSV-2 p27 protease domain. The increased level of Glu miss-integration and frameshifts linked with translational errors at CGG (Arg) codon occurs because of the high abundance of CGG usage with respect to the *E. coli* host preference (McNulty *et al.* 2003).

Previously successful attempts have been made at the expression of *Plasmodium* proteins in soluble form using the codon harmonization strategy. A candidate transmission blocking vaccines termed *Pf*48/45 was effectively expressed using this method in *E. coli*. The A+T content of the codon harmonized gene was altered from 75% (native *Pf*48/45) to 56%. This allowed the protein expression rate in *P. falciparum* to be mimicked since the translational machinery paused at the equivalent positions in *E. coli* as it would in *P. falciparum*. For the first time high yields of *Pf*48/45, in the correctly folded state were obtained. Furthermore, high immunogenicity and transmission blocking activity was observed in mice immunized with the adjuvants (Chowdhury *et al.* 2009).

An analysis of the A+T content of the n*gtpch*I, o*gtpch*I and h*gtpch*I sequences using the Bioedit sequence alignment editor version 5.0.9 (nucleotide composition) revealed a composition of 78.63%, 58.55% and 60.6%, respectively. The A+T content reduced upon codon optimization and codon harmonization as expected. Upon the induction of protein expression, an analysis of the levels of nGTPCHI, oGTPCHI and hGTPCHI were compared in different cell lines. Codon harmonization was observed to improve the expression of total and soluble protein fractions in all the cell lines based upon the relative quantity of protein bands assumed to correspond to the expressed proteins on the SDS-PAGE gels. However, these results do not prove that expression of nGTPCHI, oGTPCHI or hGTPCHI occurred. For this western analysis, mass spectrometry and enzyme activity investigations are required. Upon His-tag purification of soluble hGTPCHI, slightly greater levels of hGTPCHI were obtained in comparison to oGTPCHI and nGTPCHI. These results were confirmed by western analysis using a his-probe, which detected the his-tag attached to the target protein. However, an antibody specific to GTPCHI would have been of great value for western analysis. By determining the activity of the expressed proteins, a comparison of whether hGTPCHI produced more active protein than oGTPCHI is necessary to prove these results.

2.4.3 *E. coli* cell lines

The expression of nGTPCHI, oGTPCHI and hGTPCHI was assessed in different *E. coli* cell lines. This was evaluated in order to determine which cell line would allow for the greatest level of soluble protein expression. The *E. coli* BL21 cell lines in general are B protein expression strains that lack both the *lon* protease and the *ompT* outer membrane proteases. This cell line also contains a λDE3 lysogen that carries the gene for T7 RNA polymerase under control of the *lacUV*5 promoter. As such, the BL21 Star (DE3) cells were chosen for their high level of gene expression and easy induction. However, the problem of expressing this *Plasmodium* protein became evident during the attempt to express the nGTPCHI in these star cells. No detectable band corresponding to 48 kDa nGTPCHI was observed in these cell lines (Figure 2-16). Furthermore, the analysis of the relative quantity data generated for the oGTPCHI and hGTPCHI in Figure 2-17 revealed that the majority of the proteins were present in insoluble form as inclusion bodies (refractile aggregates of inactive, missfolded protein). Inclusion body formation can be minimized by various methods. This includes the use of enriched media for

growth, alternate host metabolism and by the co-expression with plasmid-encoded chaperones to name a few (Jonasson *et al.* 2002).

To try to improve the expression of soluble proteins, another *E. coli* cell line was assessed as an expression host. The m15 cell line has advantages such as higher expression of the lac repressor protein in comparison to the BL21 cell lines, hence upon IPTG induction, target protein expression is prolonged in comparison to the star cells. Based on the SDS-PAGE results in Figure 2-18 and Figure 2-19, attempts to improve soluble expression of nGTPCHI, oGTPCHI and hGTPCHI were observed since the levels of insoluble protein was greatly reduced in comparison to the levels of soluble protein based on the bands detected at 48 kDa. Using the m15 cell line, nGTPCHI expressed in soluble form (whereas in the star cell line no expression was observed) and at greater expression levels in comparison to the oGTPCHI. However, these results were false positives since the expression of nGTPCHI, oGTPCHI and hGTPCHI in the m15 cells could not be confirmed, as the proteins could not be detected using the western analysis of the affinity purified proteins (Figure 2-22). Furthermore, the expression levels before and then upon induction revealed similar results. Also, robust growth of a particular cell line had no implications on the level of expression obtained since it was observed that when using the m15 cell line steeper growth curves (in comparison to those of star and star 70 cells) were obtained. However, expression could not be confirmed using this cell line.

If nGTPCHI expression increased upon induction as such, the concealment of the His tag during the folding process could have explained the reason why nGTPCHI could not be detected via western analysis. General reports of weak or no interaction to metal ion affinity column caused for example by concealed His tags have been observed in other studies. This could be avoided by switching the position of the His tag (by using another vector) to the N-terminal of GTPCHI (Lee and Kim 2009). Furthermore an antibody against the target protein itself would have indicated whether nGTPCHI (as well as oGTPCHI and hGTPCHI) expressed.

2.4.4 *P. falciparum* molecular chaperone

Inclusion body formation is linked to the recombinant expression of soluble proteins (Sorensen *et al.* 2003). The levels of soluble nGTPCHI, oGTPCHI and hGTPCHI

expression in the star cells were not very significant, and the expression of the above mentioned proteins could not be confirmed when using m15 cells. Therefore, co-expression of *Plasmodium* molecular chaperone (hsp 70) with the target protein was evaluated. Previous demonstrations indicated that the co-transformation of bacterial cells with recombinant proteins and chaperones from the same specie as the target protein improved protein expression (de Marco and De Marco 2004). However, this strategy does not guarantee that the presence of molecular chaperones would improve the expression of soluble proteins. Foth *et al* (2003) and Ramya *et al* (2007) reported that *Pf*hsp70 proteins functions in the translation and export of proteins into the apicoplast (Foth *et al.* 2003; Ramya *et al.* 2007). Furthermore, the *Pf*hsp70 protein is involved in protein folding and revealed successful over-expression of protein in *E. coli* since this chaperone protected the host substrate from heat stress (Shonhai *et al.* 2007).

In order to assess whether the co-expression of GTPCHI using *Pf*hsp70 was effective, the expression of *Pf*hsp70 should be confirmed via western analysis using a *Pf*hsp70 antibody. However, no antibody was available but the expression of oGTPCHI and hGTPCHI was observed based upon the detection of the His tag attached to oGTPCHI and hGTPCHI during western analysis. A comparison of oGTPCHI and hGTPCHI expression in star and star 70 cells revealed that the expression of both oGTPCHI and hGTPCHI increased (almost 2-fold) in star cells.

Investigation of the western blot revealed the presence of a second band just below the 48 kDa band for hGTPCHI and oGTPCHI when using the star 70 cells. This band could possibly be an N-terminal truncated product of nGTPCHI. Analysis of DNA sequence (h*gtpch*I and o*gtpch*I) indicates that there are two extra start codons (AUG) present at position 55 bp (Met-19) and 162 bp (Met-75), which might have influenced protein folding and the initiation of translation. The band could also possibly correspond to an *E. coli* host protein since purification was not obtained to homogeneity. However, MS would be required to prove these observations.

Just as in star cells, nGTPCHI failed to express in the star 70 cells and this can be explained by the problem of codon bias during expression in *E. coli* cell lines. Furthermore, the presence of high order mRNA structures (e.g. hairpins) affects ribosome pausing and stacking, hindering the protein translation machinery. If this

occurs, the structural organization of the encoded protein may change (Wolin and Walter 1988; Yonath 1992). As such, the use of *Pf*hsp70 did not enhance the expression of nGTPCHI, oGTPCHI and hGTPCHI.

2.4.5 Auto-induction media

Grossman *et al.* (1998) and Studier (2005) revealed that protein expression levels increased when the pET expression system leaked at the stationary growth phase in comparison to induction with IPTG (Grossman *et al.* 1998; Studier 2005). The components of a non-inducing and inducing media controlled leaky expression. In comparison to the other expression studies, expression was followed for 48 hrs because cultures grown to saturation produce greater levels of target protein per volume culture in comparison to the conventional IPTG induction (Studier 2005). The western blot analysis (Figure 2-25), including the relative quantity data (Figure 2-26) generated for the levels of purified proteins revealed an increase in the levels of soluble hGTPCHI in comparison to soluble nGTPCHI and oGTPCHI in star 70 cells using the auto-induction media.

The m15 cell line was assessed in this study with the star 70 cell line. Expression of nGTPCHI in the m15 cell line was possible since western analysis detected the protein. Additional bands were detected on the western blot for both the hGTPCHI and nGTPCHI, which did not correspond to the 48 kDa band of GTPCHI. These bands could be *E. coli* host proteins however, MS of these specific bands needs to be conducted to determine their identity. Based upon the western blots, the use of the m15 cell line as expression hosts and auto-induction media did not improve expression of oGTPCHI and hGTPCHI in comparison to the other cell lines and media composition. When comparing the levels of purified proteins obtained using the auto-inducing media (Figure 2-24) in comparison to the YT-media (Figure 2-22), auto-induction did not enhance protein expression. Although, the use of auto-induction media was convenient, a problem with the production of soluble protein persisted. Berrow *et al*. (2006) stated that both IPTG (YT-medium) and auto-induction produce qualitatively similar results however, the use of auto-induction may produce greater levels of protein expression. Auto-induction is a simpler method for scale-up but the greater biomass obtained may hinder sample processing (Berrow *et al.* 2006).

Reasons for the expression of malarial proteins include the use of these proteins for structural research, drug screening, *in vitro* functional assays including antigen preparation for recombinant vaccines (Birkholtz *et al.* 2008a). GTP cyclohydrolase I catalyses the first reaction in the folate pathway and is therefore considered as a promising antimalarial drug target. Previous studies involved the identification and isolation of GTPCHI from *P. falciparum* (Krungkrai *et al.* 1985). Therefore, the expression of this protein was crucial for further assays to be conducted. It is accepted that proteins from *Plasmodium* are not present in sufficient quantities and therefore a heterologous expression system is required. However, the malaria parasite has features that cause expression in a heterologous host system to be problematic. As such, to ensure optimal expression of this protein the route of codon harmonization was followed. In summation, the expression of the oGTPCHI and hGTPCHI proteins were obtained and further purification could be possible by altering purification or protein isolation conditions. Further studies would involve determining the activity of GTPCHI. Current kinetic activity assays for GTPCHI is based on observing the formation of dihydroneopterin triphosphate (the product of GTPCHI).

Chapter 3
Cell-free expression of native, codon optimized and codon harmonized *P. falciparum* GTPCHI

3.1 Introduction

Proteomic research necessitates the availability of abundant proteins in their naturally folded state. Currently, *E. coli* is the preferred heterologous expression system. However, limitations with its' use include the formation of insoluble protein aggregates during the expression of eukaryotic proteins. At present, efforts have been made to improve these difficulties such as the utilization of engineered strains, fusion-tags etc. but these approaches cannot be applied universally to all proteins and thus require optimization. Owing to the prerequisite for higher expression of active protein, *in vivo* expression and cell-free protein synthesis techniques are employed. The cell-free system used most commonly is from bacteria (*E. coli*), yeasts, mammals, plants, insects or wheat germ and rabbit reticulocytes. There are limitations to the use of each technique including size limitations for chemical synthesis and *in vivo* expression is restricted to the production of proteins that are not harmful to the host cell. (Aguiar *et al.* 2004). Additionally, the quality, quantity, functionality of proteins produced and the time required for production are all dependent upon the choice of expression system used (Jonasson *et al.* 2002).

3.1.1 Expression host used for *Plasmodium* protein synthesis

As mentioned previously *P. falciparum* genes have a high A+T bias (approximately, 76% per gene), as such many of these genes encode repeated stretches of amino acid sequences, limiting the expression of *P. falciparum* proteins. The expression of *Plasmodium* proteins has never been completely problem-free. Studies by (Aguiar *et al.* 2004) revealed that when malaria genes were cloned into a gluthathione S-transferase fusion vector, as few as 39 out of 292 genes were expressed (Aguiar *et al.* 2004). Structural studies to express small cytosolic malaria proteins were conducted by Mehlin *et al.* (2006). Initially, the study involved using 1000 genes for expression however, only

300 of these genes were used and only 63 proteins expressed in soluble form (Mehlin *et al.* 2006). Additionally, Vedadi *et al*. (2007) achieved the expression of 38 proteins in soluble form from a total of 182 proteins and Mu *et al*. (2007) expressed 60.2% (65 of 108 genes) polymorphic malaria proteins (Jonasson *et al.* 2002; Mu *et al.* 2007; Vedadi *et al.* 2007).

The focus of this introduction will be placed on various expression host systems (specifically the cell free wheat germ expression system) used previously to express *Plasmodium* proteins including the advantages and disadvantages of their use.

3.1.2 Yeasts

These are single-celled fungal eukaryotes that are used for the production of "difficult" to express recombinant proteins, which require specific post-translational modifications (glycosylation) for folding. Yeasts have advantages such as being simpler to use and more cost effective in comparison to insect or mammalian cells. The most commonly used for the production of large quantity of recombinant protein are *Pichia. pastoris (P. pastoris*) and *Sacchromyces. cerevisiae* (*S. cerevisiae*) (Demain and Vaishnav 2009).

Protein expression in *S. cerevisiae* is placed under control of inducible promoters, such as a glucose-repressed alcohol dehydrogenase. The advantage of using this system is that the recombinant proteins produced are secreted (avoids inclusion body formation) from the yeast cell due to the fusion of the N-terminal to a pheromone. Previously, specific codons which are A+T rich were recognized as termination signals (i.e. polyadenylation) resulting in the formation of truncated mRNAs (encoding partial ORF in the absence of stop codons) (Certain and Sibley 2007). It is suspected that the yeast mRNA surveillance pathway would degrade the RNAs inhibiting the accumulation of the protein. This surveillance system involves the recruitment of aminoacylated-SsrA RNA to the ribosome site where a truncated mRNA is present. The nascent polypeptide chain is transferred to the Ala-charged tRNA region of SsrA. The truncated mRNA is then replaced by an ORF within SsrA. This translates until the stop codon and the tagged protein is degraded by proteases (Karzai *et al.* 2000). By mutating the genes (*hrp1* or *kap104*) required for degradation, the truncated RNA protein expression could be enhanced. *S. cerevisiae* mutants (*ura3*) were selected that were able to use the *P. falciparum* orthologue of URA3. Mutants strains were observed to express 100-fold more of *Plasmodium* proteins just by mutation of the specific genes (LaCount *et al.* 2009).

Yeast-complementation assays and PCR mutagenesis were also used to investigate the effects of mutations in DHFR in parasites resistant or sensitive to antifolate drugs (Ferlan *et al.* 2001). Furthermore, *P. falciparum* phosphoethanolamine methyltransferase was characterized similarly by scanning mutagenesis of Ala residues and determining the activity of the mutant enzyme using functional complementation. These investigations would lead to screening of potential inhibitors against these enzyme targets (Reynolds *et al.* 2008).

P. pastoris makes use of a closely regulated promoter from the alcohol oxidase I gene, which promotes foreign gene expression. Furthermore, the proteins produced can be secreted into the media making protein purification simpler (Cereghino and Cregg 2000). Glycosylation patterns are more extensive, up to 50-150 residues of N-linked mannose oligosaccharides, which normally causes a high antigenic response in patients. This can be circumvented by mutations that remove these sites in the *Plasmodium* sequence. However, the disadvantage of using *P. pastoris* is the requirement for chaperones for protein folding. Several malaria vaccine candidates have been expressed using this system. One such target is *Pfs*25, a transmission blocking vaccine. The expression study focused on the effects of overproducing *P. pastoris* protein disulfide isomerase to increase recombinant expression of *Pfs*25. Results revealed that a 3-fold increase in total protein production was obtained. The problems of using yeast expression systems occurs during the erroneous O- and N-linked glycosylation of *Plasmodium* proteins by the secretory pathway, since these native proteins are seldom glycosylated. To circumvent this, the overexpression of *P. pastoris* protein disulfide isomerase was used to decrease levels of glycosylation possibly due to the difference in tertiary structure of *Pfs*25 or the activity of the chaperone that restricts certain domains of the protein for glycosylation (Zou *et al.* 2003; Tsai *et al.* 2006). Additionally, the codon usage of the gene, use of strong promoters that causes the exhaustion of energy and precursors, gene copy number, translocation as well as folding (difference in yeast protein that assist folding from the natural host) and protein turnover affect protein expression in yeasts (Cereghino and Cregg 1999).

3.1.3 Baculovirus

Insect cells are popular as an expression system in conjunction with baculovirus technology. *Autographa californica* nuclear polyhedrosis virus is the lytic virus used most

commonly for this. The process involves cloning the specific gene into the vector, which is then co-transfected with viral DNA into insect cells (*Spodoptera frugiperda)* (Kost *et al.* 2005; Birkholtz *et al.* 2008). Baculoviruses possess a very strong polyhedrin promoter, which controls the foreign gene and infection of the insect cells enabling high expression of proteins. Various advantages exist, which include eukaryotic post-translational modifications, proper folding and high levels of protein expression (since the gene is placed under control of the viral polyhedrin promoter, allowing expression of 30% of cell protein) (Demain and Vaishnav 2009). The use of this system has been evident for the expression of a vaccine candidate, *P. falciparum* erythrocyte-binding protein (EBA-175) given the cys-rich motif within its' receptor-binding domain. Results indicated that the production of a 95% homogenous receptor-binding EBA- 175, which replicated the native EBA-175 binding to host erythrocytes and also induced antibodies to inhibit EBA-175 binding was obtained (Liang *et al.* 2000). The problem of inefficient secretion of proteins from these cells has been solved by the addition of insect secretion signals (honeybee melittin sequence). Also, incorrectly folded proteins caused by expression in the late infection cycle can be circumvented by harvesting at earlier periods (Demain and Vaishnav 2009).

Other disadvantages include glycosylation of *Plasmodium* proteins since N- and O-glycosylation occurs rarely in the parasite. Hyperglycosylation of certain proteins occur if high protein expression take place during the late viral stages of the baculovirus (Liang *et al.* 2000). Glycosylphosphotidylinositol-anchored proteins assist in the survival of the malaria parasite. Despite the presence of a glycosylphosphotidylinositol attachment site the parasite CSP was absent of this moiety because this system is not ideal for production of GPI anchored proteins. The use of the baculovirus system is also time-consuming, expensive and technically more demanding than *E. coli* (Birkholtz *et al.* 2008).

3.1.4 Plant cells

Agrobacterium tumefaciens (*A. tumefaciens*) a plant infecting bacterium, has been used to mediate gene transfer into plant cells. *A. tumefaciens* contains a low copy number of tumour inducing Ti plasmid. This plasmid transfers a segment of its DNA (T-DNA) into the plant genome via repeated sequences flanking this segment (Koprowski *et al.*, 2001). Plant cells display some advantages, such as the absence of contamination with

mammalian pathogens, the simplicity of manipulation, protein modification machinery, safety (no plant viruses are pathogenic to humans) and cost effective production (requires water, minerals and sunlight) (Demain and Vaishnav 2009).

Two plant expression systems exist: transgenic plants and plant virus vectors. Transgenic plants involves the direct expression of foreign genes, which is integrated stably into the plant genome, the genes are localized to and expressed via nuclear systems and are also inherited. In contrast, plant viruses involve transient foreign gene expression, which is the integration of the foreign gene into the viral but not the plant genome. Localization and expression is directed to the cytoplasm and amplification occurs via virus replicase and is not inherited. Expression is restricted to single transgenic plants. However, expression can be applied to all susceptible plants. The cycle time, scale-up and the cost is greater for the production of transgenic plants in comparison to plant virus vectors. Previous investigations involved the expression of various antigens as vaccine candidates such as RSV G and F proteins, the VP6 protein of rotavirus and an epitope from the major surface antigen of *P. falciparum* (PfMSP-1) (Koprowski and Yusibov 2001). Ghosh *et al*. (2002) utilized a tobacco plant to express a proteolytic product of *P. falciparum* MSP1; the same immunogenic properties were obtained in comparison to the *E. coli* expressed form of the protein. Limitations for the use of plant cells include low levels of expression and hence accumulation of expressed proteins, which may be have incorrect post-translational modification (Demain and Vaishnav 2009).

3.1.5 Mammalian cells

Mammalian expression systems are used for the expression of proteins requiring specific post-translational modifications. Since proteins are properly folded, renaturation protocols are not required. However, a disadvantage is that lower yields of proteins produced decreases the use of these cells (Demain and Vaishnav 2009). The use of mammalian cells for expression of *Plasmodium* proteins is an advantage for targeting expression to subcellular compartments and cell surfaces. Parasite surface ligands that triggers host cell binding, which aids in identifying merozoite proteins (EBA-175, AMA-1) and their related domains for binding to host erythrocyte (Tolia *et al.* 2005) were expressed in these cells. The COS-7 green monkey kidney cell-line has been used, but the common CHO-K1 and HeLa cell lines have also been utilized. In most cases, the

Plasmodium gene sequences were seldom altered to resolve the A+T-codon bias and eliminate possible N-glycosylation sites given that It was previously presumed that mammalian cells are more accommodating for the expression of A+T biased proteins. These cells are however, rarely used for functional characterization of *Plasmodium* proteins. This is because the formation of functionally compromised mutants were unsuccessful for complementation experiments and may possibly required RNAi-mediated post-transcriptional silencing (Collins *et al.* 2007; Birkholtz *et al.* 2008). Furthermore, animal cell lines do yield proteins that are active but production is time-consuming resulting in the low expression and the high cost of the culture medium (Peterson *et al.* 2002).

3.1.6 Cell-free expression system

Early investigations revealed *in vitro* translation systems do not require cell integrity for protein synthesis/ expression. In these systems mRNA transcription from the DNA template occurs *in situ* (Katzen *et al.* 2005). The first wheat germ cell-free translation system may have developed during investigations of activating protein synthesis in the imbibitions phase of seed germination (Marcus and Feeley 1964). Later, progress was made on a translation system using a commercially available wheat germ as a catalyst for protein synthesis (Erickson and Blobel 1983). At this time the wheat germ cell-free translation machinery was incapable of synthesizing large quantities of proteins required for biochemical characterization and translation only lasted a few hours. Following this, Spirin *et al.* (1988) reported protein synthesis with the use of wheat crude extracts for prolonged periods of time (20 hr) using the continuous-flow-cell-free method (Spirin *et al.* 1988). The wheat germ embryo particles were then washed extensively before homogenization to produce a stable extract that allowed translation for greater than 60 hr (Madin *et al.* 2000). At present the preparation of the wheat germ extract has improved allowing translation for more than 2 weeks (provided continuous mRNA and substrate supplies are available) and accounts for the type of wheat, the methods for seed storage, embryo particle selection and homogenization etc. (Takai *et al.* 2008). As such, successful protein production using the crude lysate is dependent on the efficiency of mRNA transcription, the supply of aminoacyl tRNA, an energy source and translation factors. This can be segregated further in that the transcription of mRNA requires template DNA, ribonucleotides and enzymes (T7 and SP6 RNA polymerase). Translation requires initiation, elongation and termination factors and constituents (tRNA,

amino acids and ATP) needed for tRNA aminoacylation. The energy system requires the presence of substrates and enzymes such as phosphoenolpyruvate (PEP) / phophoenolpyruvate kinase (PK) and creatine phosphate (CP) / creatine kinase (CK). The cell-free expression system is composed of all these constituents making it ideal for the production of proteins in milligram quantities (Shimizu *et al.* 2006).

The cell-free system has broadened avenues for protein synthesis since it allows for the rapid and simple production of proteins that are in their correctly folded states. The requirement for *in vivo* cloning is avoided since PCR gene amplicons can be utilized, permitting large-scale parallel protein expression. Additionally, post-translational modifications (disulphide bond formation) can be included and membrane proteins can be synthesized (Spirin 2004). This system makes use of a cell lysate that consists of a cocktail of ions, macromolecules, amino acid supplements and energy sources needed for translation from added DNA or RNA templates (Figure 3-1). These lysates or extracts can be prepared from a diverse array of expression hosts that include *E. coli*, yeast, *Xenopus* oocyte, mammalian cells, rabbit reticulocytes and wheat germ. Depending upon the requirements for protein folding (co-translational, post-translational modifications) adequate cellular factors are added, such as for the production of proteins with co-translational and post-translational modifications. Chemically modified amino acids can be incorporated in the nascent polypeptide chain to specific regions during the process of translation. Depending on the characteristics of the expressed protein, post-translational modifications include glycosylation, signal peptide cleavage, acetylation, phosphorylation, degradation pathway, myristoylation, prenylation, arginine methylation, and proteolytic processing (Endo and Sawasaki 2006; He 2008). The advantages of using the cell-free system include the high success of protein expression producing active proteins in the natural conformation. The production process is automable (high throughput) and proteins that are toxic to expression cell lines can be synthesized (www.genecopoeia.com/product/cellfree).

The downstream applications of using proteins synthesized via this system include protein detection, purification, characterization and structure determination (NMR or X-ray crystallography). By incorporating chemically modified amino acids to the nascent polypeptide, drug-like peptides can be produced. The cell-free system has also been utilized for the diagnosis of human genetic diseases initiated by translation-termination

mutations e.g. the protein-truncated test used to identify open reading frames (Spirin 2004). In this test various types of gene mutations including splice-site alteration, in-frame deletions (>25bp), out-of-frame deletions and nonsense mutations are analyzed rapidly with high sensitivity from large cDNA sequences. Both transcription and translation of the PCR product is achieved using an upstream primer with a RNA polymerase promoter and a translation initiation signal. The polypeptides that form are then analyzed by SDS-PAGE and compared in size to the wild-type (the gene containing a mutation would reflect a difference in the length of the protein in comparison to the wild-type when translated) (Roest *et al.* 1993). There are disadvantages that exist for this system such as, the accumulation of free phosphate from the reaction coupled to ATP, protein missfolding and the high cost involved. However, with new innovations this can be circumvented. This could include integrating membrane vesicles containing oxidative phoshorylation enzymes, ADP can be recycled to reduce free phosphate levels, the cost can be reduced by discovering an efficient energy regeneration system and miss-folded proteins can be eliminated by the addition of lysates that could possibly contain chaperones or detergents (Katzen *et al.* 2005).

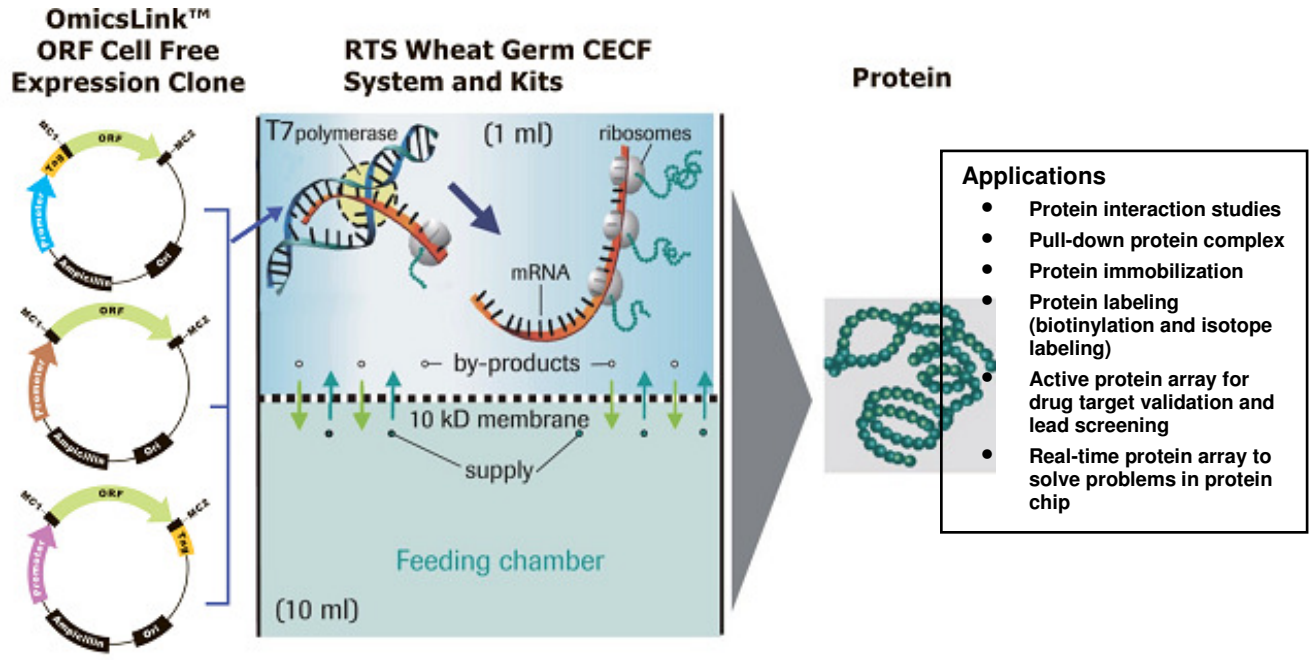


Figure 3-1: The process of cell-free synthesis and analysis of proteins (www.genecopoeia.com/product/cellfree).

Previous investigations on the overexpression of the *Plasmodium* DHFR-TS in heterologous host systems were unsuccessful at producing soluble protein in high quantity. The reason being the host would alleviate the levels of the toxic protein by losing the full-length of the coding sequence without positive selection. Genetic selection improved the expression however, accompanying the DHFR-TS would be truncated products, which were then eliminated by affinity purification. This strategy was applicable to the *Plasmodium* DHFR-TS, however this would not be useful for the production of other *Plasmodium* proteins since the expression and purification conditions varies. Mudeppa *et al.* (2007) investigated the production of a complete, active *Plasmodium* DHFR-TS using the wheat germ translational system. In general, these host cells are cultured without the target gene, thereby avoiding antiproliferative toxicity and the reaction conditions are flexible in terms of the volumes, temperature and kinetics in this system. This system therefore allows large-scale protein synthesis. For these reasons, the authors selected the wheat germ system and not rabbit reticulocytes (Mudeppa *et al.* 2007). The *Plasmodium* gene was engineered to possess an SP6 RNA polymerase and an omega (Ω) sequence for the wheat germ ribosome binding site. For large-scale protein production, a dialysis-based translation method that permits free exchange of required reagents around the cell lysate was utilized. Additionally, with the use of a longer pEU3b plasmid-derived RNA template (last longer for protein synthesis), the *Pf* DHFR-TS protein production continued for 28 hrs. A total yield of 100 µg of translation product was obtained and in soluble form. The catalytic activity of TS measured superiorly in comparison to previous attempts where only the DHFR activity was determined since TS activity is a better indication of a correctly folded protein. The protein band corresponding to DHFR-TS was visible on the SDS-PAGE gel using coomassie blue staining. This detection method (staining) was previously impossible even with a codon-optimized DHFR-TS. As such, owing to the successful use of the wheat germ expression system in the large scale production of functional DHFR-TS protein, the authors believe that this system can be applicable to the synthesis of many other *Plasmodium* proteins (Mudeppa *et al.* 2007).

Several malaria vaccine candidates such as *Pfs*25, *Pf*CSP and *Pf*AMA-1 were expressed in their native and codon optimized gene forms and their immunogenicity was found to be effective in mice models. Certain *P. falciparum* blood stage proteins were also expressed with an average yield of 1.9 g and 65% solubility (Tsuboi *et al.* 2008).

Kato *et al* (2008) expressed and purified the *P. falciparum* protein kinase 2 (PfPK2) in an attempt to elucidate the calcium signal of *Plasmodium* host invasion. PfPK2 was expressed as a GST fusion protein using wheat germ cell-free protein synthesis. The purified protein possessed protein kinase activity in a calcium-calmodulin dependent way. (Kato *et al.* 2008)

The wheat germ expression system displays a lowered preference for codon usage and as such has been used to express *Plasmodium* proteins that possess a majority A+T richness genome (like mammalian cells). Furthermore, proteins folded into their native state and large quantities were produced. This system was optimized to improve the translational initiation and mRNA stability (Endo and Sawasaki 2006; Mudeppa *et al.* 2007).

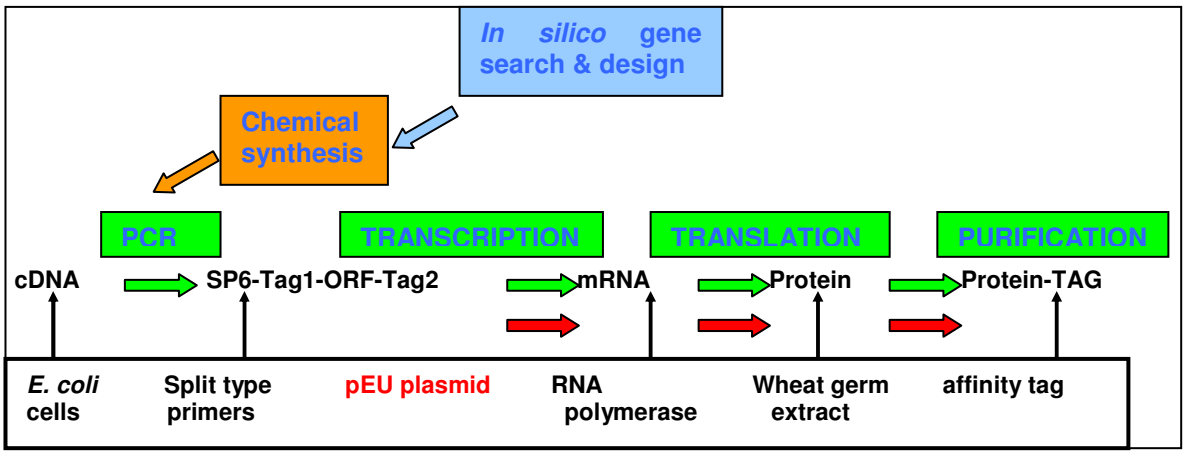


Figure 3-2: Protein production route using the wheat germ cell-free system. Initially, the target gene is selected from a DNA database. The *in silico* constructed DNA should contain an SP6 promoter, an open reading frame (ORF) coding region, a reporter, and/or a purification tag (SP6-Tag1-ORF-Tag2). The requirements for each process are indicated by the black arrow. After RNA production (transcription from DNA) this can be used directly for bilayer translation system. The red arrows indicate the second method for protein production. The target gene is cloned into pEU plasmid and the mRNA is transcribed. In the translation step, the conditions (e.g. potassium and magnesium concentrations and incubation temperature) are optimized and proteins are produced (adapted from Endo and Sawasaki. 2006).

No previous activity and kinetic properties could be inferred from the *P. falciparum* nGTPCHI, owing to the inadequate quantities of soluble protein. Chapter 2 focused on the optimization of nGTPCHI expression from the o*gtpch*I and h*gtpch*I gene using *E. coli* as a heterologous host. This chapter concentrates on the synthesis of the hGTPCHI, oGTPCHI and nGTPCHI using an alternative expression host, the wheat germ system. Since, the wheat germ expression system relies on the protein production from RNA

transcription products, the recombinant clones of n*gtpch*I, o*gtpch*I and h*gtpch*I could be used directly and did not require sub-cloning into another expression vector.

3.2 Methods and materials

3.2.1 Linearization of the pET22b- *Pfgtpch*I

Sections 2.2.11 for h*gtpch*I and 2.2.9 for o*gtpch*I and n*gtpch*I describe the sub-cloning of the inserts into pET22b, the screening procedures and section 2.2.13 the sequencing of the positive recombinant clones. Plasmid DNA was isolated using Zyppy™ Plasmid Miniprep kit (refer to section 2.2.8.2). The DNA templates (~0.5 µg/µl) of h*gtpch*I- o*gtpch*I- and n*gtpch*I-pET22b were linearized with *Bsa*I (GGTCTC(N)$_1$▼) (New England Biolabs, UK) prior to transcription to ensure the RNA produced were of distinct lengths. The linearization reaction consisted of 0.1 µg/µl plasmid DNA, NEB buffer 4 (50mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM DTT pH 7.9) and 0.2 U/µl restriction enzyme to a total volume of 50 µl. This reaction was then incubated for 16 hrs at 50°C. Heat inactivation of *Bsa*I was performed at 65°C for 20 min. To visualize complete digestion, both the intact plasmid and the digested plasmid were analyzed on a 1.5% (w/v) agarose gel (Merck-Saarchem, South Africa) containing 0.5 ng/ml ethidium bromide. The separated bands corresponding to the 6674 bp for linearized h*gtpch*I-, o*gtpch*I- and n*gtpch*I-pET22b were recovered from the agarose gels as in section 2.2.9.2. The quantity of purified, linearized plasmid DNA required for the following procedures was 10 µg or more. The DNA concentrations were determined spectrophotometrically at 260nm using the GeneQuant Pro UV/Vis Spectrophotometer (Amersham Biosciences, Sweden) as in section 2.2.3. The concentration of purified DNA obtained was 153 ng/µl, 170 ng/µl and 184 ng/µl for h*gtpch*I-, o*gtpch*I- and n*gtpch*I-pET22b, respectively.

3.2.2 Creating blunt ended plasmid DNA

Prior to RNA production, DNA should possess blunt ends to ensure that extraneous transcripts generated as a result of transcribed overhangs do not occur. However, in some cases such overhangs are generated as a result of the restriction enzyme used (*Bsa*I linearized products contain a 4 base overhang). Blunt ends were subsequently created using DNA polymerase I large (Klenow) fragment (New England Biolabs, UK)

which, as a proteolytic product of *E. coli* DNA polymerase, has no 5'-3' exonuclease activity but maintains its 3'-5' exonuclease and polymerase activities. To create blunt ends, an *in vitro* transcription reaction consisted of 10 µg/µl linear template, 5X T7 transcription buffer (proprietary mix) and 0.2 U/µg DNA polymerase I large (Klenow) fragment in 0.2 ml thin-walled PCR tubes. This was incubated at 25°C for 15 min in a Gene-AMP PCR System 9700 (Perkin Elmer, USA).

3.2.3 *In vitro* transcription

RNases are stable enzymes that cleave RNA in the absence of cofactors. These enzymes were eliminated by heat sterilization of all glassware, tips, eppendorf tubes, the use of RNase-free H_2O, the use of double latex gloves, as well as keeping RNA on ice in a reverse laminar flow cabinet. The large-scale production of high quality RNA is required for *in vitro* translation in the wheat germ expression system (Promega, USA). Large-scale RNA production was performed using the RiboMAX™ Large Scale RNA production systems using T7 (Promega, USA). Transcription was carried out using the T7 promoter of the pET22b vector containing the *P. falciparum* h*gtpch*I, o*gtpch*I and n*gtpch*I inserts such that the gene was under the control of the phage RNA polymerase promoter. Immediately after blunt end generation, transcription was induced by the addition of 0.25 mM ribonucleotide phosphate (rNTPs) rATP, rCTP, rGTP, rUTP) and T7 enzyme mix (proprietary mix consisting of T7 RNA polymerase, recombinant RNasin® ribonuclease Inhibitor recombinant inorganic pyrophosphatase and HEPES buffer pH 7.5) to a total volume of 100 µl. This was followed by incubation at 37 °C for 4 hrs in the Gene-AMP PCR System 9700 (Perkin Elmer, USA). A 0.1 U/µg (assuming that 10 µg of RNA was produced) RNase-free DNase was added to the reaction and incubated for 15 min at 37 °C to eliminate any inhibiting template DNA and other contaminating products that was not RNA. The RiboMAX™ system boasts the production 2-5 mg/ml of high quality RNA for use in cell free expression systems. It has advantages such as the incorporation of inorganic pyrophosphate, high concentrations of rNTPs and magnesium for use with either SP6 or T7 RNA polymerase and includes ribonuclease inhibitor. The synthesized RNA was thereafter purified.

3.2.4 High pure RNA cleanup

RNA purification was carried out using the RNeasy® mini kit (Qiagen, Germany). Buffer RLT was added to the reaction tube containing the RNA produced in a 1:3.5 ratio and

mixed well. This buffer contains guanidine salt, which has a high positive charge, ensuring a strong enough electrostatic interaction to occur so that the wash buffer that is used eliminates only contaminating salts, and unwanted DNA nucleotides that would affect downstream interactions. A 100% (v/v) absolute ethanol (Merck-Saarchem, South Africa) solution was added in 1: 25 ratio in order to precipitate DNA, mixed and thereafter transferred to the provided spin-columns and subjected to centrifugation at 16 000 xg for 15 sec. To wash the RNA, 500µl of buffer RPE (proprietary mix) was added to the column and centrifuged as mentioned above. This wash step was then repeated for 30 sec to ensure that no residual ethanol remained. RNA was eluted in 50 µl of RNase-free water by centrifugation at 16000xg for 1 min. To maximize the RNA concentration the eluted RNA was suspended in nuclease free water and re-loaded onto the same column and centrifuged as mentioned. RNA concentration was determined on the GeneQuant spectrometer at a wavelength of 260 nm. The concentration of RNA was determined by the formula: 44 µg/ ml x A_{260} x dilution factor (absorbance at 260nm corresponds to 44 µg of RNA/ 1 ml). Reading the absorbance at 260/280 thereafter was used to assess the levels of purity.

3.2.5 Cell-free protein expression

High yields and purity (260/280>2) of RNA of h*gtpch*I- o*gtpch*I- and n*gtpch*I-pET22b were obtained using the RiboMAX™ RNA production system (Promega, USA). The wheat germ extract plus (Promega, USA) previously stored as 30 µl aliquots at –70°C were thawed on ice and mixed gently by pipetting. In a 0.2 ml PCR tube, 30 µl of wheat germ extract plus (90 mM potassium acetate, 1.5 mM magnesium acetate, 0.5 mM spermidine, 5 mM dithiothreitol, 0.08 mM amino acid mix, 1.2 mM ATP, 0.1 mM GTP, Creatine phosphokinase 50 µg/ml and 10 mM creatine phosphate) (proprietary mix); 0.24 µg/µl purified RNA and RNase-free water to a total of 50 µl were assembled. The wheat germ extract plus (Promega, USA) system contains tRNA, ribosomes, amino acids, and initiation, elongation and termination factors needed for protein production and potassium acetate and magnesium acetate at sufficient concentrations to enhance translation of RNA. As a negative control, RNA was replaced with RNase-free water. The reaction contents were then incubated at 25 °C for 2 hrs using the Gene-AMP PCR System 9700 (Perkin Elmer, USA), to ensure controlled temperature of expression. Synthesized proteins were stored at –20°C.

3.2.6 SDS-PAGE analysis

A 4% (w/v) acrylamide-methylene-(bis) acrylamide stacking gel and a 10% (w/v) acrylamide-methylene-(bis)acrylamide, separating gel (refer to section 2.2.18) was prepared for protein analysis. The translation reaction was added to SDS sample buffer at a ratio of 1:4. This was denatured at 95 ℃ for 5 min and analyzed onto a SDS-PAGE gel. The gel was run in the Minigel G-41 system (Biometra, Germany) consisting of SDS-running buffer (0.1% (v/v) SDS, 0.25 M Tris, 192 mM Glycine buffer (pH 8.3)) at an initial 66 V followed by 100V after the proteins have separated from the stacking gel. In order to visualize the protein bands on SDS-PAGE, gels were stained, fixed and analyzed as in section 2.2.17.

3.3 Results

The wheat germ expression system was assessed in order to evaluate whether the expression levels of nGTPCHI, oGTPCHI and hGTPCHI could be improved in comparison to using *E. coli* as the expression host.

3.3.1 Linearization of the pET22b- *Pfgtpch*I

For the transcription of RNA by RNA polymerase (supplied by the RiboMAX^{TM} system) using the T7 promoter of pET22b, the plasmids containing the inserts were linearized using restriction enzymes that would cleave the plasmid only once and at a site closest to the T7 promoter site. As such plasmid DNA of h*gtpch*I- o*gtpch*I- and n*gtpch*I-pET22b were isolated and digestions were performed using *Bsa*I as indicated in Figure 3.3. Linearized plasmid DNA (h*gtpch*I- o*gtpch*I- and n*gtpch*I-pET22b) was excised from the 1.5 % (w/v) gel and purified.

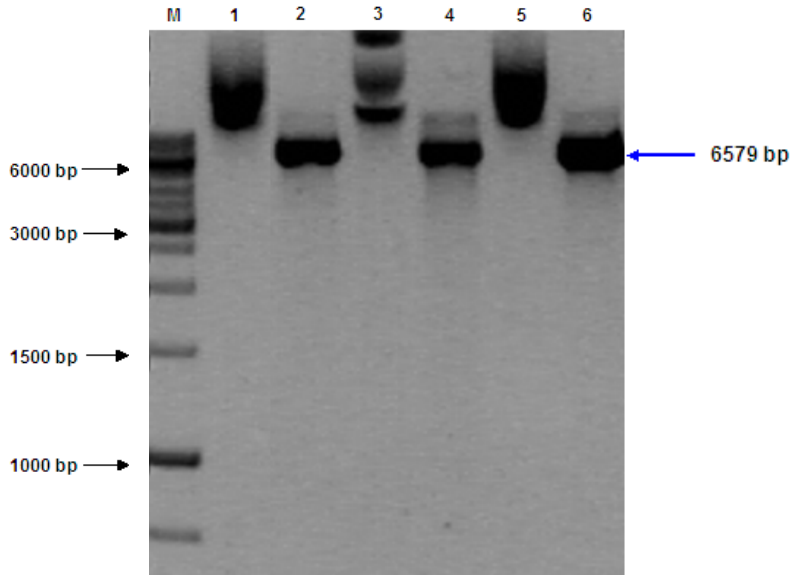


Figure 3-3: The linearization of h*gtpch*l-pET22b, o*gtpch*l-pET22b and n*gtpch*l-pET22b. M represents the molar mass marker (Genruler 1kb DNA ladder, Fermentas Luthania). The expected bands (bp) when digested and loaded on a 1.5 % (w/v) agarose gel after gel electophoresis. Lanes 1, 3 & 5 indicates the presence of undigested plasmid DNA h*gtpch*l-pET22b, o*gtpch*l-pET22b and n*gtpch*l-pET22b and lanes 2, 4 & 6 represent the digested plasmid DNA with *Bsa*I. The blue arrow indicated the linearized plasmid DNA.

3.3.2 RNA yield and purity

RNA concentrations were determined on the GeneQuant spectrometer at a wavelength of 260 nm. The concentration of RNA indicated in Table 3-1 was determined by the formula: 44 µg/ ml x A_{260} x dilution factor (absorbance at 260nm corresponds to 44 µg of RNA/ 1 ml). Reading the absorbance at 260/280 thereafter was used to assess the levels of purity.

Table 3-1: The concentration of purified RNA and the purity (260/280) obtained for h*gtpch*l-pET22b, o*gtpch*l-pET22b and n*gtpch*l-pET22b

	[RNA] µg/µl	260/280
n*gtpch*l-pET22b	68.80	1.8
o*gtpch*l-pET22b	88.34	2.0
h*gtpch*l-pET22b	72.63	2.0

The purity factor of RNA produced was 1.8, signifying no protein contamination or RNA degradation.

3.3.3 Cell free expression system

A total of 12µg of purified RNA was then required for protein synthesis using the wheat germ extract plus. Total protein samples from the crude lysate were analyzed on a 10 % (w/v) SDS-PAGE gel.

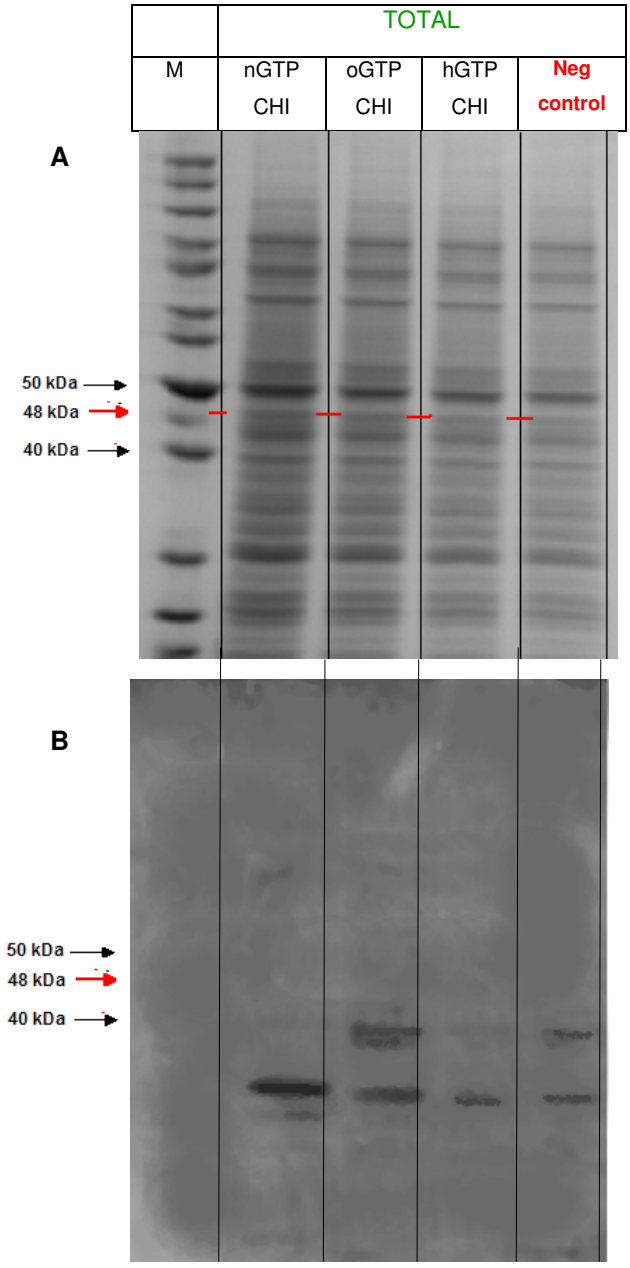


Figure 3-4: Expression of the nGTPCHI, oGTPCHI and hGTPCHI proteins in wheat germ extracts where A) 10% (w/v) SDS-PAGE and B) western blot analysis M: Protein Molecular weight marker (Fermentas, Luthania) hGTPCHI, oGTPCHI and nGTPCHI samples were loaded as total wheat germ extracts. The red arrow depicts the position of the expected GTPCHI band of 48kDa. His- tagged proteins were detected using His Probe-HRP (horseradish peroxidase) and detected via chemiluminescence.

No expression of nGTPCHI, oGTPCHI and hGTPCHI proteins was observed based on the comparison of expression of the negative control (Figure 3-4 A). The expression profiles of all lanes are similar and bands on the western analysis (Figure 3-4 B) detected proteins smaller than 48 kDa (size of GTPCHI). This indicates that the wheat germ extract plus system was unsuccessful for the expression of nGTPCHI, oGTPCHI and hGTPCHI.

3.4 Discussion

It has been commonly documented that the *E. coli* translational system is capable of sustaining the expression of prokaryotic and smaller eukaryotic proteins. However, specific domains of proteins are inclined to fold incorrectly using *E. coli* as a host and this causes the formation of inclusion bodies (Carrio and Villaverde 2002). This is the case with the expression of certain *Plasmodium* proteins, as such the wheat germ cell-free protein expression system was used for its practicality in use to produce *Plasmodium* proteins since it is eukaryotic in nature. The wheat germ extract plus (Promega, USA) was utilized for *P.falciparum* nGTPCHI, oGTPCHI and hGTPCHI protein expression. Briefly, this entailed the linearization of plasmid DNA containing the T7 promoter region. This linearized DNA was used as a template for transcription to RNA. Adequate amounts of plasmid DNA were isolated, linearized appropriately, purified directly from the agarose gel to eliminate residual undigested DNA (Figure 3-3). Purified RNA was added to the wheat germ extract plus and protein synthesis was initiated at 37 ℃. The purity of RNA was detected to be of high quality as shown in Table 3-1. Protein synthesis was initiated according to the manufacturers instructions. However, based on Figure 3-4, which was the SDS-PAGE analysis of the synthesized proteins neither nGTPCHI, oGTPCHI nor hGTPCHI proteins were produced using this wheat germ expression system. A negative control containing no mRNA was used to distinguish for protein expression. The expression profile of nGTPCHI, oGTPCHI and hGTPCHI were replicas of the negative control and protein bands detected via western analysis were not that of GTPCHI, indicating GTPCHI protein synthesis was unsuccessful. The detected bands could be analyzed via mass spectrometry to determine if they were in fact wheat germ proteins. Furthermore, a positive control for the wheat germ system is essential for further use of this system for protein synthesis

Based on the concentration and the purity of the RNA obtained after it was synthesized and thereafter purified, the process of transcription may not have been problematic this could been assessed by agarose gel electrophoresis. The failure of nGTPCHI, oGTPCHI and hGTPCHI synthesis could be at the protein production level. Either *Plasmodium* proteins fail to express using the wheat germ expression system or experimental conditions need optimization.

In order to optimize experimental conditions, other studies that were successful at protein synthesis when using the wheat germ system should be evaluated. The wheat germ extract plus system (Promega, USA) has been successful at providing a platform for detecting protein interactions at a fast rate. Porter *et al*. (2007) attempted to reassemble split-protein reporters (e.g. green fluorescent proteins and lactamases) in order to determine the ability of these reporters to function in the wheat germ cell-free system. This system allowed for rapid and sensitive determination of biomolecular interactions without the need for laborious experimental methods (cell culture, purification etc.). Furthermore, mRNA transcription was also conducted using the RiboMAX Large Scale RNA Production System-T7 (Promega, USA). However, the mRNA sequence possessed a Kozak sequence, 32 bp hairpin loop and the method used for translation and concentrations of components varied since biomolecular interactions were studied (Porter *et al.* 2007). Another study included the expression of various different enzymes (glutathione-S-transferase, β- galactosidase and luciferase) ranging from 26-135 kDa using the wheat germ extract plus system (Promega, USA). In comparison to conventional wheat germ extract the wheat germ extract plus synthesized 8.5- to 17-fold more protein (http://www.promega.com/pnotes/90.pdf).

This indicates that the wheat germ extract plus system (Promega, USA) has been used for the synthesis of other proteins. However, in both articles by (Porter *et al.* 2007) and Hurst *et al*. the concentrations of mRNA used were 0.5 pmol and 0.16µg/µl, respectively and it has been suggested to determine the optimal mRNA concentration required for translation by using a range of mRNA concentrations. Even though the purity of mRNA before translation was of high quality, the concentration of template mRNA used may have resulted in no expression. The wheat germ extract plus system (Promega, USA) suggests the usage of 6-12 µg mRNA in a final volume of 50µl and lower mRNA concentrations were not evaluated.

Not only is the concentration of template important but It has also been observed that in the wheat germ cell-free system, the template mRNA molecules are rapidly degraded during translation and this could be confirmed by gel electrophoresis (Takai *et al.* 2008). This problem can be circumvented by the use of the bilayer translation reaction. The bilayer translation reaction allows the continuous supply of substrates and the diffusion of by-products through a layer between the translation and substrate. High protein yields are obtained and reactions lasted longer than the conventional method (Sawasaki *et al.* 2002). *Pfs*25, *Pf*CSP, and *Pf*AMA1 proteins are *Plasmodium* vaccine candidates that were synthesized using this bilayer translation reaction. Results revealed that approximately 75% of these genes expressed proteins in soluble form. Furthermore, the authors expressed proteins from 567 cDNA clones, from sporozoites, merozoite and gametocytic developmental stages and a surprising 84% were obtained in high quality and soluble form (Tsuboi *et al.* 2009). These high levels of *Plasmodium* protein expression studies were performed using an alternate wheat germ kit, the Endext technology wheat germ expression premium kit (Cell free Sciences, Japan) (discussion to follow).

The Wheat Germ Extract Plus (Promega, USA) has been improved for translation of mRNAs that have uncapped 5´ ends, capped or those possessing 5' viral leader enhancers. In theory nuclear processing of mRNAs adds a cap structure to the 5' end and a poly (A) tail to the 3' end, which aids as binding sites for the cap-binding protein and poly (A)-binding protein. A eukaryotic initiation factor (eIF 4E) then binds the 5' cap and poly (A) tail, which together interacts with eIF4G forming a closed loop. These contributions including the addition of a poly (A) tail or 3´ viral enhancer to the 3´ end extensively increased protein expression (Gallie 1998). As such, the protein-coding regions can be cloned into a Wheat Germ Flexi® Vector (Promega, USA) that can be used with the Wheat Germ Extract Plus (Promega, USA) system and would ensure the incorporation of 5´ and 3´ elements into the mRNA. This was observed during expression studies conducted by Hurst *et al*., where mRNA templates with the barley yellow dwarf virus (BYDV) sequence enhanced the synthesis of GST more than 3-fold than templates without the enhancer sequence or with poly (A) tails (Hurst *et al*. Promega notes). The BYDV enhancer region stimulate translation in this system since

these elements; originally from the RNA plant virus interact with one another synergistically to form a closed loop (Guo *et al.* 2001).

The malaria vaccine candidates, which were successfully synthesized using the wheat germ system, involved cloning the target genes within a pEU-E01-GST (a vector with an N-terminal GST tag followed by a tobacco etch virus protease cleavage site) and these vectors were designed specifically for expression using the Endext technology wheat germ expression premium kit (Cell free Sciences, Japan) (Tsuboi *et al.* 2008). The general technique involves the removal of the 5'-G cap and the poly(A)-tail (inhibits translation initiation) of the mRNA's. Proteins can also contain an affinity tag (to allow protein purification) and a pEU plasmid (plasmid of Ehime University) containing the gene to be transcribed (Endo and Sawasaki 2006). The 5'- and 3' untranslated regions (UTR's) of eukaryotic mRNA are crucial for the control of gene regulation. As such, in the absence of the 5'cap and 3' tail mRNA translation was improved by introducing a 5' enhancer element (71 nucleotides called an Ω sequence of the tobacco mosaic virus). The pEU plasmid originated from pSP65 and was designed as such to incorporate the 5' Ω sequence and a long 3'UTR. This expression vector allowed for the production of proteins in milligram quantities and the plasmid did not require linearization before transcription (Takai *et al.* 2008). Furthermore, the translation system can be fine tuned in terms of ion concentrations and incubation temperature to improve expression (Endo and Sawasaki 2006; Mudeppa *et al.* 2007).

The transcription mixture utilized by Tsuboi *et al.* (2008) for successful expression of malaria vaccine candidates involved the use of plasmid DNA, HEPES-KOH magnesium acetate, spermidine, dithiothreitol, each nucleoside triphosphates, SP6 RNA polymerase and RNasin that was incubated for 6 hrs at 37℃. Transcribed RNA was then resuspended with wheat germ extract that was supplemented with creatine kinase in a 6-well plate and incubated at 26℃ for 12 hrs (supplied by CFS Co., Ltd., Matsuyama, Japan) (Tsuboi *et al.* 2008) (refer to Appendix H). A major difference noted was the additional incubation time of both the transcription and translation reactions using this system and this may account for the results obtained.

DHFR-TS (folate biosynthesis enzyme) was also produced with high yields in soluble form using the Endext technology wheat germ expression premium kit (Cell free

Sciences, Japan) and a dialysis-based translation method (it permits free exchange of required reagents around the cell lysate). The *Plasmodium* gene was engineered to possess an SP6 RNA polymerase, an Ω sequence for the wheat germ ribosome binding site and a longer pEU3b plasmid-derived RNA template (last longer for protein synthesis) (Mudeppa *et al.* 2007). However, the wheat germ extract plus (Promega, USA) is also capable of synthesizing high levels of proteins (based on previous studies) and by repeating the synthesis using altered mRNA conditions, the synthesis of GTPCHI could be possible. Since the template mRNA rapidly degrades during translation when using the wheat cell-free system, a dialysis-based translation method can be used. Furthermore, protein synthesis can be continued for longer periods of time hence greater quantities of proteins will be produced. The *Plasmodium gtpch*I gene can also be cloned into a Wheat Germ Flexi® Vector (Promega, USA) to incorporate the elements required to improve translation and lower concentrations (pmol) of mRNA template can be assessed for optimal protein production.

Chapter 4: Concluding discussion and future perspectives

4.1 Insights into *Plasmodium* biology

Malaria is a problematic and complicated disease requiring an array of prospects to eradicate the disease and challenges faced (Wells and Poll 2010). The current focus of restraining malaria is based on vector control regimes, the development of an effective vaccine, accurate diagnosis of malaria infected patients and drug discovery (Tripathi *et al.* 2005). As such, novel and effective antimalarial medicines are required that go beyond standard treatments to breach transmission from mosquito to human and from human to human (Wells and Poll 2010).

The launch of systems biology approach is crucial for the discovery of novel antimalarial drugs and their targets (Birkholtz *et al.* 2006). Systems biology involves the dissection of data generated from the genome, proteome and metabolome of not only the *Plasmodium* parasite but also its' two hosts (Fraunholz 2005; Hall and Carlton 2005). The development of x-omic strategies in the field of malaria eradication is still at infancy, which explains why detail of parasite biology is still limited. The data generated extends from understanding the mechanism of action of antimalarial compounds, identifying druggable genes (families of genes/gene products susceptible as drug targets), allowing lead optimization, verifying drug targets, annotating hypothetical proteins and regulation of gene expression (Birkholtz *et al.* 2008b).

In 2002 the full genome sequence of *P. falciparum* (strain 3D7) was determined (Gardner *et al.* 2002). It was found that sixty percent of proteins in *Plasmodium* bears no sequence similarity to other known species but some are attractive as antimalarial drug targets. Many challenges arise in antiparasitic drug discovery. Probable targets should be validated (chemically or genetically), structurally and biochemically examined and accepted for high-throughput screening (Birkholtz *et al.* 2008b). A tactic used to discover antimalarial agents involved identifying compounds that control the target's (e.g.

enzyme) activity *in vivo* (Plouffe *et al.* 2008). Previously, many compounds discovered as active in some antimalarial screens, have low activity in other screens. Therefore, the aim was to build up the infrastructure, knowledge and high throughput screening assays for antimalarial drugs that would eventually allow different compounds to inhibit various stages of the life cycle (stage-specific) and pathways (fatty acid biosynthesis, folate biosynthesis) based on drug-resistant strains and their phenotypic sensitivities (Plouffe *et al.* 2008). X-omics research applied in the field of antimalarial drug discovery is promising however, a limiting factor hindering the advancement of these technologies is the lack of available funding (Hayes *et al.* 2006).

4.2 Targeting folate metabolism

The primary role of folate metabolism in the *Plasmodium* parasite is to supply reduced 5,10-methylenetetrahydrofolate for the thymidylate cycle and thus DNA replication. Even though the parasite is capable of salvaging host folates and hence can bypass *de novo* folate synthesis, the use of well-known sulfa drugs (e.g. sulfonamides) to inhibit DHPS (the fourth enzyme in the biosynthetic pathway) has established that other enzymes of the folate pathway can be explored as potential targets (Hyde 2007). Folate metabolism has proven to be a good target in antimalarial drug discovery since it offers many opportunities for selective inhibition of vital processes in the malaria parasite (Hyde *et al.* 2008).

Recently, studies have reported that the enzyme catalyzing the second reaction (the conversion of 7,8-dihydroneopterin triphosphate to 6- hydroxymethyl-7,8- dihydropterin) of *de novo* folate synthesis is mediated by PTPS and not DHNA since the gene was missing from the *Plasmodium* genome (Dittrich *et al.* 2008; Hyde *et al.* 2008). PTPS is usually involved in tetrahydrobiopterin synthesis in other organisms (e.g. humans) unrelated to other apicomplexans indicating that convergent evolution has occurred. However, the *P. falciparum* PTPS differs (glu instead of cys) in the active site region as well as other structural differences in comparison to its human counterpart and could therefore provide an opportunity for differential inhibition (Hyde *et al.* 2008; Pribat *et al.* 2009).

HPPK catalyzes the third reaction in folate biosynthesis and with DHPS forms a bifunctional protein. Using the structure of *Saccharomyces cerevisiae* HPPK-DHPS, the

P. falciparum structure was derived using homology modelling. Investigations revealed the active site of HPPK and DHPS was highly conserved among *P. falciparum* and certain species. Therefore, effective inhibitors can be designed based on the structure of HPPK-DHPS (de Beer *et al.* 2006).

Other targets of folate biosynthesis include the bifunctional enzyme, DHFS-FPGS. DHFS in conjunction with FPGS mediates the final step in the *de novo* pathway (the conversion of dihydropterate to DHF upon addition of a L-glutamate residue to dihydropteroate) (Nzila *et al.* 2005). Characterization of *P. falciparum* DHFS-FPGS was recently initiated when a cDNA fragment encoding the gene was cloned from the K1 *P.falciparum* isolate into the expression vector. Since the behavior of *Plasmodium* FPGS differs in comparison to humans, the binding of the anticancer drug, methotrexate was evaluated as a substrate for *Plasmodium* FPGS. Inhibition studies revealed the drug was ineffective at reducing FPGS and DHFS levels but had an effect on parasite growth due to the binding of the drug to DHFR. Novel antifolate drugs should be developed in the future since the bifunctional enzyme is a good drug candidate (Wang *et al.* 2010).

The thymidylate cycle can also be probed for antimalarial drug discovery since inhibition studies conducted on *P. falciparum* SHMT using folate analogs that were aimed at inhibiting DHFR or TS also inhibited SHMT (Maenpuen *et al.* 2009; Pang *et al.* 2009). Therefore, despite the high similarity at the primary sequence level to the human isoforms, differential inhibition could be possible (Franca *et al.* 2005).

At present the *P. falciparum* DHFR-TS is the only enzyme in this pathway for which the crystal structures have been inferred and recently studies of new DHFR inhibitors designed specifically based on data obtained from the crystal structures have emerged (Fogel *et al.* 2008).

4.3 Probing into *P. falciparum* GTPCHI

GTPCHI catalyzes the first reaction in the biosynthesis of folates and was initially detected in *P. falciparum* extracts some 25 years ago (Lee *et al.* 2001). In other organisms GTPCHI catalyses the rate limiting step in *de novo* folate synthesis, in *Dictyostelium discoidium* control of GTPCHI activity is associated with progression in development (Witter *et al.* 1996) and in *P. knowlesi* the specific activity of this enzyme

was observed to alter at different life stages (Krungkrai *et al.* 1985). The human host possesses GTPCHI activity (even though *de novo* folate synthesis is absent), on the contrary the requirement for GTPCHI is for the synthesis of reduced biopterin (cofactor required for aromatic amino-acid hydroxylation and nitric oxide synthesis) (Tayeh and Marletta 1989). The *P. falciparum* GTPCHI C-terminal domain is well conserved revealing a 39% identity to other homologues. The active site as such, consists of 3 loops, 14 amino acid residues (13 of the residues are conserved to other homologues) and 4 monomers interact with the GTP substrate (Ichinose *et al.* 1995). Additionally, residues on the outskirts (N-terminal domain) of the active site have been proposed to alter GTPCHI activity. These residues in *Plasmodium* are quite divergent from the host. Furthermore, *P. falciparum* possess an elongated N-terminal residue in comparison to other organisms that is composed particularly of lysine and asparagines residues. This domain is hypothesized to function in protein-protein interactions for regulatory activity (Witter *et al.* 1996) and is found on the surface of the decameric holoenzyme. Since this N-terminal domain diverges from other homologues and plays a role in regulation, inhibitors can therefore be designed to target the divergent N- terminal and it may be possible to selectivity inhibit the *Plasmodium* GTPCHI (Ichinose *et al.* 1995; Lee *et al.* 2001).

Efforts to determine the structure of GTPCHI and other studies are hindered due to the difficulties in acquiring adequate quantities of soluble, active *P. falciparum* via heterologous expression. Recently studies on GTPCHI have been revived since the *gtpch*I gene was detected to amplify almost 11-fold in genomes of *Plasmodium* strains resistant to antifolate due to mutations in *dhps* and *dhfr* genes. This could be a mechanism for which the parasite compensates for the reduced flux of folate intermediates (Kidgell *et al.* 2006; Nair *et al.* 2008).

4.4 Expression of soluble GTPCHI

Possible antimalarial targets need to be validated for downstream examination (Birkholtz *et al.* 2008a). However, this proves to be challenging because of insufficient quantities of putative *Plasmodium* drug targets when using a heterologous expression system. The *Plasmodium* genome is A+T rich (approximately 76%) this results in defective expression or expression of the protein as insoluble inclusion bodies because proteins fold incorrectly (as is the case with codon bias) (Aguiar *et al.* 2004). Recombinant protein

expression assumes that the product formed is equivalent to the wild type protein. As such, there are various strategies available to improve soluble protein expression.

The simplicity and rapidity of the T7 expression system makes *E. coli* the host of choice for almost all structural genomics groups. This is because high-throughput cloning and screening produces expression of many target proteins with ease (Mehlin *et al.* 2006). *E.coli* was therefore considered as the preliminary host for *P. falciparum* GTPCHI expression studies. The nGTPCHI gene sequence was analyzed and manipulated to overcome the problem associated with *Plasmodium* codon bias using codon optimization and codon harmonization. Codon optimization entails substituting codons to match the codon preference in the expression host (*E. coli*). Whereas, codon harmonization involves substituting synonymous codons from *E. coli* that have codon usage frequencies that is similar or equal to those in *P. falciparum*. Codon harmonization would therefore ensure that the translational machinery is equivalent to that of *P. falciparum* since the low usage codons in *P. falciparum gtpch*I are altered to the low usage codons used by the heterologous expression host (*E. coli*). Harmonization also takes into account the presence of predicted pause sites, which were visible at distinct regions throughout the gene. This ensures that ribosomes pause and allow for correct folding of the target protein (Angov and Lyon 2006).

The codon optimized gene was obtained from Prof Hyde (University of Manchester, UK) thus synthesis was not required. However, the codon harmonized *gtpch*I gene was analyzed before synthesis to ensure no adverse RNA structures were present to alter the rate of translation (by ribosomes) and hence the folding of the target protein (Thanaraj and Argos 1996). As such, a set of repeat of AAA close to the translation initiation region was altered since hairpin RNA structures positioned close to the 5' end reduces the initiation of translation by obstructing ribosome loading. By weakening the hairpin structure, (by altering codons to lower the ΔG) translation efficiency would improve (Kozak 1989). Once ribosomes pass the initiation sites, they are then capable of disrupting base-pairing in the coding region of the gene and translation continues (Paulus *et al.* 2004).

Chapter 2 focussed particularly on a comparative expression study, conducted using the native gene (amplified from genomic DNA), codon optimized and codon harmonized

versions of the gene. These studies therefore, entailed expression in two different *E. coli* cell lines (BL21 Star (DE3) and M15pREP4), using two different media compositions (2X YT and auto-inducing media) and co-expression with a *Plasmodium* molecular chaperone (*Pf*hsp70). Codon harmonization was observed to slightly increase the total, insoluble, soluble as well as purified protein fractions of hGTPCHI in comparison to oGTPCHI and nGTPCHI using different media conditions (auto-induction media and YT-media) and with the co-expression of *Pf*hsp70. This indicates that codon harmonization improved expression of the target protein but not the levels of soluble proteins in comparison to the insoluble proteins.

Previously studies demonstrated that the co-transformation of bacterial cells with recombinant proteins and chaperones from the same species as the target protein improved target protein expression (de Marco and De Marco 2004). As such, the target proteins were co-expressed with *Pf*hsp70 but the presence of the *Plasmodium* chaperone did not increase the levels of purified proteins (hGTPCHI and oGTPCHI) as confirmed by western analysis. A second band below the 48 kDa hGTPCHI and oGTPCHI bands were identified on the western blot. These bands could possibly be an N-terminal deletion product since upon analysis of DNA sequences (h*gtpch*I and o*gtpch*I) there are two extra start codons (AUG) position further down from the original start codon that might have influenced the initiation of translation. Additionally, the band could also possibly correspond to an *E. coli* host protein since purification was not obtained to homogeneity. However, mass spectrometry would be required to prove these observations. This strategy revealed that the presence of molecular chaperones did not improve the expression of soluble proteins as suggested by Foth *et al* (2003) and Ramya *et al* (2007).

As expected the nGTPCHI protein did not express in BL21 Star (DE3) cells or with the co-expression of *P. falciparum* molecular chaperone (*Pf*hsp70) due to codon bias since the translational system of *E. coli* stalls when rare codons in *P. falciparum* genes are encountered (Baca and Hol 2000). In comparison to the other expression studies where expression was induced for a maximum of 4 hrs with IPTG, for bacterial cultures grown in auto-induction media, protein expression was followed for 48 hrs because cultures were grown to saturation to produce greater levels of target protein per volume culture (Studier 2005). Although, the auto-induction media was convenient, It was clear that the

change in media composition did not improve the levels of soluble proteins. Berrow *et al*. (2006) stated that both IPTG (YT-medium) and auto-induction produce qualitatively similar results however, the use of auto-induction may produce greater levels of protein expression. Auto-induction is a simpler method for scale-up but the greater biomass obtained may hinder sample processing (Berrow *et al.* 2006).

Overexpression of genes is essential for the production of adequate quantity of recombinant protein used for protein structure and function studies. Therefore, instead of using the model host *E. coli*, the cell-free wheat germ expression system was used to reveal whether expression of the target protein could improve. Codon harmonization or optimization would not alter the outcome of expression since the wheat germ system was observed to have little preference for codon usage and was able to translate A+T rich sequences (Endo and Sawasaki 2002). Additionally, the Endext technology wheat germ expression (Cell free Sciences, Japan) system boasted the expression of soluble proteins and was successful at the expression of functional *P. falciparum* DHFR-TS (Mudeppa *et al.* 2007), several malaria vaccine candidates (such as *Pfs*25, *Pf*CSP and *Pf*AMA-1), certain *P. falciparum* blood stage proteins (Tsuboi *et al.* 2008) and a *P. falciparum* protein kinase 2 (PfPK2) (Kato *et al.* 2008).

However, when comparing the expression of nGTPCHI, oGTPCHI and hGTPCHI using the wheat germ system results indicated no expression of each of the proteins since the protein band corresponding to GTPCHI (48 kDa) could not be detected via western analysis and a comparison of expression to the negative control revealed similar results. This discrepancy to the expected results initiated the proposal of several hypotheses based upon previous successful investigations where the wheat germ extract plus (Promega, USA) was used for protein synthesis and where *Plasmodium* proteins were synthesized. Therefore, either *Plasmodium* GTPCHI fails to express using the wheat germ expression system or experimental conditions need optimization.

The expression of various different enzymes using the wheat germ extract plus system (Promega, USA) was performed (http://www.promega.com/pnotes/90.pdf; Porter *et al.* 2007). However, in comparison to *gtpch*I mRNA templates a difference noted was that the mRNA templates possessed a BYDV sequence that enhanced the protein synthesis more than 3-fold than templates without the enhancer sequence

(http://www.promega.com/pnotes/90.pdf). Furthermore, high levels of *Plasmodium* protein expression studies were performed using Endext technology wheat germ expression premium kit (Cell free Sciences, Japan). For this a pEU plasmid incorporating a 5' enhancer element that also allowed for the production of proteins in milligram quantities and the plasmid does not require linearization before transcription (Takai *et al.* 2008). An option of cloning the protein-coding regions into the Wheat Germ Flexi® Vectors, which would also incorporate these 5´ and 3´ elements into the mRNA could be useful for GTPCHI synthesis.

Even though the purity of mRNA before translation was high, a factor that may have resulted in no expression is the concentration of template mRNA used. As such, a variety of template concentrations needs to be assessed. It has also been observed that in the wheat cell-free system, the template mRNA is rapidly degraded during translation (Takai *et al.* 2008), which may explain why protein synthesis did not occur. As such, expression can be conducted using the bilayer translation reaction, which is based on a continuous supply of substrates and the diffusion of by-products (Sawasaki *et al.* 2002). This would ensure a continuous supply of mRNA template (substrate) is available, so the degraded mRNA is replenished and translation can continue for longer periods of time.

These results reveal that no universal method for heterologous expression can be applied and each protein has to be evaluated individually (Birkholtz *et al.* 2008). The aim of the thesis was to express *Plasmodium falciparum* GTPCHI. A comparative analysis of nGTPCHI, oGTPCHI and hGTPCHI was evaluated using different expression systems and conditions. Results indicated levels of soluble protein were obtained using in *E. coli* cells however, purification was not to homogeneity.

4.5 Future perspectives

4.5.1 Improving the yield of soluble proteins

It has been established that the majority of proteins express in an insoluble form, which are sequestered in inclusion bodies (Carrio and Villaverde 2002). Based on western analysis that was conducted on the total, soluble, insoluble and purified GTPCHI (results not shown) expressed in *E. coli*, a considerable level (greater than the soluble fraction) of GTPCHI expression was observed in the insoluble fractions. There are various other

techniques to improving the expression of soluble GTPCHI. This includes protein refolding strategies, the use of solubility-enhancing tags and gentle lysis of host cells. These tags (maltose-binding protein and nutilisation substance A) support the solubility of their fusion partners (Waugh 2005). Since protein aggregation in bacteria is reversible, insoluble protein fractions (inclusion bodies) can be isolated and refolded to the correct conformation (Villaverde and Carrio 2003). However, isolation is a delicate process as unwanted contaminates and membranes have been observed to co-precipitate with the inclusion bodies. Conditions for solubilization and disulphide bond disruption involves treatment with high concentration of chaotropic reagents and anionic detergents and metal ions chelated to inhibit oxidation of cysteine residues. This is then followed by dialysis, filtration or dilutions to promote refolding of the soluble protein and to eliminate the reagents that were added (Vallejo and Rinas 2004). Furthermore, the methods that were utilized for the disruption of bacterial cells were harsh and this could have decreased the yield of soluble proteins. The use of EDTA should have been included to release proteins from the bacterial cells before the addition of lysozyme since poor cell lysis would result in protein isolating in the insoluble fractions. This entails a 30 min incubation of bacterial cells with EDTA at 37 ºC prior to adding lysozyme. As such, the EmulsiXex-C3 (Avestin, Inc.) or the MicroXuidizer (Micro- Xuidics) which, allows gentle cell lysis of a range of culture volumes and does not require the addition of lysozyme or DNase, (to assist lysis) could be considered instead. However, the main disadvantage of these instruments is the high cost involved (Peti and Page 2007).

4.5.2 Downstream use of GTPCHI

For verification of the protein obtained, mass spectrometry could be conducted on the protein bands obtained for GTPCHI. The target protein was detected via western analysis however, based on the his tag attached to the C-terminal of the protein. Co-expression of the target protein with *Pf*hsp70 did not give better levels of expression as expected. Therefore, western analysis using an antibodies for *Pf*hsp70 and GTPCHI could be used to detect the expression of *Pf*hsp70 and GTPCHI, respectively. Furthermore, purification of GTPCHI could be improved by the addition of glycerol or a detergent (e.g. Tween-20) since this would eliminate nonspecific interactions that associate with GTPCHI and also by increasing the concentration of imidazole in the wash buffer contaminating proteins should be removed during the washing of the his-column before elution of GTPCHI.

Once purified GTPCHI is obtained, kinetic activity assays would be conducted to determine the activity of GTPCHI. Kinetic activity assays for GTPCHI is based on observing the formation of dihydroneopterin triphosphate (the product of GTPCHI). From the starting product, GTP neopterin 3'-triphosphate (NTP) is produced (after iodine oxidation of dihydroneopterin triphosphate) (Blau and Niederwieser 1983). Thereafter, the active, correctly folded protein obtained in sufficient quantities will be utilized for NMR investigation and X-ray crystallography to obtain the three-dimensional structure. This structure will permit the design of novel antimalarial inhibitors to evolve what is now prevalent as antimalarial drug resistance and open avenues for new drug regimes (Hayes *et al.* 2006).

Appendix A

The following sequence alignment differentiates between codons of the n*gtpch*I gene sequence, the o*gtpch*I gene sequence and the h*gtpch*I gene sequence. The codon optimized gene sequence was altered to an *E. coli* codon preference and the codon harmonized gene sequence was adapted to the *P. falciparum* codon usage frequency (Refer to section 2.1.2.2). Note the amino acid sequence remains identical for each gene sequence with the change of codons.

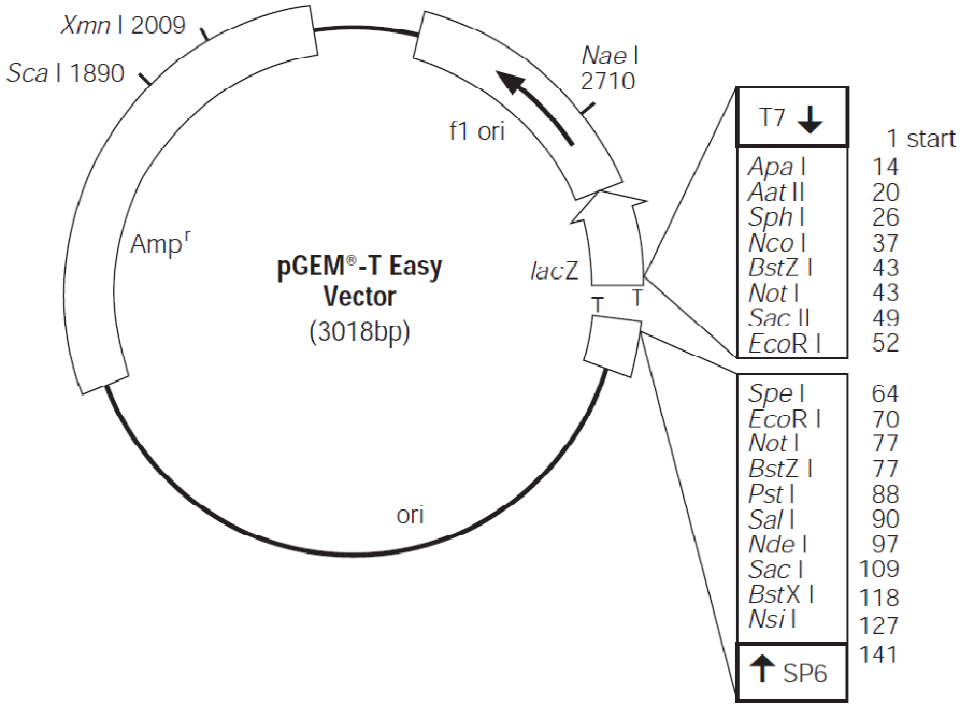
```
                  10        20        30        40        50        60
          ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI   atgtataaatatacgtcaataaacaaatctgataaaatatatgaaacacataatatggaa
ogtpchI   atgtacaaatacacctctatcaacaagtccgacaagatctacgaaacccacaacatggaa
hgtpchI   aTGTACAAATACACGTCTATCAATAAAAGCGACAAAATCTACGAAACCCACAACATGGAA
amino acid  ~~M~~Y~~K~~Y~~T~~S~~I~~N~~K~~S~~D~~K~~I~~Y~~E~~T~~H~~N~~M~~E

                   70        80        90       100       110       120
          ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI   gaaaaaaaaaaaaaaggtaataataacaatttctcaggattattaaataatgaaatcgat
ogtpchI   gagaagaagaaaaagggtaacaacaacaacttctctagcctgctgaacaacgaaatcgac
hgtpchI   GAAAAAAagAAAAAAGGCAACAACAATAACTTTTCTGGCCTGCTGAACAACGAAATAGAC
amino acid  ~~E~~K~~K~~K~~K~~G~~N~~N~~N~~N~~F~~S~~G~~L~~L~~N~~N~~E~~I~~D

                  130       140       150       160       170       180
          ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI   gataataataaaaggaaaagttaaaaaatagtatatctaaaatgtatagtaaccataag
ogtpchI   gataataacaagaaagaaaaactgaaaaacagcatctctaaaatgtacagcaaccacaag
hgtpchI   GACAACAACAAAAAGGAAAAGCTGAAAAACTCTATCAGCAAAATGTACTCTAATCACAAG
amino acid  ~~D~~N~~N~~K~~K~~E~~K~~L~~K~~N~~S~~I~~S~~K~~M~~Y~~S~~N~~H~~K

                  190       200       210       220       230       240
          ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI   aatagagaaaattttaacgaatgtgagaaggaagaccttgttgtgatagatgaaaaggat
ogtpchI   aaccgcgaaaacttcaacgaatgtgaaaaagaagacctggttgtaatcgacgaaaaagac
hgtpchI   AACCGTGAAAACTTCAATGAATGCGAGAAGGAAGATCTCGTTGTCATCGACGAAAAGGAC
amino acid  ~~N~~R~~E~~N~~F~~N~~E~~C~~E~~K~~E~~D~~L~~V~~V~~I~~D~~E~~K~~D

                  250       260       270       280       290       300
          ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI   aataataaaaagaaaaaaaaaaatatgacaaatacatttgaacaagataataattataat
ogtpchI   aacaacaagaagaagaagaagaacatgaccaacaccttcgaacaggataacaactacaac
hgtpchI   AACAACAAAAAGAAAAAAAAAAACATGACCAACACCTTCGAACAGGACAACAACTACAAC
amino acid  ~~N~~N~~K~~K~~K~~K~~N~~M~~T~~N~~T~~F~~E~~Q~~D~~N~~N~~Y~~N

                  310       320       330       340       350       360
          ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI   atgaatgataataaaaggttaggtagctttttaaaataaatgataaatgtgaatctatt
ogtpchI   atgaacgacaacaaacgtctgggttccttcttcaaaatcaacgacaagtgtgaatcgatc
hgtpchI   ATGAACGACAACAAACGCCTGGGCTCGTTCTTCAAAATCAACGACAAATGCGAAAGCATT
amino acid  ~~M~~N~~D~~N~~K~~R~~L~~G~~S~~F~~F~~K~~I~~N~~D~~K~~C~~E~~S~~I
```

```
                370       380       390       400       410       420
           ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI    aatgagaatgttaataatattaataaacagtctcttaaggattctatactatttgacaat
ogtpchI    aacgaaaacgtgagcaacatcaacaaacagtctctgaaagactccatcctgttcgacaac
hgtpchI    AACGAGAACGTTAACAACATTAACAAACAAAGCCTCAAGGACAGCATCCTCTTCGATAAC
amino acid ~~N~~E~~N~~V~~N~~N~~I~~N~~K~~Q~~S~~L~~K~~D~~S~~I~~L~~F~~D~~N

                430       440       450       460       470       480
           ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI    ataaatgagaaagaatattttaacgaaaccaaagaagaaaataaggaaggaaataaaagt
ogtpchI    atcaacgaaaaggaatacttcaacgaaaccaaagaagaaaacaaggaaggtaacaagagc
hgtpchI    ATCAACGAGAAAGAATACTTCAATGAAACGAAAGAAGAAAACAAGGAAGGCAACAAATCT
amino acid ~~I~~N~~E~~K~~E~~Y~~F~~N~~E~~T~~K~~E~~E~~N~~K~~E~~G~~N~~K~~S

                490       500       510       520       530       540
           ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI    aatgatatagaaaagataaattgtatgaaagtgaagaagaaaactgttaaaaagaataaa
ogtpchI    aacgatatcgaaaagatcaactgcatgaaagttaagaaaaagaccgttaagaaaaacaag
hgtpchI    AACGACATCGAAAAGATCAACTGCATGAAAGTCAAGAAGAAAACTGTTAAAAAGAACAAA
amino acid ~~N~~D~~I~~E~~K~~I~~N~~C~~M~~K~~V~~K~~K~~K~~T~~V~~K~~K~~N~~K

                550       560       570       580       590       600
           ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI    aagaagattaataaaattataaacaataagaataaaatatctaaatcaaatgacatagaa
ogtpchI    aagaagatcaacaagatcatcaacaacaagaacaaaatcagcaagtctaacgatatcgaa
hgtpchI    AAGAAGATTAACAAAATTATCAATAACAAGAACAAAATCAGCAAATCTAACGATATCGAA
amino acid ~~K~~K~~I~~N~~K~~I~~I~~N~~N~~K~~N~~K~~I~~S~~K~~S~~N~~D~~I~~E

                610       620       630       640       650       660
           ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI    gaacaaataattaatattagtaaacatatatataaaatattaaatatatccaaattacca
ogtpchI    gaacagatcatcaacatctccaaacacatctacaagatcctgaacatctctaagcttccg
hgtpchI    GAACAGATCATTAACATTTCTAAACACATCTACAAAATCCTGAACATCTCGAAACTGCCG
amino acid ~~E~~Q~~I~~I~~N~~I~~S~~K~~H~~I~~Y~~K~~I~~L~~N~~I~~S~~K~~L~~P

                670       680       690       700       710       720
           ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI    aaatgtgatatattaaaaagaacaaatagaagatatgctgaaacatttttatatttaact
ogtpchI    aagtgtgatatcctgaaacgtactaaccgtcgttacgcggaaaccttcctgtacctgact
hgtpchI    AAATGCGACATCCTGAAACGTACCAACCGTCGTTACGCTGAAACCTTCCTGTACCTGACT
amino acid ~~K~~C~~D~~I~~L~~K~~R~~T~~N~~R~~R~~Y~~A~~E~~T~~F~~L~~Y~~L~~T

                730       740       750       760       770       780
           ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI    aatggttataatctggatatagaacaaataataaaaagatctttatataaaaggatgtat
ogtpchI    aacggttacaacctggacatcgaacagattattaaacgttccctgtacaagcgtatgtac
hgtpchI    AACGGCTACAACTTAGACATCGAACAGATCATCAAACGTAGCCTGTACAAACGCATGTAC
amino acid ~~N~~G~~Y~~N~~L~~D~~I~~E~~Q~~I~~I~~K~~R~~S~~L~~Y~~K~~R~~M~~Y

                790       800       810       820       830       840
           ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI    aaaaataattcaataatcaaagttacaggtatacatatatattcattatgtaaacaccat
ogtpchI    aagaacaactctatcatcaaagtgaccggtatccacatctactccctgtgtaaacaccac
hgtpchI    AAAAACAACTCTATCATAAAAGTTACCGGCATCCACATCTACTCTCTGTGCAAACATCAC
amino acid ~~K~~N~~N~~S~~I~~I~~K~~V~~T~~G~~I~~H~~I~~Y~~S~~L~~C~~K~~H~~H
```

```
                  850       860       870       880       890       900
           ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI    cttttacctttgaaggtacatgtgatattgagtatatacccaataaatatattatcggg
ogtpchI    ctgcttccattcgaaggtacctgcgatatcgaatacatcccgaacaaatacatcatcggt
hgtpchI    CTCCTGCCATTCGAAGGCACCTGCGACATTGAGTACATCCCTAACAAATACATTATAGGG
amino acid ~~L~~L~~P~~F~~E~~G~~T~~C~~D~~I~~E~~Y~~I~~P~~N~~K~~Y~~I~~I~~G

                  910       920       930       940       950       960
           ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI    ttatctaaattttcaagaatagttgatgtcttttctagaagattacaattacaagaagat
ogtpchI    ctgtctaagttctctcgtatcgtggatgtattctctcgtcgtctgcaactgcaagaagat
hgtpchI    CTGAGCAAATTCTCTCGTATCGTTGACGTCTTCAGCCGTCGTCTGCAGCTGCAGGAAGAC
amino acid ~~L~~S~~K~~F~~S~~R~~I~~V~~D~~V~~F~~S~~R~~R~~L~~Q~~L~~Q~~E~~D

                  970       980       990       1000      1010      1020
           ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI    ttaactaacgatatttgtaacgctttaaaaaaatacttaaaaccattatatattaaagta
ogtpchI    ctgaccaacgacatctgcaacgcgctgaaaaaatacctgaagccgctgtacatcaaggtt
hgtpchI    CTGACTAATGACATTTGCAATGCTCTGAAAAAATATCTGAAACCGCTGTACATTAAAGTT
amino acid ~~L~~T~~N~~D~~I~~C~~N~~A~~L~~K~~K~~Y~~L~~K~~P~~L~~Y~~I~~K~~V

                  1030      1040      1050      1060      1070      1080
           ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI    tctattgtagctaaacatttatgtataaatatgaggggagttaaagagcacgatgctaaa
ogtpchI    tccatcgttgctaagcacctgtgcatcaacatgcgtggtgtgaaggaacacgacgcgaag
hgtpchI    AGCATTGTTGCTAAACACCTGTGCATCAACATGCGCGGCGTTAAAGAGCATGACGCTAAA
amino acid ~~S~~I~~V~~A~~K~~H~~L~~C~~I~~N~~M~~R~~G~~V~~K~~E~~H~~D~~A~~K

                  1090      1100      1110      1120      1130      1140
           ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI    actataacgtatgcatcttataaagcagaaaaagagaatcctacagttcattctttaaat
ogtpchI    actatcacttacgcttcctacaaagctgaaaaagaaaacccgactgtacactccctgaac
hgtpchI    ACTATCACGTACGCTAGCTACAAAGCTGAAAAAGAGAACCCAACCGTTCACAGCCTGAAC
amino acid ~~T~~I~~T~~Y~~A~~S~~Y~~K~~A~~E~~K~~E~~N~~P~~T~~V~~H~~S~~L~~N

                  1150      1160      1170
           ....|....|....|....|....|....|
ngtpchI    attgactctttctgtggaaaatttaaattag
ogtpchI    atcgactcctccgttgaaaacctgaactaa
hgtpchI    ATTGATAGCAGCGTCGAAAACCTGAACTGA
amino acid ~~I~~D~~S~~S~~V~~E~~N~~L~~N~~*
```

Appendix B

pGEM®- T Easy vector map



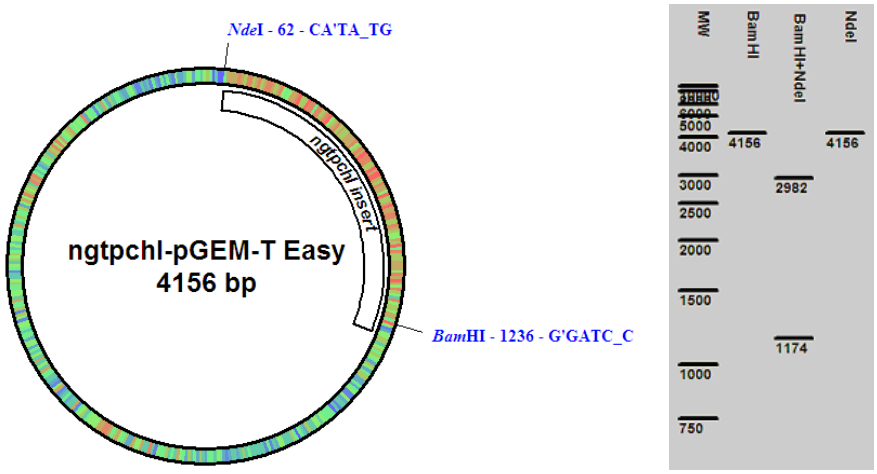
pGEM®-T Easy Vector Sequence reference points:	
T7 RNA Polymerase transcription initiation site	1
SP6 RNA Polymerase transcription initiation site	141
T7 RNA Polymerase promoter	3002-6
SP6 RNA Polymerase promoter	136-158
multiple cloning site	10-128
lacZ start codon	180
lac operon sequences	2839-2999, 166-395
lac operator	100-216
β-lactamase coding region	1337-2197
phage f1 region	2383-2838
binding site of pUC/M13 Forward Sequencing Primer	2959-2975
binding site of pUC/M13 Reverse Sequencing Primer	176-192

The multiple cloning site which, is positioned within the lacZ gene is the A/T cloning site (blue arrow). The red arrow indicates the ampicillin resistance gene encoded by the β-lactamse coding region.

Appendix B.1

n*gtpch*I- pGEM®-T Easy vector map

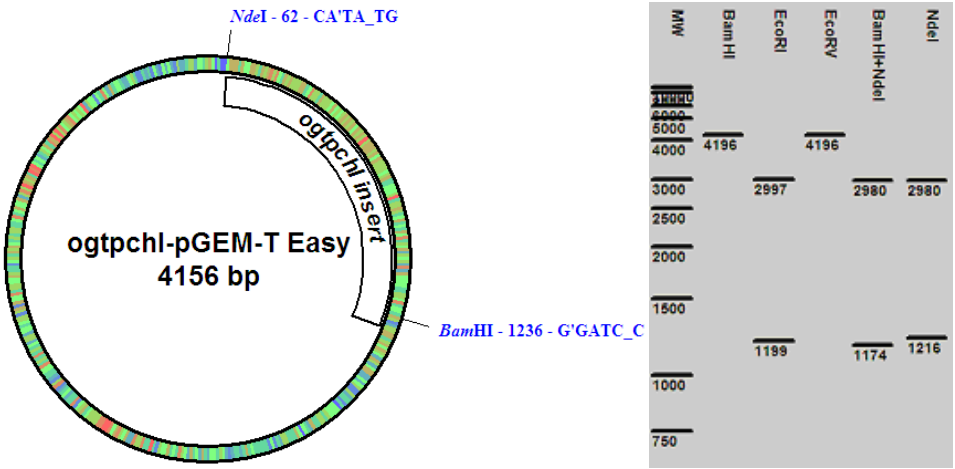
The plasmid vector map and predicted bands after restriction enzyme digestion for the release of the n*gtpch*I insert from pGEM®-T Easy vector. The restriction enzyme (*BamH*I-*Nde*I) sites flanking the insert gene are indicated in blue.



Appendix B.2

o*gtpch*I- pGEM®-T Easy vector map

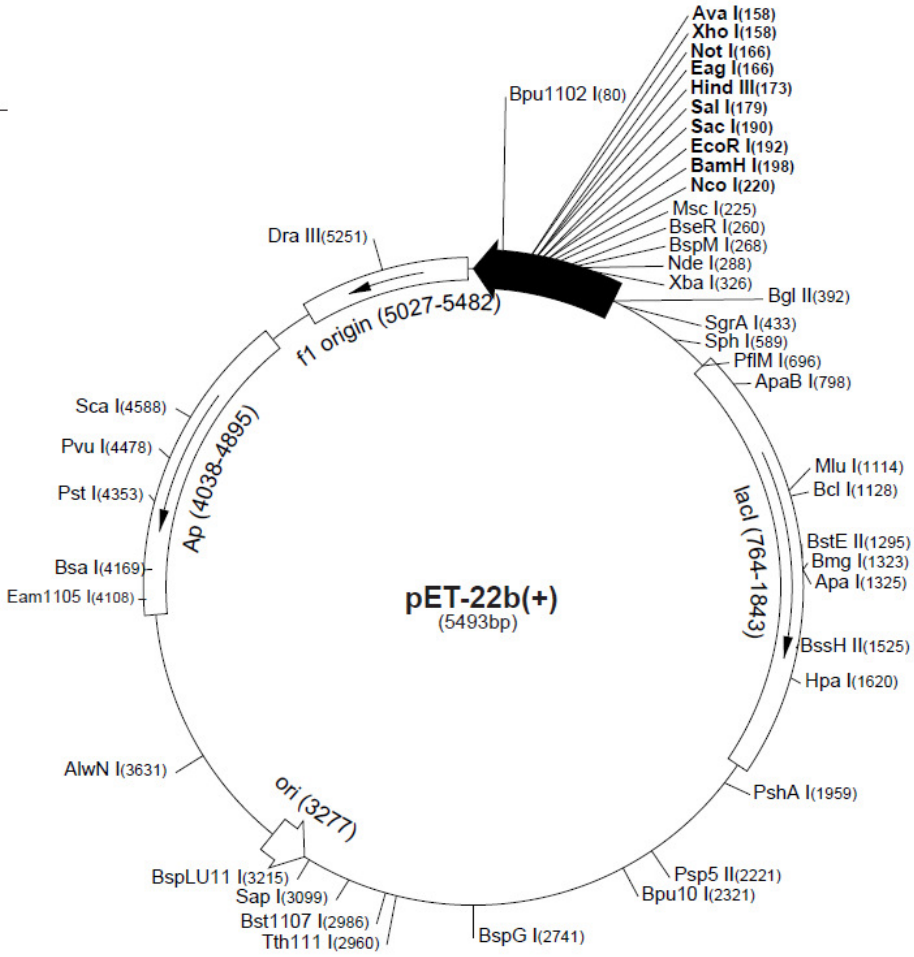
The plasmid vector map and predicted bands after restriction enzyme digestion for the release of the o*gtpch*I insert from pGEM®-T Easy vector. The restriction enzyme sites flanking the insert gene are indicated in blue (*BamH*I-*Nde*I).



Appendix C

pET-22b vector map

pET-22b(+) sequence landmarks	
T7 promoter	361-377
T7 transcription start	360
pelB coding sequence	224-289
Multiple cloning sites	
(*Nco* I - *Xho* I)	158-225
His•Tag coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	764-1843
pBR322 origin	3277
bla coding sequence	4038-4895
f1 origin	5027-5482

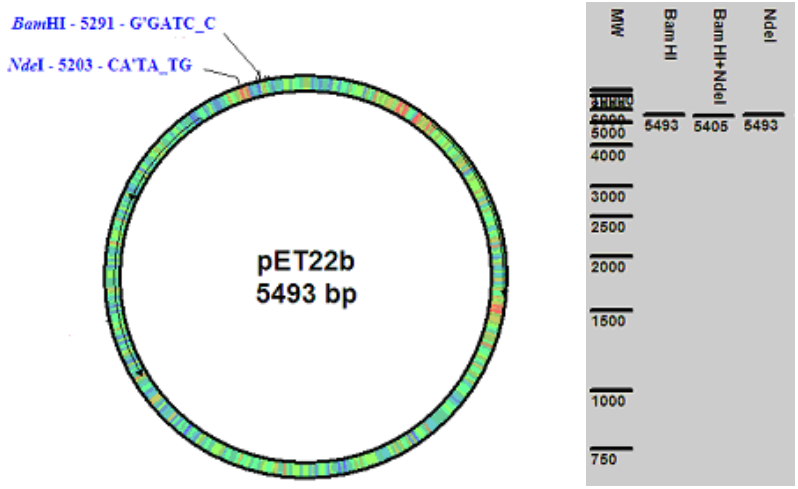


The T7 promoter and T7 terminator sites indicate vector primer positions. The pET-22b vector was utilized for His-tag protein expression. The position of the His-tag coding region is indicated (pink arrow). The ampicillin resistance gene is also indicated (red arrow).

Appendix C.1

Restriction enzyme digestion of pET22b vector

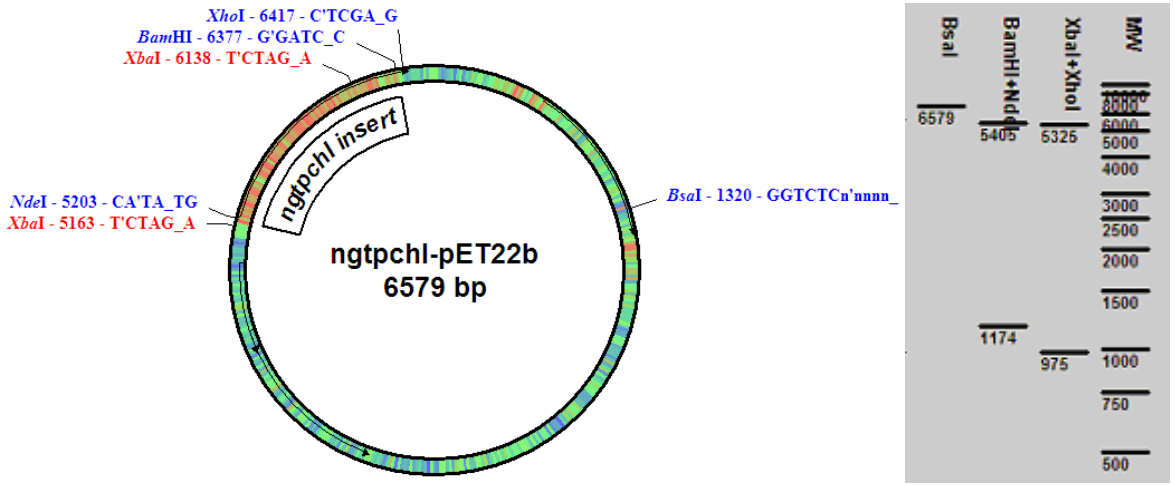
The plasmid vector map and predicted bands after restriction enzyme digestion for ligation of the pET22b vector with the n- o- h-*gtpch*I insert. The restriction enzyme sites for vector cleavage are indicated in blue (*BamH*I-*Nde*I).



Appendix C.2

n*gtpch*I-pET22b vector

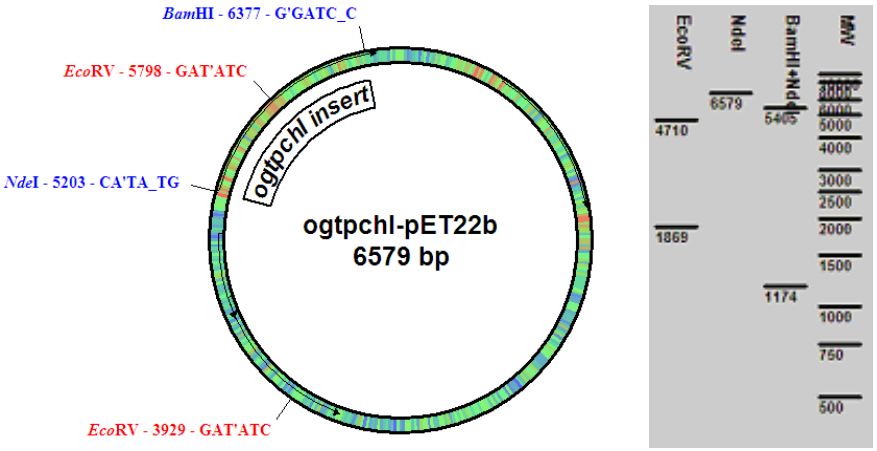
The plasmid vector map and predicted bands after restriction enzyme digestion for the release of the n*gtpch*I insert from pET22b vector. The restriction enzyme sites flanking the insert gene are indicated in blue (*BamH*I-*Nde*I).



Appendix C.3

o*gtpch*I-pET22b vector

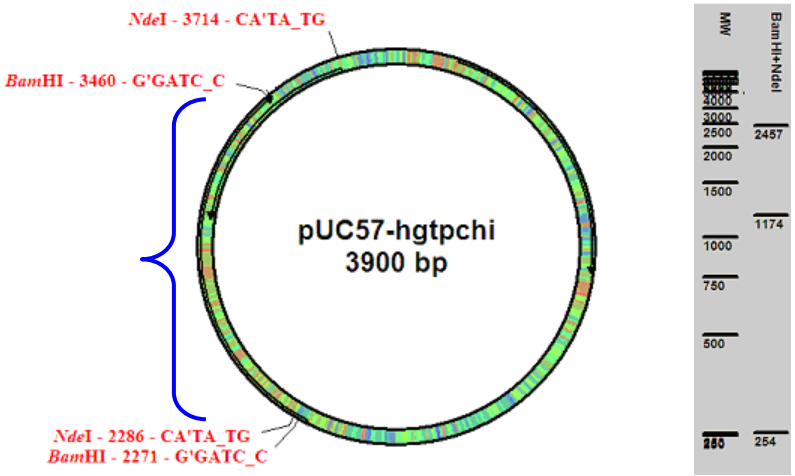
The plasmid vector map and predicted bands after restriction enzyme digestion for the release of the o*gtpch*I insert from pET22b vector. The restriction enzyme sites flanking the insert gene are indicated in blue (*BamH*I-*Nde*I).



Appendix C.4

h*gtpch*I- pUC57 vector

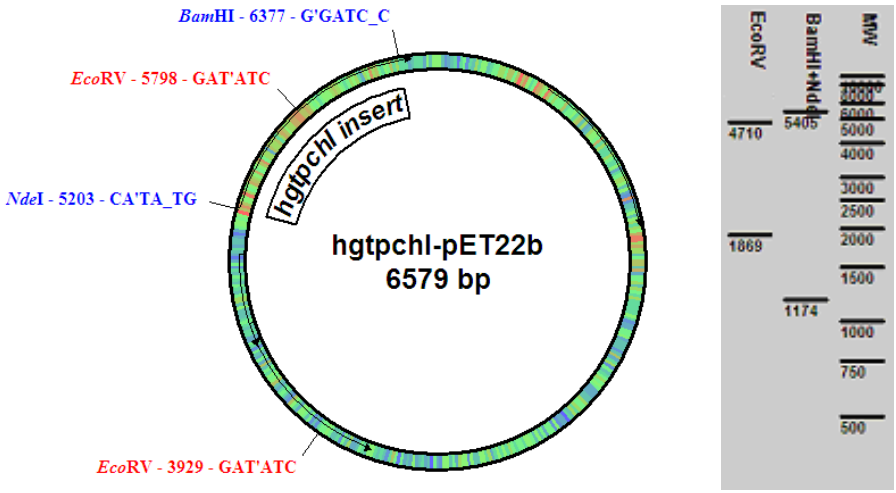
The plasmid vector map and predicted bands after restriction enzyme digestion for the release of the h*gtpch*I insert from pUC57 vector. The restriction enzyme sites flanking the insert gene (indicated as a blue, left brace) are indicated in red (*BamH*I-*Nde*I).Appendix C.4 *BamH*I and *Nde*I restriction vector map of insert.



Appendix C.5

h*gtpch*I-pET22b vector

The plasmid vector map and predicted bands after restriction enzyme digestion for the release of the h*gtpch*I insert from pET22b vector. The restriction enzyme sites flanking the insert gene are indicated in blue (*BamH*I-*Nde*I).



Appendix D

The n*gtpch*I gene sequence and the alignment of sequences obtained from the nucleotide sequencing results (Refer to section 2.2.13) after sub-cloning n*gtpch*I into the pET22b expression vector.

```
                    10         20         30         40         50         60
               ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI        CATatgtataaatatacgtcaataaacAaatctgataaaatatatgaaacacataatatg
Forward pr     CATATGTATAAATATAGG~~AAAAAACAAATCTGATAAAATATATGAAACACATAATATG
Reverse pr     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Amino acid     H~~M~~Y~~K~~Y~~T~~S~~I~~N~~K~~S~~D~~K~~I~~Y~~E~~T~~H~~N~~M~~

                    70         80         90        100        110        120
               ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI        gaagaaaaaaaaaaaaaggtaataataacaatttctcaggattattaaataatgaaatc
Forward pr     GAAGAAAAAAAAAAAAAGGTAATAATAACAATTTCTCAGGATTATTAAATAATGAAATC
Reverse pr     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Amino acid     E~~E~~K~~K~~K~~K~~G~~N~~N~~N~~N~~F~~S~~G~~L~~L~~N~~N~~E~~I~~

                   130        140        150        160        170        180
               ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI        gatgataataataaaaggaaaagttaaaaaatagtatatctaaaatgtatagtaaccat
Forward pr     GATGATAATAATAAAAAGGAAAAGTTAAAAAATAGTATATCTAAAATGTATAGTAACCAT
Reverse pr     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Amino acid     D~~D~~N~~N~~K~~K~~E~~K~~L~~K~~N~~S~~I~~S~~K~~M~~Y~~S~~N~~H~~
```

```
                 190       200       210       220       230       240
            ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI     aagaatagagaaaattttaacgaatgtgagaaggaagaccttgttgtgatagatgaaaag
Forward pr  AAGAATAGAGAAAATTTTAACGAATGTGAGAAGGAAGACCTTGTTGTGATAGATGAAAAG
Reverse pr  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Amino acid  K~~N~~R~~E~~N~~F~~N~~E~~C~~E~~K~~E~~D~~L~~V~~V~~I~~D~~E~~K~~

                 250       260       270       280       290       300
            ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI     gataataataaaaagaaaaaaaaaaatatgacaaatacatttgaacaagataataattat
Forward pr  GATAATAATAAAAAGAAAAAAAAAA~TATGACAAATACATTTGAACAAGATAATAATTAT
Reverse pr  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Amino acid  D~~N~~N~~K~~K~~K~~K~~K~~N~~M~~T~~N~~T~~F~~E~~Q~~D~~N~~N~~Y~~

                 310       320       330       340       350       360
            ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI     aatatgaatgataataaaaggttaggtagctttttaaaataaatgataaatgtgaatct
Forward pr  AATATGAATGATAATAAAAGGTTAGGTAGCTTTTTTAAAATAAATGATAAATGTGAATCT
Reverse pr  AATATGAATGATAATAAAAGGTTAGGTAGCTTTTTTAAAATAAATGATAAATGTGAATCT
Amino acid  N~~M~~N~~D~~N~~K~~R~~L~~G~~S~~F~~F~~K~~I~~N~~D~~K~~C~~E~~S~~

                 370       380       390       400       410       420
            ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI     attaatgagaatgttaataatattaataaacagtctcttaaggattctatactatttgac
Forward pr  ATTAATGAGAATGTTAATAATATTAATAAACAGTCTCTTAAGGATTCTATACTATTTGAC
Reverse pr  ATTAATGAGAATGTTAATAATATTAATAAACAGTCTCTTAAGGATTCTATACTATTTGAC
Amino acid  I~~N~~E~~N~~V~~N~~N~~I~~N~~K~~Q~~S~~L~~K~~D~~S~~I~~L~~F~~D~~

                 430       440       450       460       470       480
            ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI     aatataaatgagaaagaatattttaacgaaaccaaagaagaaaataaggaaggaaataaa
Forward pr  AATATAAATGAGAAAGAATATTTTAACGAAACCAAAGAAGAAAATAAGGAAGGAAATAAA
Reverse pr  AATATAAATGAGAAAGAATATTTTAACGAAACCAAAGAAGAAAATAAGGAAGGAAATAAA
Amino acid  N~~I~~N~~E~~K~~E~~Y~~F~~N~~E~~T~~K~~E~~E~~N~~K~~E~~G~~N~~K~~

                 490       500       510       520       530       540
            ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI     agtaatgatatagaaaagataaattgtatgaaagtgaagaagaaaactgttaaaaagaat
Forward pr  AGTAATGATATAGAAAAGATAAATTGTATGAAAGTGAAGAAGAAAACTGTTAAAAAGAAT
Reverse pr  AGTAATGATATAGAAAAGATAAATTGTATGAAAGTGAAGAAGAAAACTGTTAAAAAGAAT
Amino acid  S~~N~~D~~I~~E~~K~~I~~N~~C~~M~~K~~V~~K~~K~~K~~T~~V~~K~~K~~N~~

                 550       560       570       580       590       600
            ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI     aaaaagaagattaataaaattataaacaataagaataaaatatctaaatcaaatgacata
Forward pr  AAAAAGAAGATTAATAAAATTATAAACAATAAGAATAAAATATCTAAATCAAATGACATA
Reverse pr  AAAAAGAAGATTAATAAAATTATAAACAATAAGAATAAAATATCTAAATCAAATGACATA
Amino acid  K~~K~~K~~I~~N~~K~~I~~I~~N~~N~~K~~N~~K~~I~~S~~K~~S~~N~~D~~I~~

                 610       620       630       640       650       660
            ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI     gaagaacaaataattaatattagtaaacatatatataaaatattaaatatatccaaatta
Forward pr  GAAGAACAAATAATTAATATTAGTAAACATATATATAAAATATTAAATATATCCAAATTA
Reverse pr  GAAGAACAAATAATTAATATTAGTAAACATATATATAAAATATTAAATATATCCAAATTA
Amino acid  E~~E~~Q~~I~~I~~N~~I~~S~~K~~H~~I~~Y~~K~~I~~L~~N~~I~~S~~K~~L~~
```

```
                  670        680        690        700        710        720
              ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI       ccaaaatgtgatatattaaaaagaacaaatagaagatatgctgaaacatttttatattta
Forward pr    CCAAAATGTGATATATTAAAAAGAACAAATAGAAGATATGCTGAAACATTTTTATATTTA
Reverse pr    CCAAAATGTGATATATTAAAAAGAACAAATAGAAGATATGCTGAAACATTTTTATATTTA
Amino acid    P~~K~~C~~D~~I~~L~~K~~R~~T~~N~~R~~R~~Y~~A~~E~~T~~F~~L~~Y~~L~~

                  730        740        750        760        770        780
              ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI       actaatggttataatctggatatagaacaaataataaaaagatctttatataaaaggatg
Forward pr    ACTAATGGTTATAATCTGGATATAGAACAAATAATAAAAAGATCTTTATATAAAAGGATG
Reverse pr    ACTAATGGTTATAATCTGGATATAGAACAAATAATAAAAAGATCTTTATATAAAAGGATG
Amino acid    T~~N~~G~~Y~~N~~L~~D~~I~~E~~Q~~I~~I~~K~~R~~S~~L~~Y~~K~~R~~M~~

                  790        800        810        820        830        840
              ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI       tataaaaataattcaataatcaaagttacaggtatacatatatattcattatgtaaacac
Forward pr    TATAAAAATAATTCAATAATCAAAGTTACAGGTATACATATATATTCATTATGTAAACAC
Reverse pr    TATAAAAATAATTCAATAATCAAAGTTACAGGTATACATATATATTCATTATGTAAACAC
Amino acid    Y~~K~~N~~N~~S~~I~~I~~K~~V~~T~~G~~I~~H~~I~~Y~~S~~L~~C~~K~~H~~

                  850        860        870        880        890        900
              ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI       catcttttacctttgaaggtacatgtgatattgagtatatacccaataaatatattatc
Forward pr    CATCTTTTACCTTTTGAAGGTACATGTGATATTGAGTATATACCCAATAAATATATTATC
Reverse pr    CATCTTTTACCTTTTGAAGGTACATGTGATATTGAGTATATACCCAATAAATATATTATC
Amino acid    H~~L~~L~~P~~F~~E~~G~~T~~C~~D~~I~~E~~Y~~I~~P~~N~~K~~Y~~I~~I~~

                  910        920        930        940        950        960
              ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI       gggttatctaaattttcaagaatagttgatgtcttttctagaagattacaattacaagaa
Forward pr    GGGTTATCTAAATTTTCAAGAATAGTTGATGTCTTTTCTAGAAGATTACA~~TACA~GAA
Reverse pr    GGGTTATCTAAATTTTCAAGAATAGTTGATGTCTTTTCTAGAAGATTACAATTACAAGAA
Amino acid    G~~L~~S~~K~~F~~S~~R~~I~~V~~D~~V~~F~~S~~R~~R~~L~~Q~~L~~Q~~E~~

                  970        980        990        1000       1010       1020
              ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI       gatttaactaacgatatttgtaacgctttaaaaaaatacttaaaaccattatatattaaa
Forward pr    GATTTA~CTA~CGATATTTGTAACGCTTTAAAAAAATACT~~~AAACATTATAAATTAAA
Reverse pr    GATTTAACTAACGATATTTGTAACGCTTTAAAAAAATACTTAAAACCATTATATATTAAA
Amino acid    D~~L~~T~~N~~D~~I~~C~~N~~A~~L~~K~~K~~Y~~L~~K~~P~~L~~Y~~I~~K~~

                  1030       1040       1050       1060       1070       1080
              ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI       gtatctattgtagctaaacatttatgtataaatatgaggggagttaaagagcacgatgct
Forward pr    GTATCTATTGTAGCTAAACATTTATGTATAAATATTGAGGGAG~TAAAGAGAACGATGTC
Reverse pr    GTATCTATTGTAGCTAAACATTTATGTATAAATATGAGGGGAGTTAAAGAGCACGATGCT
Amino acid    V~~S~~I~~V~~A~~K~~H~~L~~C~~I~~N~~M~~R~~G~~V~~K~~E~~H~~D~~A~~

                  1090       1100       1110       1120       1130       1140
              ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI       aaaactataacgtatgcatcttataaagcagaaaaagagaatcctacagttcattcttta
Forward pr    TTAAACTTTAACGTATGCATCTTATAAAGCCGAAAAGAATCTCACAGTTCATCTTAATTG
Reverse pr    AAAACTATAACGTATGCATCTTATAAAGCAGAAAAAGAGAATCCTACAGTTCATTCTTTA
Amino acid    K~~T~~I~~T~~Y~~A~~S~~Y~~K~~A~~E~~K~~E~~N~~P~~T~~V~~H~~S~~L~~
```

```
                1150       1160       1170       1180       1190       1200
           ....|....|....|....|....|....|....|....|....|....|....|....|
ngtpchI    aatattgactcttctgtggaaaatttaaatAagCAggatccgaattcgagctccgtcgac
Forward pr AACTCTCTCTGTGTGAGGA
Reverse pr AATATTGACTCTTCTGTGGAAAATTTAAATAAGCAGGATCCGAATTCGAGCTCCGTCGAC
Amino acid N~~I~~D~~S~~S~~V~~E~~N~~L~~N~~R~~Q~~D~~P~~N~~S~~S~~S~~V~~D~~

                1210       1220       1230       1240
           ....|....|....|....|....|....|....|..
ngtpchI    aagcttgcggccgcactcgagcaccaccaccaccaccactga
Forward pr
Reverse pr AAGCTTGC~GCCGCACTCGAGCACCACCACCACCACCACTGA
Amino acid K~~L~~A~~A~~A~~L~~E~~H~~H~~H~~H~~H~~H~~*
```

Appendix E

The o*gtpchI* gene sequence and the alignment of sequences obtained from the nucleotide sequencing results (Refer to section 2.2.13) after sub-cloning o*gtpchI* into the pET22b expression vector.

```
                  10         20         30         40         50         60
           ....|....|....|....|....|....|....|....|....|....|....|....|
ogtpchI    CATatgtacaaatacacctctatcaacaagtccgacaagatctacgaaacccacaacatg
Forward pr CATATGTACAAATACACGTCAATCAATAAAAGCGACAAAATCTACGAAACCCACAACATG
Reverse pr ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Amino acid H~~M~~Y~~K~~Y~~T~~S~~I~~N~~K~~S~~D~~K~~I~~Y~~E~~T~~H~~N~~M~~

                  70         80         90        100        110        120
           ....|....|....|....|....|....|....|....|....|....|....|....|
ogtpchI    gaagagaagaagaaaaagggtaacaacaacaacttctctagcctgctgaacaacgaaatc
Forward pr GAAGAAAAAAAGAAAAAAGGCAACAACAATAACTTTTCTGGCCTGCTGAACAACGAAATA
Reverse pr ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Amino acid E~~E~~K~~K~~K~~K~~G~~N~~N~~N~~N~~F~~S~~G~~L~~L~~N~~N~~E~~I~~

                 130        140        150        160        170        180
           ....|....|....|....|....|....|....|....|....|....|....|....|
ogtpchI    gacgataataacaagaaagaaaaactgaaaaacagcatctctaaaatgtacagcaaccac
Forward pr GACGACAACAACAAAAAGGAAAAGCTGAAAAACTCTATCAGCAAAATGTACTCTAATCAC
Reverse pr ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Amino acid D~~D~~N~~N~~K~~K~~E~~K~~L~~K~~N~~S~~I~~S~~K~~M~~Y~~S~~N~~H~~

                 190        200        210        220        230        240
           ....|....|....|....|....|....|....|....|....|....|....|....|
ogtpchI    aagaaccgcgaaaacttcaacgaatgtgaaaaagaagacctggttgtaatcgacgaaaaa
Forward pr AAGAACCGTGAAAACTTCAATGAATGCGAGAAGGAAGATCTCGTTGTCATCGACGAAAAG
Reverse pr ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Amino acid K~~N~~R~~E~~N~~F~~N~~E~~C~~E~~K~~E~~D~~L~~V~~V~~I~~D~~E~~K~~
```

```
                250       260       270       280       290       300
           ....|....|....|....|....|....|....|....|....|....|....|....|
ogtpchI    gacaacaacaagaagaagaagaagaacatgaccaacaccttcgaacaggataacaactac
Forward pr GACAACAACAAAAAGAAAAAAAAAAACATGACCAACACCTTCGAACAGGACAACAACTAC
Reverse pr ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Amino acid D~~N~~N~~K~~K~~K~~K~~K~~N~~M~~T~~N~~T~~F~~E~~Q~~D~~N~~N~~Y~~

                310       320       330       340       350       360
           ....|....|....|....|....|....|....|....|....|....|....|....|
ogtpchI    aacatgaacgacaacaaacgtctgggttccttcttcaaaatcaacgacaagtgtgaatcg
Forward pr AACATGAACGACAACAAACGCCTGGGCTCGTTCTTCAAAATCAACGACAAATGCGAAAGC
Reverse pr ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~AAATCAACGACAAATGCGAAAGC
Amino acid N~~M~~N~~D~~N~~K~~R~~L~~G~~S~~F~~F~~K~~I~~N~~D~~K~~C~~E~~S~~

                370       380       390       400       410       420
           ....|....|....|....|....|....|....|....|....|....|....|....|
ogtpchI    atcaacgaaaacgtgagcaacatcaacaaacagtctctgaaagactccatcctgttcgac
Forward pr ATTAACGAGAACGTTAACAACATTAACAAACAAAGCCTCAAGGACAGCATCCTCTTCGAT
Reverse pr ATTAACGAGAACGTTAACAACATTAACAAACAAAGCCTCAAGGACAGCATCCTCTTCGAT
Amino acid I~~N~~E~~N~~V~~N~~N~~I~~N~~K~~Q~~S~~L~~K~~D~~S~~I~~L~~F~~D~~

                430       440       450       460       470       480
           ....|....|....|....|....|....|....|....|....|....|....|....|
ogtpchI    aacatcaacgaaaaggaatacttcaacgaaaccaaagaagaaaacaaggaaggtaacaag
Forward pr AACATCAACGAGAAAGAATACTTCAATGAAACGAAAGAAGAAAACAAGGAAGGCAACAAA
Reverse pr AACATCAACGAGAAAGAATACTTCAATGAAACGAAAGAAGAAAACAAGGAAGGCAACAAA
Amino acid N~~I~~N~~E~~K~~E~~Y~~F~~N~~E~~T~~K~~E~~E~~N~~K~~E~~G~~N~~K~~

                490       500       510       520       530       540
           ....|....|....|....|....|....|....|....|....|....|....|....|
ogtpchI    agcaacgatatcgaaaagatcaactgcatgaaagttaagaaaaagaccgttaagaaaaac
Forward pr TCTAACGACATCGAAAAGATCAACTGCATGAAAGTCAAGAAGAAAACTGTTAAAAAGAAC
Reverse pr TCTAACGACATCGAAAAGATCAACTGCATGAAAGTCAAGAAGAAAACTGTTAAAAAGAAC
Amino acid S~~N~~D~~I~~E~~K~~I~~N~~C~~M~~K~~V~~K~~K~~K~~T~~V~~K~~K~~N~~

                550       560       570       580       590       600
           ....|....|....|....|....|....|....|....|....|....|....|....|
ogtpchI    aagaagaagatcaacaagatcatcaacaacaagaacaaaatcagcaagtctaacgatatc
Forward pr AAAAAGAAGATTAACAAAATTATCAATAACAAGAACAAAATCAGCAAATCTAACGATATC
Reverse pr AAAAAGAAGATTAACAAAATTATCAATAACAAGAACAAAATCAGCAAATCTAACGATATC
Amino acid K~~K~~K~~I~~N~~K~~I~~I~~N~~N~~K~~N~~K~~I~~S~~K~~S~~N~~D~~I~~

                610       620       630       640       650       660
           ....|....|....|....|....|....|....|....|....|....|....|....|
ogtpchI    gaagaacagatcatcaacatctccaaacacatctacaagatcctgaacatctctaagctt
Forward pr GAAGAACAGATCATTAACATTTCTAAACACATCTACAAAATCCTGAACATCTCGAAACTG
Reverse pr GAAGAACAGATCATTAACATTTCTAAACACATCTACAAAATCCTGAACATCTCGAAACTG
Amino acid E~~E~~Q~~I~~I~~N~~I~~S~~K~~H~~I~~Y~~K~~I~~L~~N~~I~~S~~K~~L~~

                670       680       690       700       710       720
           ....|....|....|....|....|....|....|....|....|....|....|....|
ogtpchI    ccgaagtgtgatatcctgaaacgtactaaccgtcgttacgcggaaaccttcctgtacctg
Forward pr CCGAAATGCGACATCCTGAAACGTACCAACCGTCGTTACGCTGAAACCTTCCTGTACCTG
Reverse pr CCGAAATGCGACATCCTGAAACGTACCAACCGTCGTTACGCTGAAACCTTCCTGTACCTG
Amino acid P~~K~~C~~D~~I~~L~~K~~R~~T~~N~~R~~R~~Y~~A~~E~~T~~F~~L~~Y~~L~~
```

```
                730       740       750       760       770       780
           ....|....|....|....|....|....|....|....|....|....|....|....|
ogtpchI    actaacggttacaacctggacatcgaacagattattaaacgttccctgtacaagcgtatg
Forward pr ACTAACGGCTACAACTTAGACATCGAACAGATCATCAAACGTAGCCTGTACAAACGCATG
Reverse pr ACTAACGGCTACAACTTAGACATCGAACAGATCATCAAACGTAGCCTGTACAAACGCATG
Amino acid T~~N~~G~~Y~~N~~L~~D~~I~~E~~Q~~I~~I~~K~~R~~S~~L~~Y~~K~~R~~M~~

                790       800       810       820       830       840
           ....|....|....|....|....|....|....|....|....|....|....|....|
ogtpchI    tacaagaacaactctatcatcaaagtgaccggtatccacatctactccctgtgtaaacac
Forward pr TACAAAAACAACTCTATCATAAAAGTTACCGGCATCCACATCTACTCTCTGTGCAAACAT
Reverse pr TACAAAAACAACTCTATCATAAAAGTTACCGGCATCCACATCTACTCTCTGTGCAAACAT
Amino acid Y~~K~~N~~N~~S~~I~~I~~K~~V~~T~~G~~I~~H~~I~~Y~~S~~L~~C~~K~~H~~

                850       860       870       880       890       900
           ....|....|....|....|....|....|....|....|....|....|....|....|
ogtpchI    cacctgcttccattcgaaggtacctgcgatatcgaatacatcccgaacaaatacatcatc
Forward pr CACCTCCTGCCATTCGAAGGCACCTGCGACATTGAGTACATCCCTAACAAATACATTATA
Reverse pr CACCTCCTGCCATTCGAAGGCACCTGCGACATTGAGTACATCCCTAACAAATACATTATA
Amino acid H~~L~~L~~P~~F~~E~~G~~T~~C~~D~~I~~E~~Y~~I~~P~~N~~K~~Y~~I~~I~~

                910       920       930       940       950       960
           ....|....|....|....|....|....|....|....|....|....|....|....|
ogtpchI    ggtctgtctaagttctctcgtatcgtggatgtattctctcgtcgtctgcaactgcaagaa
Forward pr GGGCTGAGCAAATTCTCTCGTATCGTTGACGTCTTCAGCCGTCGTCTGCAGCTGCAGGAA
Reverse pr GGGCTGAGCAAATTCTCTCGTATCGTTGACGTCTTCAGCCGTCGTCTGCAGCTGCAGGAA
Amino acid G~~L~~S~~K~~F~~S~~R~~I~~V~~D~~V~~F~~S~~R~~R~~L~~Q~~L~~Q~~E~~

                970       980       990       1000      1010      1020
           ....|....|....|....|....|....|....|....|....|....|....|....|
ogtpchI    gatctgaccaacgacatctgcaacgcgctgaaaaaatacctgaagccgctgtacatcaag
Forward pr GACCTGACTAATGACATTTGCAT~GCTCTGAAAAAATATCTGAAACCG
Reverse pr GACCTGACTAATGACATTTGCAATGCTCTGAAAAAATATCTGAAACCGCTGTACATTAAA
Amino acid D~~L~~T~~N~~D~~I~~C~~N~~A~~L~~K~~K~~Y~~L~~K~~P~~L~~Y~~I~~K~~

                1030      1040      1050      1060      1070      1080
           ....|....|....|....|....|....|....|....|....|....|....|....|
ogtpchI    gtttccatcgttgctaagcacctgtgcatcaacatgcgtggtgtgaaggaacacgacgcg
Forward pr
Reverse pr GTTAGCATTGTTGCTAAACACCTGTGCATCAACATGCGCGGCGTTAAAGAGCATGACGCT
Amino acid V~~S~~I~~V~~A~~K~~H~~L~~C~~I~~N~~M~~R~~G~~V~~K~~E~~H~~D~~A~~

                1090      1100      1110      1120      1130      1140
           ....|....|....|....|....|....|....|....|....|....|....|....|
ogtpchI    aagactatcacttacgcttcctacaaagctgaaaaagaaaacccgactgtacactccctg
Forward pr
Reverse pr AAAACTATCACGTACGCTAGCTACAAAGCTGAAAAAGAGAACCCAACCGTTCACAGCCTG
Amino acid K~~T~~I~~T~~Y~~A~~S~~Y~~K~~A~~E~~K~~E~~N~~P~~T~~V~~H~~S~~L~~

                1150      1160      1170      1180      1190      1200
           ....|....|....|....|....|....|....|....|....|....|....|....|
ogtpchI    aacatcgactcctccgttgaaaacctgaacAAACaggatccgaattcgagctccgtcgac
Forward pr
Reverse pr AACATTGATAGCAGCGTCGAAAACCTGAACAGACAGGATCCGAATTCGAGCTCCGTCGAC
Amino acid N~~I~~D~~S~~S~~V~~E~~N~~L~~N~~R~~Q~~D~~P~~N~~S~~S~~S~~V~~D~~
```

```
                1210      1220      1230      1240
           ....|....|....|....|....|....|....|..
ogtpchI    aagcttgcggccgcactcgagcaccaccaccaccaccactga
Forward pr
Reverse pr AAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGA
Amino acid K~~L~~A~~A~~A~~L~~E~~H~~H~~H~~H~~H~~H~~*
```

Appendix F

The h*gtpch*I gene sequence and the alignment of sequences obtained from the nucleotide sequencing results (Refer to section 2.2.13) after sub-cloning h*gtpch*I into the pET22b expression vector.

```
                  10        20        30        40        50        60
           ....|....|....|....|....|....|....|....|....|....|....|....|
hgtpchI    ~~~cataTGTACAAATACACGTCTATCAATAAAAGCGACAAAATCTACGAAACCCACAAC
forward pr …..caTATGTACAAATA~~~~~~~~~~~~~~~~~GCGACAaAATCTACGAAACCCACAA~
reverse pr ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
amino acid ~~~~~H~~M~~Y~~K~~Y~~T~~S~~I~~N~~K~~S~~D~~K~~I~~Y~~E~~T~~H~~N

                  70        80        90       100       110       120
           ....|....|....|....|....|....|....|....|....|....|....|....|
hgtpchI    ATGGAAGAAAAAAAagAAAAAAGGCAACAACAATAACTTTTCTGGCCTGCTGAACAACGAA
forward pr A~GGA~~~~~YAAAGAAAAAAGGCAACAACAATAACTTTTCTGGCCTGCTGAACAACGAA
reverse pr ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
amino acid ~~M~~E~~E~~K~~K~~K~~K~~G~~N~~N~~N~~N~~F~~S~~G~~L~~L~~N~~N~~E

                 130       140       150       160       170       180
           ....|....|....|....|....|....|....|....|....|....|....|....|
hgtpchI    ATAGACGACAACAACAAAAAGGAAAAGCTGAAAAACTCTATCAGCAAAATGTACTCTAAT
forward pr ATAGACGACAACAACAAAWAGGAAAAGCTGAAAAACTCTATCAGCAAAATGTACTCTAAT
reverse pr ~~~~~~~~~~~~~~~~~~~~~~~~~~~CTGAAAAACTCTATCAGCAAAATGTACTCTAAT
amino acid ~~I~~D~~D~~N~~N~~K~~K~~E~~K~~L~~K~~N~~S~~I~~S~~K~~M~~Y~~S~~N

                 190       200       210       220       230       240
           ....|....|....|....|....|....|....|....|....|....|....|....|
hgtpchI    CACAAGAACCGTGAAAACTTCAATGAATGCGAGAAGGAAGATCTCGTTGTCATCGACGAA
forward pr CACAAGAACCGTGAAAACTTCAATGAATGCGAGAAGGAAGATCTCGTTGTCATCGACGAA
reverse pr CACA~GA~~CGTGAAA~CTCC~ATGA~TGCGAGAACGAAGATCTCGTCGTCATCGACGAA
amino acid ~~H~~K~~N~~R~~E~~N~~F~~N~~E~~C~~E~~K~~E~~D~~L~~V~~V~~I~~D~~E

                 250       260       270       280       290       300
           ....|....|....|....|....|....|....|....|....|....|....|....|
hgtpchI    AAGGACAACAACAAAAAGAAAAAAAAAAAACATGACCAACACCTTCGAACAGGACAACAAC
forward pr AAGGACAACAACAAAAAGAAAAAAAAAAAACATGACCAACACCTTCGAACAGGACAACAAC
reverse pr AAGGACAACA~CAAAAAGAAAAAAAAAAAACATGACCAACAC~TTCGAACAGGACAACAAC
amino acid ~~K~~D~~N~~N~~K~~K~~K~~K~~N~~M~~T~~N~~T~~F~~E~~Q~~D~~N~~N
```

```
                310       320       330       340       350       360
           ....|....|....|....|....|....|....|....|....|....|....|....|
hgtpchI    TACAACATGAACGACAACAAACGCCTGGGCTCGTTCTTCAAAATCAACGACAAATGCGAA
forward pr TACAACATGAACGACAACAAACGCCTGGGCTCGTTCTTCAAAATCAACGACAAATGCGAA
reverse pr TACAACATGAACGACAACAAACGCCTGGGCTCGTTCTTCAAAATCAACGACAAATGCGAA
amino acid ~~Y~~N~~M~~N~~D~~N~~K~~R~~L~~G~~S~~F~~F~~K~~I~~N~~D~~K~~C~~E

                370       380       390       400       410       420
           ....|....|....|....|....|....|....|....|....|....|....|....|
hgtpchI     AGCATTAACGAGAACGTTAACAACATTAACAAACAAAGCCTCAAGGACAGCATCCTCTTC
forward pr AGCATTAACGAGAACGTTAACAACATTAACAAACAAAGCCTCAAGGACAGCATCCTCTTC
reverse pr AGCATTAACGAGAACGTTAACAACATTAACAAACAAAGCCTCAAGGACAGCATCCTCTTC
amino acid ~~S~~I~~N~~E~~N~~V~~N~~N~~I~~N~~K~~Q~~S~~L~~K~~D~~S~~I~~L~~F

                430       440       450       460       470       480
           ....|....|....|....|....|....|....|....|....|....|....|....|
hgtpchI     GATAACATCAACGAGAAAGAATACTTCAATGAAACGAAAGAAGAAAACAAGGAAGGCAAC
forward pr GATAACATCAACGAGAAAGAATACTTCAATGAAACGAAAGAAGAAAACAAGGAAGGCAAC
reverse pr GATAACATCAACGAGAAAGAATACTTCAATGAAACGAAAGAAGAAAACAAGGAAGGCAAC
amino acid ~~D~~N~~I~~N~~E~~K~~E~~Y~~F~~N~~E~~T~~K~~E~~E~~N~~K~~E~~G~~N

                490       500       510       520       530       540
           ....|....|....|....|....|....|....|....|....|....|....|....|
hgtpchI     AAATCTAACGACATCGAAAAGATCAACTGCATGAAAGTCAAGAAGAAAAACTGTTAAAAAG
forward pr AAATCTAACGACATCGAAAAGATCAACTGCATGAAAGTCAAGAAGAAAAACTGTTAAAAAG
reverse pr AAATCTAACGACATCGAAAAGATCAACTGCATGAAAGTCAAGAAGAAAAACTGTTAAAAAG
amino acid ~~K~~S~~N~~D~~I~~E~~K~~I~~N~~C~~M~~K~~V~~K~~K~~K~~T~~V~~K~~K

                550       560       570       580       590       600
           ....|....|....|....|....|....|....|....|....|....|....|....|
hgtpchI     AACAAAAAGAAGATTAACAAAATTATCAATAACAAGAACAAAATCAGCAAATCTAACGAT
forward pr AACAAAAAGAAGATTAACAAAATTATCAATAACAAGAACAAAATCAGCAAATCTAACGAT
reverse pr AACAAAAAGAAGATTAACAAAATTATCAATAACAAGAACAAAATCAGCAAATCTAACGAT
amino acid ~~N~~K~~K~~K~~I~~N~~K~~I~~I~~N~~N~~K~~N~~K~~I~~S~~K~~S~~N~~D

                610       620       630       640       650       660
           ....|....|....|....|....|....|....|....|....|....|....|....|
hgtpchI     ATCGAAGAACAGATCATTAACATTTCTAAACACATCTACAAAATCCTGAACATCTCGAAA
forward pr ATCGAAGAACAGATCATTAACATTTCTAAACACATCTACAAAATCCTGAACATCTCGAAA
reverse pr ATCGAAGAACAGATCATTAACATTTCTAAACACATCTACAAAATCCTGAACATCTCGAAA
amino acid ~~I~~E~~E~~Q~~I~~I~~N~~I~~S~~K~~H~~I~~Y~~K~~I~~L~~N~~I~~S~~K

                670       680       690       700       710       720
           ....|....|....|....|....|....|....|....|....|....|....|....|
hgtpchI     CTGCCGAAATGCGACATCCTGAAACGTACCAACCGTCGTTACGCTGAAACCTTCCTGTAC
forward pr CTGCCGAAATGCGACATCCTGAAACGTACCAACCGTCGTTACGCTGAAACCTTCCTGTAC
reverse pr CTGCCGAAATGCGACATCCTGAAACGTACCAACCGTCGTTACGCTGAAACCTTCCTGTAC
amino acid ~~L~~P~~K~~C~~D~~I~~L~~K~~R~~T~~N~~R~~R~~Y~~A~~E~~T~~F~~L~~Y

                730       740       750       760       770       780
           ....|....|....|....|....|....|....|....|....|....|....|....|
hgtpchI     CTGACTAACGGCTACAACTTAGACATCGAACAGATCATCAAACGTAGCCTGTACAAACGC
forward pr CTGACTAACGGCTACAACTTAGACATCGAACAGATCATCAAACGTAGCCTGTACAAACGC
reverse pr CTGACTAACGGCTACAACTTAGACATCGAACAGATCATCAAACGTAGCCTGTACAAACGC
amino acid ~~L~~T~~N~~G~~Y~~N~~L~~D~~I~~E~~Q~~I~~I~~K~~R~~S~~L~~Y~~K~~R
```

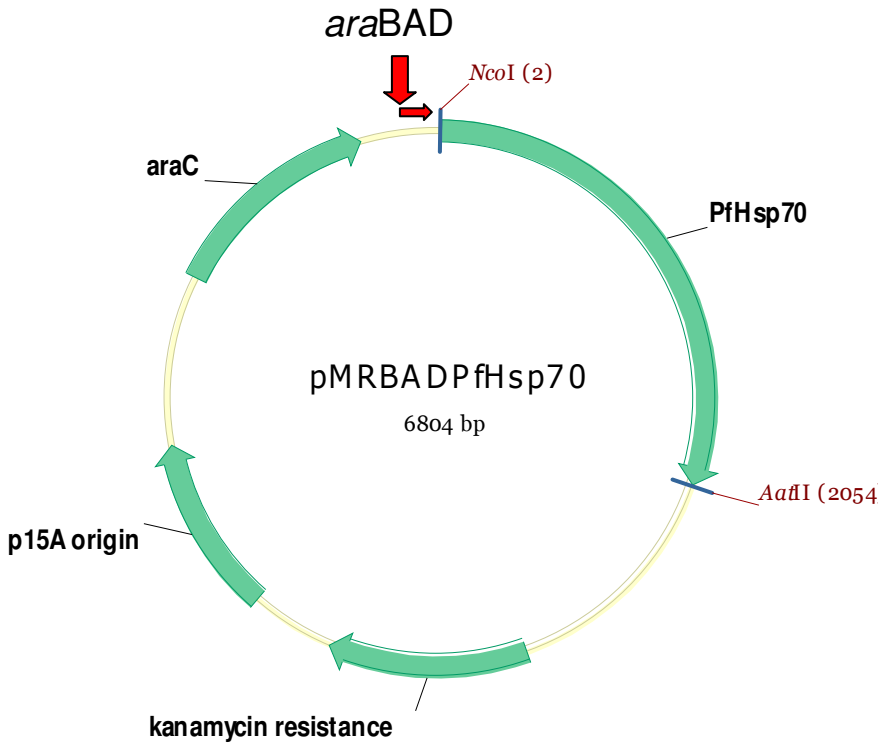
```
                  790       800       810       820       830       840
              ....|....|....|....|....|....|....|....|....|....|....|....|
hgtpchI       ATGTACAAAAACAACTCTATCATAAAAGTTACCGGCATCCACATCTACTCTCTGTGCAAA
forward pr    ATGTACAAAAACAACTCTATCATAAAAGTTACCGGCATCCACATCTACTCTCTGTGCAAA
reverse pr    ATGTACAAAAACAACTCTATCATAAAAGTTACCGGCATCCACATCTACTCTCTGTGCAAA
amino acid    ~~M~~Y~~K~~N~~N~~S~~I~~I~~K~~V~~T~~G~~I~~H~~I~~Y~~S~~L~~C~~K

                  850       860       870       880       890       900
              ....|....|....|....|....|....|....|....|....|....|....|....|
hgtpchI       CATCACCTCCTGCCATTCGAAGGCACCTGCGACATTGAGTACATCCCTAACAAATACATT
forward pr    CATCACCTCCTGCCATTCGAAGGCACCTGCGACATTGAGTACATCCCTAACAAATACATT
reverse pr    CATCACCTCCTGCCATTCGAAGGCACCTGCGACATTGAGTACATCCCTAACAAATACATT
amino acid    ~~H~~H~~L~~L~~P~~F~~E~~G~~T~~C~~D~~I~~E~~Y~~I~~P~~N~~K~~Y~~I

                  910       920       930       940       950       960
              ....|....|....|....|....|....|....|....|....|....|....|....|
hgtpchI       ATAGGGCTGAGCAAATTCTCTCGTATCGTTGACGTCTTCAGCCGTCGTCTGCAGCTGCAG
forward pr    ATAGGGCTGAGCAAATTCTCTCGTATCGTTGACGTCTTCAGCCGTCGTCTGCAGCTGCAG
reverse pr    ATAGGGCTGAGCAAATTCTCTCGTATCGTTGACGTCTTCAGCCGTCGTCTGCAGCTGCAG
amino acid    ~~I~~G~~L~~S~~K~~F~~S~~R~~I~~V~~D~~V~~F~~S~~R~~R~~L~~Q~~L~~Q

                  970       980       990       1000      1010      1020
              ....|....|....|....|....|....|....|....|....|....|....|....|
hgtpchI       GAAGACCTGACTAATGACATTTGCAATGCTCTGAAAAAATATCTGAAACCGCTGTACATT
forward pr    GAAGACCTGACTAATGACATTTGCAATGCTCTGAAAAA~TATCTGAA~~CGCTGTACATT
reverse pr    GAAGACCTGACTAATGACATTTGCAATGCTCTGAAAAAATATCTGAAACCGCTGTACATT
amino acid    ~~E~~D~~L~~T~~N~~D~~I~~C~~N~~A~~L~~K~~K~~Y~~L~~K~~P~~L~~Y~~I

                  1030      1040      1050      1060      1070      1080
              ....|....|....|....|....|....|....|....|....|....|....|....|
hgtpchI       AAAGTTAGCATTGTTGCTAAACACCTGTGCATCAACATGCGCGGCGTTAAAGAGCATGAC
forward pr    AA~GGTAGCATGG~TGCTAA~CAC~TGTGCATCAACATGCGCGGCGTK
reverse pr    AAAGTTAGCATTGTTGCTAAACACCTGTGCATCAACATGCGCGGCGTTAAAGAGCATGAC
amino acid    ~~K~~V~~S~~I~~V~~A~~K~~H~~L~~C~~I~~N~~M~~R~~G~~V~~K~~E~~H~~D

                  1090      1100      1110      1120      1130      1140
              ....|....|....|....|....|....|....|....|....|....|....|....|
hgtpchI       GCTAAAACTATCACGTACGCTAGCTACAAAGCTGAAAAAGAGAACCCAACCGTTCACAGC
forward pr
reverse pr    GCTAAAACTATCACGTACGCWAGCTACAAAGCTGAAAAAGAGAACCCAACCGTTCACAGC
amino acid    ~~A~~K~~T~~I~~T~~Y~~A~~S~~Y~~K~~A~~E~~K~~E~~N~~P~~T~~V~~H~~S

                  1150      1160      1170      1180      1190      1200
              ....|....|....|....|....|....|....|....|....|....|....|....|
hgtpchI       CTGAACATTGATAGCAGCGTCGAAAACCTGAACaGACaggatccgaattcgagctccgtc
forward pr
reverse pr    CTGAACATTGATAGCAGCGTCSAAKKCCTTC~CA~ACAGGATCCGAATTCGAGCTCCGTC
amino acid    ~~L~~N~~I~~D~~S~~S~~V~~E~~N~~L~~N~~R~~Q~~D~~P~~N~~S~~S~~S~~V

                  1210      1220      1230      1240
              ....|....|....|....|....|....|....|....|
hgtpchI       gacaagcttgcggccgcactcgagcaccaccaccaccaccactga
forward pr
reverse pr    GA~AAG~~~~~~~~~~CACTCGAGCACCACCACCWCCACCACTGA
amino acid    ~~D~~K~~L~~A~~A~~A~~L~~E~~H~~H~~H~~H~~H~~H~~*
```

Appendix G

Vector map for chaperone protein (*Pf*hsp70)

(Stephans L., Rhodes University)

Plasmid	Chaperone	Promoter	Inducer	Resistant marker
pMRBAD-*Pf*Hsp70	*Pf*Hsp70	araBAD	L- arabinose	kanamycin



Appendix H

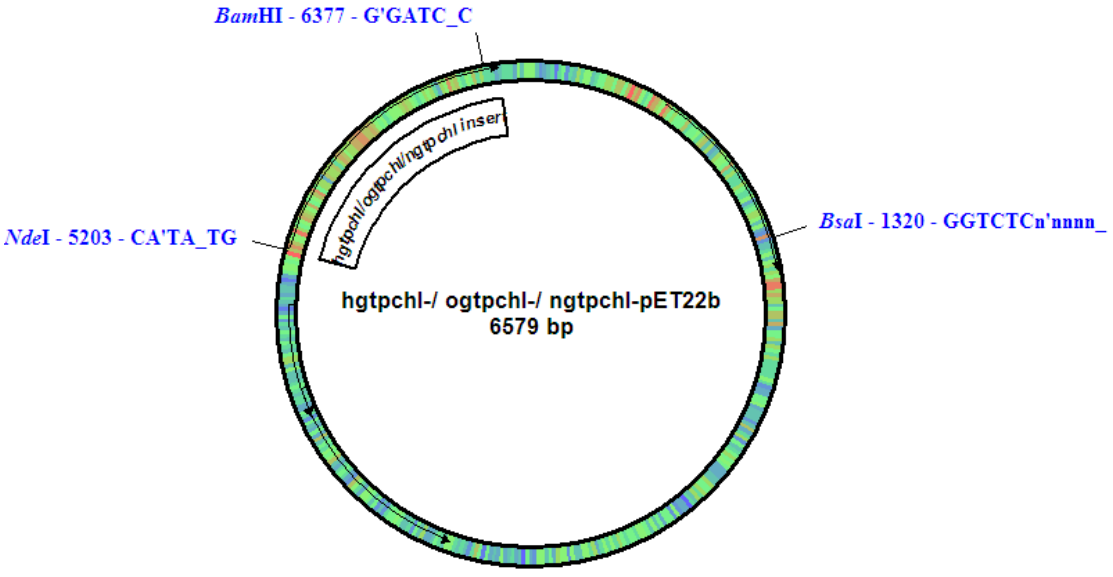
Table 1: The components of the Promega (USA) and Endext technology (Cell free Sciences, Japan) Wheat Germ kit used for protein synthesis.

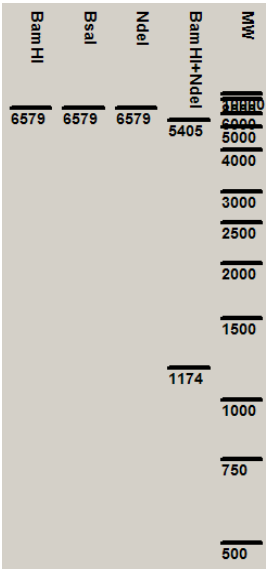
Components	[Components] of Promega, Wheat Germ kit (mM)	[Components] of Endext technology, Wheat Germ kit (mM)
Potassium acetate	90	100
Magnesium acetate	1.5	2.7
Spermidine	0.5	0.4
Dithiothreitol	5	2.5
Amino acid mix	0.08	0.3
ATP	1.2	1.2
GTP	0.1	0.25
Creatine phosphate	10	16
HEPES-KOH	-	30
Creatine phosphokinase	50µg/ml	**-**

Appendix I

*gtpch*I-pET22b vector

The plasmid vector map and predicted bands after restriction enzyme digestion (*Bsa*I) for linearization of the *gtpch*I-pET22b vector. The restriction enzyme sites flanking the insert gene are indicated in blue (BamHI-NdeI).

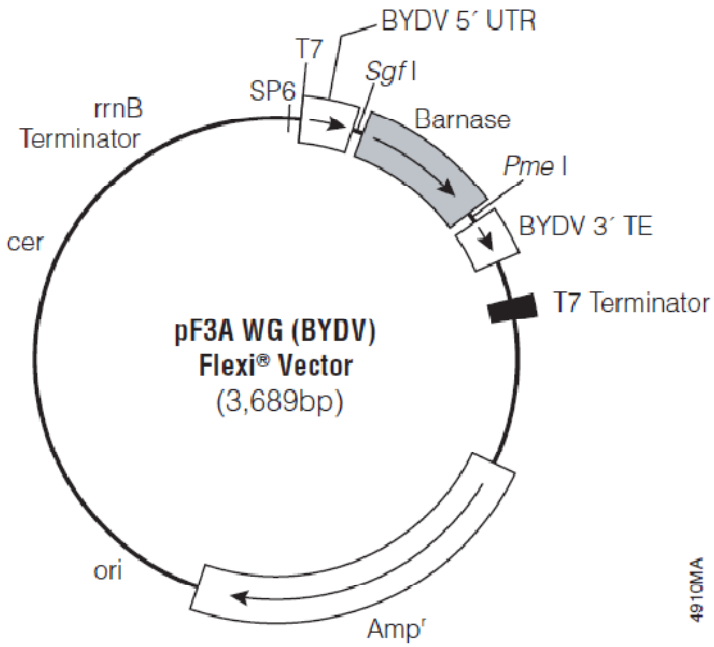




Appendix J

Pf3A WG (BYDV) Flexi® Vector

Wheat Germ Flexi® Vectors are used for expression of proteins in wheat germ extract. They incorporate sequences from the barley yellow dwarf virus (BYDV), upstream and downstream of the target protein coding region.



References

Abdulla, S., Oberholzer, R., Juma, O., *et al.* (2008) **Safety and Immunogenicity of RTS,S/AS02D Malaria Vaccine in Infants**. *New England Journal of Medicine*. **359**, 2533-2544

Acharya, P., Kumar, R., and Tatu, U. (2007) **Chaperoning a cellular upheaval in malaria: Heat shock proteins in Plasmodium falciparum**, *Molecular and Biochemical Parasitology 153*, 85-94.

Adams, M. M., Damani, P., Perl, N. R., *et al*. (2009) **Design and Synthesis of Potent Quillaja Saponin Vaccine Adjuvants**. *Journal of the American Chemical Society*. **132**, 1939-1945

Aguiar, J. C., LaBaer, J., Blair, P. L., *et al* (2004) **High-throughput generation of P. falciparum functional molecules by recombinational cloning**. *Genome Research*. 14, 2076-2082

Amexo, M., Tolhurst, R., Barnish, G. and Bates, I. (2004) **Malaria misdiagnosis: effects on the poor and vulnerable**. *Lancet*. **364**, 1896-1898

Angov, E., Hillier, C. J., Kincaid, R. L., and Lyon, J. A. (2008) **Heterologous Protein Expression Is Enhanced by Harmonizing the Codon Usage Frequencies of the Target Gene with those of the Expression Host**, *Plos One 3*.

Angov, E., and Lyon, J. A. (2006) **Overcoming obstacles in protein expression using codon harmonization**, *Parasite Immunology 28*, 250-250.

Angov, E., Aufiero, B. M., Turgeon, A. M., *et al*. (2003) **Development and pre-clinical analysis of a Plasmodium falciparum Merozoite Surface Protein-1(42) malaria vaccine**, *Molecular and Biochemical Parasitology 128*, 195-204.

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

Banerjee, A. K., Arora, N. and Murty, U. S. N. (2009) **Structural model of the Plasmodium falciparum Thioredoxin reductase: a novel target for antimalarial drugs**. *Journal of Vector Borne Diseases*. **46**, 171-183

Banerjee, A., Lagardere, B. and Grp Pediatrie, T. (2001) **Prevention of malaria**. *Archives De Pediatrie*. **8**, 275S-277S

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

Bannister, L. H., Hopkins, J. M., Fowler, *et al*. (2000) **A brief illustrated guide to the ultrastructure of Plasmodium falciparum asexual blood stages**. *Parasitology Today*. **16**, 427-433

Bell, A. (1998) **Microtubule inhibitors as potential antimalarial agents**. *Parasitology Today*. **14**, 234-240

Bentley, W. E., and Kompala, D. S. (1990) **optimal induction of protein-synthesis in recombinant bacterial cultures**, *Annals of the New York Academy of Sciences 589*, 121-138.

Bermingham, A. and Derrick, J. P. (2002) **The folic acid biosynthesis pathway in bacteria: evaluation of potential for antibacterial drug discovery**. *Bioessays*. **24**, 637-648

Berrow, N. S., Bussow, K., Coutard, B., *et al* (2006) **Recombinant protein expression and solubility screening in Escherichia coli: a comparative study**, *Acta Crystallographica Section D-Biological Crystallography 62*, 1218-1226.

Blau, N. and Niederwieser, A. (1983) **guanosine triphosphate cyclohydrolase-i assay in human and rat-liver using high-performance liquid-chromatography of neopterin phosphates and guanine-nucleotides**. *Analytical Biochemistry*. **128**, 446-452

Birkholtz, L. M., Blatch, G., Coetzer, T. L., *et al*. (2008a) **Heterologous expression of plasmodial proteins for structural studies and functional annotation**, *Malaria Journal 7*.

Birkholtz, L., van Brummelen, A. C., Clark, K., *et al* (2008) **Exploring functional genomics for drug target and therapeutics discovery in Plasmodia**. *Acta Tropica*. **105**, 113-123

Birkholtz, L. M., Bastien, O., Wells, G., *et al* (2006) **Integration and mining of malaria molecular, functional and pharmacological data: how far are we from a chemogenomic knowledge space?** *Malaria Journal*. **5**

Botha, M., Pesce, E. R., and Blatch, G. L. (2007) **The Hsp40 proteins of Plasmodium falciparum and other apicomplexa: Regulating chaperone power in the parasite and the host**, *International Journal of Biochemistry & Cell Biology 39*, 1781-1803.

Boddey, J. A., Hodder, A. N., Gunther, S., *et al*. (2010) **An aspartyl protease directs malaria effector proteins to the host cell**. *Nature*. **463**, 627-U652

Bornman, R., Barnhoorn, I. and Aneck-Hahn, N. (2009) **A pilot study on the occurrence of endocrine disruptive chemicals in a DDT-sprayed area**. *WRC Report*, xiv + 38 pp.

Bradford, M. M. (1976) **rapid and sensitive method for quantitation of microgram quantities of protein utilizing principle of protein-dye binding**, *Analytical Biochemistry 72*, 248-254.

Bray, P. G., Martin, R. E., Tilley, L., Ward, S. A., Kirk, K. and Fidock, D. A. (2005) **Defining the role of PfCRT in Plasmodium falciparum chloroquine resistance**. *Molecular Microbiology*. **56**, 323-333

Burg, A. W. and Brown, G. M. (1968) **biosynthesis of folic acid .8. purification and properties of enzyme that catalyzes production of formate from carbon atom 8 of guanosine triphosphate**. *Journal of Biological Chemistry*. **243**, 2349-&

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

Carrio, M. M. and Villaverde, A. (2003) **Role of molecular chaperones in inclusion body formation**. *Febs Letters*. **537**, 215-221

Cereghino, J. L. and Cregg, J. M. (2000) **Heterologous protein expression in the methylotrophic yeast Pichia pastoris**. *Fems Microbiology Reviews*. 24, 45-66

Cereghino, G. P. L. and Cregg, J. M. (1999) **Applications of yeast in biotechnology: protein production and genetic analysis.** *Current Opinion in Biotechnology*. 10, 422-427

Certain, L. K. and Sibley, C. H. (2007) **Plasmodium falciparum: A novel method for analyzing haplotypes in mixed infections**. *Experimental Parasitology*. 115, 233-241

Choi, S. R., Mukherjee, P. and Avery, M. A. (2008) **The fight against drug-resistant malaria: Novel plasmodial targets and antimalarial drugs**. *Current Medicinal Chemistry*. **15**, 161-171

Cheetham, M. E., Jackson, A. P., and Anderton, B. H. (1994) **Regulation of 70-Kda Heat-Shock-Protein Atpase Activity and Substrate-Binding by Human Dnaj-Like Proteins, Hsj1a and Hsj1b**, *European Journal of Biochemistry 226*, 99-107.

Chowdhury, D. R., Angov, E., Kariuki, T., and Kumar, N. (2009) **A Potent Malaria Transmission Blocking Vaccine Based on Codon Harmonized Full Length Pfs48/45 Expressed in Escherichia coli**, *Plos One 4*.

Christopherson, R. I., Seymour, K. K. and Yeo, A. E. T. (1998) **Pyrimidine antagonists and antifolates as antimalarial drugs**. *Purine and Pyrimidine Metabolism in Man Ix*. **431**, 705-709

Collins, C. R., Withers-Martinez, C., Bentley, G. A., *et al* (2007) F**ine mapping of an epitope recognized by an invasion-inhibitory monoclonal antibody on the malaria vaccine candidate apical membrane antigen 1.** *Journal of Biological Chemistry*. 282, 7431-7441

Coward, J. K. and McGuire, J. J. (2008) **mechanism-based inhibitors of folylpoly-gamma-glutamate synthetase and gamma-glutamyl hydrolase: control of folylpoly-gamma-glutamate homeostasis as a drug target**. *Folic Acid and Folates*. **79**, 347-373

Cowman, A. F., Galatis, D. and Thompson, J. K. (1994) **Selection for Mefloquine Resistance in Plasmodium-Falciparum Is Linked to Amplification of the Pfmdr1 Gene and Cross-Resistance to Halofantrine and Quinine**. *Proceedings of the National Academy of Sciences of the United States of America*. **91**, 1143-1147

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. *Bioorganic & Medicinal Chemistry*. **14**, 4433-4443

Delfino, R. T., Santos, O. A. and Figueroa-Villar, J. D. (2002) **Type 2 antifolates in the chemotherapy of falciparum malaria**. *Journal of the Brazilian Chemical Society*. **13**, 727-741

Delfino, R. T., Santos, O. A. and Figueroa-Villar, J. D. (2002) **Molecular modeling of wild-type and antifolate resistant mutant Plasmodium falciparum DHFR**. *Biophysical Chemistry*. **98**, 287-300

Demain, A. L. and Vaishnav, P. (2009) **Production of recombinant proteins by microbes and higher organisms**. *Biotechnology Advances*. 27, 297-306

de Marco, A., and De Marco, V. (2004) **Bacteria co-transformed with recombinant proteins and chaperones cloned in independent plasmids are suitable for expression tuning**, *Journal of Biotechnology 109*, 45-52.

Dhanawat, M., Das, N., Nagarwal, R. C. and Shrivastava, S. K. (2009) **Antimalarial Drug Development: Past to Present Scenario**. *Mini-Reviews in Medicinal Chemistry*. **9**, 1447-1469

Dittrich, S., Mitchell, S. L., Blackborough, A. M., Wang, P., Sims, P. F. G. and Hyde, J. E. (2008) **An atypical ortholog of 6-pyruvoyltetrahydropterin synthase provides the missing link in the folate biosynthesis of malaria parasites**. *International Journal for Parasitology*. **38**, S42-S43

Ditzen, M., Pellegrino, M. and Vosshall, L. B. (2008) **Insect odorant receptors are molecular targets of the insect repellent DEET**. *Science*. **319**, 1838-1842

Dobano, C., Sedegah, M., Rogers, W. O., Kumar, S., Zheng, H., Hoffman, S. L., and Doolan, D. L. (2009) **Plasmodium: Mammalian codon optimization of malaria plasmid DNA vaccines enhances antibody responses but not T cell responses nor protective immunity**, *Experimental Parasitology 122*, 112-123.

Dong, C. K., Patel, V., Yang, J. C., Dvorin, *et al*. (2009) **Type II NADH dehydrogenase of the respiratory chain of Plasmodium falciparum and its inhibitors**. *Bioorganic & Medicinal Chemistry Letters*. **19**, 972-975

Draper, S. J., Goodman, A. L., Biswas, S., *et al*. (2009) **Recombinant Viral Vaccines Expressing Merozoite Surface Protein-1 Induce Antibody- and T Cell-Mediated Multistage Protection against Malaria**. *Cell Host & Microbe*. **5**, 95-105

Draper, S. J., Moore, A. C., Goodman, A. L., *et al*. (2008) **Effective induction of high-titer antibodies by viral vector vaccines**. *Nature Medicine*. **14**, 819-821

El Bissati, K., Zufferey, R., Witola, W. H., Carter, N. S., Ullman, B. and Ben Mamoun, C. (2006) **The plasma membrane permease PfNT1 is essential for purine salvage in the human malaria parasite Plasmodium falciparum**. *Proceedings of the National Academy of Sciences of the United States of America*. **103**, 9286-9291

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

Endo, Y. and Sawasaki, T. (2002) **Highly efficient cell-free protein synthesis system prepared from wheat embryos**. *Seikagaku*. **74**, 326-330

Erickson, A. H. and Blobel, G. (1983) **cell-free translation of messenger-rna in a wheat-germ system.** *Methods in Enzymology*. 96, 38-50

Ferlan, J. T., Mookherjee, S., Okezie, I. N., Fulgence, L. and Sibley, C. H. (2001) **Mutagenesis of dihydrofolate reductase from Plasmodium falciparum: analysis in Saccharomyces cerevisae of triple mutant alleles resistant to pyrimethamine or WR99210.** *Molecular and Biochemical Parasitology*. 113, 139-150

Fidock, D. A., Eastman, R. T., Ward, S. A. and Meshnick, S. R. (2008) **Recent highlights in antimalarial drug resistance and chemotherapy research**. *Trends in Parasitology*. **24**, 537-544

Fidock, D. A., Nomura, T. and Wellems, T. E. (1998) **Cycloguanil and its parent compound proguanil demonstrate distinct activities against Plasmodium falciparum malaria parasites transformed with human dihydrofolate reductase**. *Molecular Pharmacology*. **54**, 1140-1147

Flaherty, K. M., Delucaflaherty, C., and McKay, D. B. (1990) **3-Dimensional Structure of the Atpase Fragment of a 70k Heat-Shock Cognate Protein**, *Nature 346*, 623-628

Flick, K., Ahuja, S., Chene, A., Bejarano *et al*. (2004) **Optimized expression of Plasmodium falciparum erythrocyte membrane protein I domains in Escherichia coli**, *Malaria Journal 3*.

Fogel, G. B., Cheung, M., Pittman, E. and Hecht, D. (2008) **In silico screening against wild-type and mutant Plasmodium falciparum dihydrofolate reductase**. *Journal of Molecular Graphics & Modelling*. **26**, 1145-1152

Foley, M. and Tilley, L. (1998) **Quinoline antimalarials: Mechanisms of action and resistance and prospects for new agents**. *Pharmacology & Therapeutics*. **79**, 55-87

Foth, B. J., Ralph, S. A., Tonkin, C. J., *et al* (2003) **Dissecting apicoplast targeting in the malaria parasite Plasmodium falciparum**, *Science 299*, 705-708.

Franca, T. C. C., Pascutti, P. G., Ramalho, T. C. and Figueroa-Villar, J. D. (2005) **A three-dimensional structure of Plasmodium falciparum serine hydroxymethyltransferase in complex with glycine and 5-formyl-tetrahydrofolate. Homology modeling and molecular dynamics**. *Biophysical Chemistry*. **115**, 1-10

Fraunholz, M. J. (2005) **Systems biology in malaria research**. *Trends in Parasitology*. **21**, 393-395

Frederich, M., Dogne, J. M., Angenot, L. and De Mol, P. (2002) **New trends in anti-malarial agents**. *Current Medicinal Chemistry*. **9**, 1435-1456

Frellsen, J., Moltke, I., Thiim, M., *et al*. (2009) **A Probabilistic Model of RNA Conformational Space**, *Plos Computational Biology 5*.

Galinski, M. R. and Barnwell, J. W. (2009) **Monkey malaria kills four humans**. *Trends in Parasitology*. **25**, 200-204

Gallie, D. R. (1998) **A tale of two termini: A functional interaction between the termini of an mRNA is a prerequisite for efficient translation initiation.** *Gene*. 216, 1-11

Gardiner, D. L., McCarthy, J. S. and Trenholme, K. R. (2005) **Malaria in the post-genomics era: light at the end of the tunnel or just another train?** *Postgraduate Medical Journal*. **81**, 505-509

Gardner, M. J., Hall, N., Fung, E., White, O., *et al* (2002) **Genome sequence of the human malaria parasite Plasmodium falciparum**. *Nature*. **419**, 498-511

Gaspin, C., and Westhof, E. (1995) **an interactive framework for rna secondary structure prediction with a dynamical treatment of constraints,** *Journal of Molecular Biology 254*, 163-174.

Gravel, P., Sanchez, J. C., Walzer, C., *et al* (1995) **human bleed platelet protein map established by 2-dimensional polyacrylamide-gel electrophoresis,** *Electrophoresis 16*, 1152-1159.

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

Griswold, K. E., Mahmood, N. A., Iverson, B. L., and Georgiou, G. (2003) **Effects of codon usage versus putative 5 '-mRNA structure on the expression of Fusarium solani cutinase in the Escherichia coli cytoplasm**, *Protein Expression and Purification 27*, 134-142.

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

Guo, L., Allen, E. M. and Miller, W. A. (2001) **Base-pairing between untranslated regions facilitates translation of uncapped, nonpolyadenylated viral RNA.** *Molecular Cell*. 7, 1103-1109

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

Hajduk, P. J., Huth, J. R. and Tse, C. (2005) **Predicting protein druggability**. *Drug Discovery Today*. **10**, 1675-1682

Haldar, K., Murphy, S. C., Milner, D. A. and Taylor, T. E. (2007) **Malaria: Mechanisms of erythrocytic infection and pathological correlates of severe disease**. *Annual Review of Pathology-Mechanisms of Disease*. **2**, 217-249

Hall, N. and Carlton, J. (2005) **Comparative genomics of malaria parasites**. *Current Opinion in Genetics & Development*. **15**, 609-613

Hannig, G., and Makrides, S. C. (1998) **Strategies for optimizing heterologous protein expression in Escherichia coli**, *Trends in Biotechnology 16*, 54-60.

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

He, M. Y. (2008) **Cell-free protein synthesis: applications in proteomics and biotechnology**. *New Biotechnology*. 25, 126-132

Hempelmann, E., Motta, C., Hughes, R., Ward, S. A. and Bray, P. G. (2003) **Plasmodium falciparum: sacrificing membrane to grow crystals?** *Trends in Parasitology*. **19**, 23-26

Hillier, C. J., Ware, L. A., Barbosa, A., *et al*. (2005) P**rocess development and analysis of liver-stage antigen 1, a preerythrocyte-stage protein-based vaccine for Plasmodium falciparum**, *Infection and Immunity 73*, 2109-2115.

Hogeweg, P., and Hesper, B. (1984) **energy directed folding of rna sequences,** *Nucleic Acids Research 12*, 67-74.

Holder, A. A. (2009) **Malaria Vaccines: Where Next?** *Plos Pathogens*. **5**

Holder, A. A., Lockyer, M. J. and Hardy, G. W. (1988) **a hybrid gene to express protein epitopes from both sporozoite and merozoite surface-antigens of plasmodium-falciparum**. *Parasitology*. **97**, 373-382

Holt, R. A., Subramanian, G. M., *et al* (2002) **The genome sequence of the malaria mosquito Anopheles gambiae**. *Science*. **298**, 129-+

Hyde, J. E. (2008) **Antifolate resistance in Africa and the 164-dollar question**. *Transactions of the Royal Society of Tropical Medicine and Hygiene*. **102**, 301-303

Hyde, J. E., Dittrich, S., Wang, P., *et al*. (2008) **Plasmodium falciparum: a paradigm for alternative folate biosynthesis in diverse microorganisms?** *Trends in Parasitology*. **24**, 502-508

Hyde, J. E. (2007) **Drug-resistant malaria - an insight**. In Febs Journal ed.)^eds.). pp. 4688-4698

Hyde, J. E. (2007) **Targeting purine and pyrimidine metabolism in human apicomplexan parasites**. *Current Drug Targets*. **8**, 31-47

Hyde, J. E. (2002) **Mechanisms of resistance of Plasmodium falciparum to antimalarial drugs**. *Microbes and Infection*. **4**, 165-174

Ichinose, H., Ohye, T., Segawa, M., *et al.* (1995) **Gtp Cyclohydrolase-I Gene in Hereditary Progressive Dystonia with Marked Diurnal Fluctuation**. *Neuroscience Letters*. **196**, 5-8

Ikemura, T., and Aota, S. (1985) **Codon-Choice Patterns of Higher Eukaryotes - the Correlation with the Cellular Transfer-Rna Contents,** *Japanese Journal of Genetics 60*, 594-594.

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

Japrung, D., Leartsakulpanich, U., Chusacultanachai, S. and Yuthavong, Y. (2007) **Conflicting requirements of Plasmodium falciparum dihydrofolate reductase mutations conferring resistance to pyrimethamine-WR99210 combination**. *Antimicrobial Agents and Chemotherapy*. **51**, 4356-4360

Jonasson, P., Liljeqvist, S., Nygren, P. A., and Stahl, S. (2002) **Genetic design for facilitated production and recovery of recombinant proteins in Escherichia coli**, *Biotechnology and Applied Biochemistry 35*, 91-105.

Karzai, A. W., Roche, E. D. and Sauer, R. T. (2000) T**he SsrA-SmpB system for protein tagging, directed degradation and ribosome rescue**. *Nature Structural Biology*. 7, 449-455

Kato, K., Sudo, A., Kobayashi, K., Tohya, Y. and Akashi, H. (2008) **characterization of plasmodium falciparum protein kinase 2**. *American Journal of Tropical Medicine and Hygiene*. 79, 412

Katzen, F., Chang, G. and Kudlicki, W. (2005) T**he past, present and future of cell-free protein synthesis.** *Trends in Biotechnology*. 23, 150-156

Kidgell, C., Volkman, S. K., Daily, J., *et al* (2006) **A systematic map of genetic variation in Plasmodium falciparum**. *PLoS Pathog*. **2**, e57

Kelley, W. L. (1998) **The J-domain family and the recruitment of chaperone power**, *Trends in Biochemical Sciences 23*, 222-227.

Kidgell, C., Volkman, S. K., Daily, J., *et al*. (2006) **A systematic map of genetic variation in Plasmodium falciparum**. *PLoS Pathog*. **2**, e57

Kolinsky, M. A., Crabtree, M. J. and Gross, S. S. (2004) **GFRP tonically inhibits production of tetrahydrobiopterin and nitric oxide in marine endothelial cells**. *Nitric Oxide-Biology and Chemistry*. **11**, 86-86

Kolinsky, M. A. and Gross, S. S. (2004) **The mechanism of potent GTP cyclohydrolase I inhibition by 2,4-diamino-6-hydroxypyrimidine - Requirement of the GTP cyclohydrolase I feedback regulatory protein**. *Journal of Biological Chemistry*. **279**, 40677-40682

Konings, D. A. M., and Hogeweg, P. (1989) **pattern-analysis of rna secondary structure - similarity and consensus of minimal-energy folding**, *Journal of Molecular Biology 207*, 597-614.

Koprowski, H. and Yusibov, V. (2001) **The green revolution: plants as heterologous expression vectors**. *Vaccine*. 19, 2735-2741

Kost, T. A., Condreay, J. P. and Jarvis, D. L. (2005) **Baculovirus as versatile vectors for protein expression in insect and mammalian cells.** *Nature Biotechnology*. 23, 567-575

Kozak, M. (1989) **Circumstances and Mechanisms of Inhibition of Translation by Secondary Structure in Eukaryotic Messenger-Rnas**, *Molecular and Cellular Biology 9*, 5134-5142.

Krungkrai, J., Yuthavong, Y. and Webster, H. K. (1985) **Guanosine Triphosphate Cyclohydrolase in Plasmodium-Falciparum and Other Plasmodium Species**. *Molecular and Biochemical Parasitology*. **17**, 265-276

LaCount, D. J., Vignali, M., Chettier, R., *et al*. (2005) **A protein interaction network of the malaria parasite Plasmodium falciparum**. *Nature*. **438**, 103-107

Laemmli, U. K. (1970) **cleavage of structural proteins during assembly of head of bacteriophage-t4**, *Nature 227*, 680-&.

Lee, J., and Kim, S. H. (2009) **High-throughput T7 LIC vector for introducing C-terminal poly-histidine tags with variable lengths without extra sequences**, *Protein Expression and Purification 63*, 58-61.

Lee, C. S., Salcedo, E., Wang, Q., *et al*. (2001) **Characterization of three genes encoding enzymes of the folate biosynthetic pathway in Plasmodium falciparum**, *Parasitology 122*, 1-13.

LaCount, D., Schoenfeld, L. W. and Fields, S. (2009) **Selection of yeast strains with enhanced expression of Plasmodium falciparum proteins**. *Molecular and Biochemical Parasitology*. 163, 119-122

Laufer, M. K. (2009) **Monitoring antimalarial drug efficacy: current challenges**. *Curr Infect Dis Rep*. **11**, 59-65

Liang, H., Narum, D. L., Fuhrmann, S. R., Luu, T. and Sim, B. K. L. (2000) **A recombinant baculovirus-expressed Plasmodium falciparum receptor-binding domain of erythrocyte binding protein EBA-175 biologically mimics native protein.** *Infection and Immunity*. 68, 3564-3568

Looareesuwan, S., Chulay, J. D., Canfield, C. J., Hutchinson, D. B. A. and Malarone Clinical Trials Study, G. (1999) **Malarone (R) (atovaquone and proguanil hydrochloride): A review of its clinical development for treatment of malaria**. *American Journal of Tropical Medicine and Hygiene*. **60**, 533-541

Lopez, P. J., Marchand, I., Joyce, S. A., and Dreyfus, M. (1999) **The C-terminal half of RNase E, which organizes the Escherichia coli degradosome, participates in mRNA degradation but not rRNA processing in vivo**, *Molecular Microbiology 33*, 188-199.

Madin, K., Sawasaki, T., Ogasawara, T. and Endo, Y. (2000) **A highly efficient and robust cell-free protein synthesis system prepared from wheat embryos: Plants apparently contain a suicide system directed at ribosomes**. *Proceedings of the National Academy of Sciences of the United States of America*. 97, 559-564

Maenpuen, S., Sopitthummakhun, K., Yuthavong, Y., Chaiyen, P. and Leartsakutpanich, U. (2009) **Characterization of Plasmodium falciparum serine hydroxymethyltransferase-A potential antimalarial target**. *Molecular and Biochemical Parasitology*. **168**, 63-73

Marcus, A. and Feeley, J. (1964) **activation of protein synthesis in imbibition phase of seed germination**. *Proceedings of the National Academy of Sciences of the United States of America*. 51, 1075-&

Marsh, K., Forster, D., Waruiru, C., *et al*. (1995) **Indicators of Life-Threatening Malaria in African Children**. *New England Journal of Medicine*. **332**, 1399-1404

Martin, R. E., Marchetti, R. V., Cowan, A. I., *et al*. (2009) **Chloroquine Transport via the Malaria Parasite's Chloroquine Resistance Transporter**. *Science*. **325**, 1680-1682

Martinelli, A., Moreira, R. and Cravo, P. V. L. (2008a) **Malaria combination therapies: Advantages and shortcomings**. *Mini-Reviews in Medicinal Chemistry*. **8**, 201-212

Matambo, T. S., Odunuga, O. O., Boshoff, A., and Blatch, G. L. (2004) **Overproduction, purification, and characterization of the Plasmodium falciparum heat shock protein 70**, *Protein Expression and Purification 33*, 214-222.

Mathews, D. H. (2006) **Predicting RNA secondary structure by free energy minimization**, *Theoretical Chemistry Accounts 116*, 160-168.

McNulty, D. E., Claffee, B. A., Huddleston, M. J., and Kane, J. F. (2003) **Mistranslational errors associated with the rare arginine codon CGG in Escherichia coli**, *Protein Expression and Purification 27*, 365-374.

Mehlin, C., Boni, E., Buckner, F. S., *et al*. (2006) **Heterologous expression of proteins from Plasmodium falciparum: Results from 1000 genes**, *Molecular and Biochemical Parasitology 148*, 144-160.

Moody, A. (2002) **Rapid diagnostic tests for malaria parasites**. *Clinical Microbiology Reviews*. **15**, 66-+

Mordueller, B. (2010) **New medicines for malaria**. *Wiener Klinische Wochenschrift*. **122**, 19-22

Mu, J. B., Awadalla, P., Duan, J. H., *et al* (2007) **Genome-wide variation and identification of vaccine targets in the Plasmodium falciparum genome**. *Nature Genetics*. 39, 126-130

Mudeppa, D. G., Pang, C. K. T., Tsuboi, T., *et al* (2007) **Cell-free production of functional Plasmodium falciparum dihydrofolate reductase-thymidylate synthase**. *Molecular and Biochemical Parasitology*. 151, 216-219

Muller, I. B., Hyde, J. E. and Wrenger, C. (2009) **Vitamin B metabolism in Plasmodium falciparum as a source of drug targets**. *Trends in Parasitology*. **26**, 35-43

Muller, I. B., Das Gupta, R., Luersen, K., *et al*. (2008) **Assessing the polyamine metabolism of Plasmodium falciparum as chemotherapeutic target**. *Molecular and Biochemical Parasitology*. **160**, 1-7

Muller, I. B., Knokel, J., Groves, M. R., *et al*. (2008) **Validation of the vitamin B6 biosynthesis as a potential drug targets in malaria**. *International Journal for Parasitology*. **38**, S43-S43

Nair, S., Miller, B., Barends, M., *et al*. (2008) **Adaptive Copy Number Evolution in Malaria Parasites**. *Plos Genetics*. **4**

Nar, H., Huber, R., Auerbach, G., *et al*. (1995) **Active site topology and reaction mechanism of GTP cyclohydrolase I**. *Proceedings of the National Academy of Sciences of the United States of America*. **92**, 12120-12125

Naughton, J. A., Hughes, R., Bray, P. and Bell, A. (2008) **Accumulation of the antimalarial microtubule inhibitors trifluralin and vinblastine by Plasmodium falciparum**. *Biochemical Pharmacology*. **75**, 1580-1587

Nduati, E., Diriye, A., Ommeh, S., *et al*. (2008) **Effect of folate derivatives on the activity of antifolate drugs used against malaria and cancer**. *Parasitology Research*. **102**, 1227-1234

Neuhoff, V., Arold, N., 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.

Nomura, T., Tazawa, M., Ohtsuki, M., *et al*. (1998) **Enzymes related to catecholamine biosynthesis in Tetrahymena pyriformis. Presence of GTP cyclohydrolase I**. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology*. **120**, 753-760

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

Olliaro, P. (2001) **Mode of action and mechanisms of resistance for antimalarial drugs**. *Pharmacology & Therapeutics*. **89**, 207-219

Olliaro, P. L. and Yuthavong, Y. (1999) **An overview of chemotherapeutic targets for antimalarial drug discovery**. *Pharmacology & Therapeutics*. **81**, 91-110

Olszewski, K. L., Morrisey, J. M., Wilinski, D., *et al*. (2009) **Host-Parasite Interactions Revealed by Plasmodium falciparum Metabolomics**. *Cell Host & Microbe*. **5**, 191-199

Pang, C. K. T., Hunter, J. H., Gujjar, R., *et al*. (2009) **Catalytic and ligand-binding characteristics of Plasmodium falciparum serine hydroxymethyltransferase**. *Molecular and Biochemical Parasitology*. **168**, 74-83

Pasternak, N. D. and Dzikowski, R. (2009) **PfEMP1: An antigen that plays a key role in the pathogenicity and immune evasion of the malaria parasite Plasmodium falciparum**. *International Journal of Biochemistry & Cell Biology*. **41**, 1463-1466

Paulus, M., Haslbeck, M., and Watzele, M. (2004) **RNA stem-loop enhanced expression of previously non-expressible genes**, *Nucleic Acids Research 32*.

Pesce, E. R., Acharya, P., Tatu, U., Nicoll *et al*. (2008) **The Plasmodium falciparum heat shock protein 40, Pfj4, associates with heat shock protein 70 and shows similar heat induction and localisation patterns**, *International Journal of Biochemistry & Cell Biology 40*, 2914-2926.

Peti, W. and Page, R. (2007) **Strategies to maximize heterologous protein expression in Escherichia coli with minimal cost**. *Protein Expression and Purification*. **51**, 1-10

Peters, W. (1970) **therapeutic agents against drug resistant malaria**. *Journal of Parasitology*. **56**, 266

Peterson, D. S., Gao, Y., Asokan, K. and Gaertig, J. (2002) **The circumsporozoite protein of Plasmodium falciparum is expressed and localized to the cell surface in the free-living ciliate Tetrahymena thermophila**. *Molecular and Biochemical Parasitology*. 122, 119-126

Peti, W., and Page, R. (2007) **Strategies to maximize heterologous protein expression in Escherichia coli with minimal cos**t, *Protein Expression and Purification 51*, 1-10.

Plouffe, D., Brinker, A., McNamara, C., *et al*. (2008) **In silico activity profiling reveals the mechanism of action of antimalarials discovered in a high-throughput screen**. *Proceedings of the National Academy of Sciences of the United States of America*. **105**, 9059-9064

Porter, J. R., Stains, C. I., Segal, D. J. and Ghosh, I. (2007) **Split beta-lactamase sensor for the sequence-specific detection of DNA methylation**. *Analytical Chemistry*. **79**, 6702-6708

Pribat, A., Jeanguenin, L., Lara-Nunez, *et al*. (2009) **6-Pyruvoyltetrahydropterin Synthase Paralogs Replace the Folate Synthesis Enzyme Dihydroneopterin Aldolase in Diverse Bacteria**. *Journal of Bacteriology*. **191**, 4158-4165

Porter, J. R., Stains, C. I., Segal, D. J. and Ghosh, I. (2007) **Split beta-lactamase sensor for the sequence-specific detection of DNA methylation.** *Analytical Chemistry*. 79, 6702-6708

Pribat, A., Jeanguenin, L., Lara-Nunez, A., *et al*. (2009) **6-Pyruvoyltetrahydropterin Synthase Paralogs Replace the Folate Synthesis Enzyme Dihydroneopterin Aldolase in Diverse Bacteria**. *Journal of Bacteriology*. **191**, 4158-4165

Ramya, T. N. C., Karmodiya, K., Surolia, A., and Surolia, N. (2007) **15-deoxyspergualin primarily targets the trafficking of apicoplast proteins in Plasmodium falciparum**, *Journal of Biological Chemistry 282*, 6388-6397.

Reynolds, J. M., Takebe, S., Choi, J. Y., *et al* (2008) **Biochemical and genetic analysis of the phosphoethanolamine methyltransferase of the human malaria parasite Plasmodium falciparum**. *Journal of Biological Chemistry*. 283, 7894-7900

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

Riley, E. M., Wagner, G. E., Ofori, M. F., *et al*. (2000) **Lack of association between maternal antibody and protection of African infants from malaria infection**. *Infection and Immunity*. **68**, 5856-5863

Ringe, D., and Petsko, G. A. (2003) **The 'glass transition' in protein dynamics: what it is, why it occurs, and how to exploit it**, *Biophysical Chemistry 105*, 667-680.

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

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

Roest, P. A. M., Roberts, R. G., Sugino, S., Vanommen, G. J. B. and Dendunnen, J. T. (1993) **Protein Truncation Test (Ptt) for Rapid Detection of Translation-Terminating Mutations.** *Human Molecular Genetics*. 2, 1719-1721

Rychlik, W., Rhoads, R. E., and Spencer, W. J. (1990) **determination of optimum annealing temperature for pcr**, *FASEB Journal 4*, A2287.

Sahdev, S., Khattar, S. K., and Saini, K. S. (2008) **Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies**, *Molecular and Cellular Biochemistry 307*, 249-264.

Sahu, N. K., Sahu, S. and Kohli, D. V. (2008) **Novel molecular targets for antimalarial drug development**. *Chemical Biology & Drug Design*. **71**, 287-297

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

Sawasaki, T., Hasegawa, Y., Tsuchimochi, M., *et al* (2002) **A bilayer cell-free protein synthesis system for high-throughput screening of gene products**. *Febs Letters*. 514, 102-105

Schellen.K and Coatney, G. R. (1961) **influence of antimalarial drugs on nucleic acid synthesis in plasmodium gallinaceum and plasmodium berghei**. *Biochemical Pharmacology*. **6**, 143-&

Sheng, Y., Khanam, N., Tsaksis, Y., Shi *et al*. (2008) **Mutagenesis of folylpolyglutamate synthetase indicates that dihydropteroate and tetrahydrofolate bind to the same site**. *Biochemistry*. **47**, 2388-2396

Sherman, I. W. (2009) **Reflections on a century of malaria biochemistry**. *Adv Parasitol*. **67**, 1-402

Sherman, I. W. (1979) **Biochemistry of Plasmodium (Malarial Parasites)**. *Microbiological Reviews*. **43**, 453-495

Shonhai, A., Boshoff, A., and Blatch, G. L. (2007) **The structural and functional diversity of Hsp70 proteins from Plasmodium falciparum**, *Protein Science 16*, 1803-1818.

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

Singh, S., Alam, M. M., Pal-Bhowmick, I., Brzostowski, J. A. and Chitnis, C. E. (2010) **Distinct external signals trigger sequential release of apical organelles during erythrocyte invasion by malaria parasites**. *PLoS Pathog*. **6**, e1000746

Sorensen, H. P., Sperling-Petersen, H. U., and Mortensen, K. K. (2003) **Production of recombinant thermostable proteins expressed in Escherichia coli: completion of protein synthesis is the bottleneck**, *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences 786*, 207-214.

Sorensen, H. P., and Mortensen, K. K. (2005) **Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli**, *Microbial Cell Factories 4*.

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

Spirin, A. S. (2004) **High-throughput cell-free systems for synthesis of functionally active proteins**. *Trends in Biotechnology*. 22, 538-545

Spirin, A. S., Baranov, V. I., Ryabova, L. A., Ovodov, S. Y. and Alakhov, Y. B. (1988) **a continuous cell-free translation system capable of producing polypeptides in high-yield**. *Science*. 242, 1162-1164

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

Suh, W. C., Burkholder, W. F., Lu, C. Z., *et al*. (1998) **Interaction of the Hsp70 molecular chaperone, DnaK, with its cochaperone DnaJ**, *Proceedings of the National Academy of Sciences of the United States of America 95*, 15223-15228.

Tahar, R., de Pecoulas, P. E., Basco, L. K., *et al*. (2001) **Kinetic properties of dihydrofolate reductase from wild-type and mutant Plasmodium vivax expressed in Escherichia coli**. *Molecular and Biochemical Parasitology*. **113**, 241-249

Takai, K., Sawasaki, T. and Endo, Y. (2008) **development of key technologies for high-throughput cell-free protein production with the extract from wheat embryos.** *Structural Genomics, Pt A*. 75, 53-+

Takken, W. and Knols, B. G. J. (2009) **Malaria vector control: current and future strategies**. *Trends in Parasitology*. **25**, 101-104

Takyar, S., Hickerson, R. P., and Noller, H. F. (2005) **mRNA helicase activity of the ribosome**, *Cell 120*, 49-58.

Tangpukdee, N., Duangdee, C., Wilairatana, P. and Krudsood, S. (2009) **Malaria Diagnosis: A Brief Review**. *Korean Journal of Parasitology*. **47**, 93-102

Tayeh, M. A. and Marletta, M. A. (1989) **Macrophage Oxidation of L-Arginine to Nitric-Oxide, Nitrite, and Nitrate - Tetrahydrobiopterin Is Required as a Cofactor**. *Journal of Biological Chemistry*. **264**, 19654-19658

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

Thanaraj, T. A., and Argos, P. (1996) **Protein secondary structural types are differentially coded on messenger RNA**, *Protein Science 5*, 1973-1983.

Thony, B., Auerbach, G. and Blau, N. (2000) **Tetrahydrobiopterin biosynthesis, regeneration and functions**. *Biochemical Journal*. **347**, 1-16

Thony, B. and Blau, N. (1997) **Mutations in the GTP cyclohydrolase I and 6-pyruvoyl-tetrahydropterin synthase genes**. *Human Mutation*. **10**, 11-20

Tolia, N. H., Enemark, E. J., Sim, B. K. L. and Joshua-Tor, L. (2005) **Structural basis for the EBA-175 erythrocyte invasion pathway of the malaria parasite Plasmodium falciparum** (vol 122, pg 183, 2005). *Cell*. 122, 485-485

Tripathi, R. P., Mishra, R. C., Dwivedi, N., *et al*. (2005) **Current status of malaria control**. *Current Medicinal Chemistry*. **12**, 2643-2659

Tsai, C. W., Duggan, P. F., Shimp, R. L., Miller, L. H. and Narum, D. L. (2006) **Overproduction of Pichia pastoris or Plasmodium falciparum protein disulfide isomerase affects expression, folding and O-linked glycosylation of a malaria vaccine candidate expressed in P-pastoris.** *Journal of Biotechnology*. 121, 458-470

Tsuboi, T., Takeo, S., Sawasaki, T., Torii, M. and Endo, Y. (2009) **An Efficient Approach to the Production of Vaccines Against the Malaria Parasite**. *Cell-Free Protein Production: Methods and Protocols*, 73-83

Tsuboi, T., Takeo, S., Iriko, H., *et al*. (2008) **Wheat germ cell-free system-based production of malaria proteins for discovery of novel vaccine candidates.** *Infection and Immunity*. 76, 1702-1708

Tsuboi, T., Takeo, S., Iriko, H., Jin, L., *et al* (2008) **Expression of malaria vaccine candidates using a wheat germ cell-free protein synthesis system without codon optimisation.** *International Journal for Parasitology*. 38, S77-S77

Urscher, M., Przyborski, J. M., Imoto, M. and Deponte, M. (2010) **Distinct subcellular localization in the cytosol and apicoplast, unexpected dimerization and inhibition of Plasmodium falciparum glyoxalases**. *Molecular Microbiology*. **76**, 92-103

van Dooren, G. G., Stimmler, L. M. and McFadden, G. I. (2006) **Metabolic maps and functions of the Plasmodium mitochondrion**. *Fems Microbiology Reviews*. **30**, 596-630Witkowski, B., Berry, A. and Benoit-Vical, F. (2009) **Resistance to antimalarial compounds: Methods and applications**. *Drug Resistance Updates*. **12**, 42-50

Vallejo, L. F. and Rinas, U. (2004) **Optimized procedure for renaturation of recombinant human bone morphogenetic protein-2 at high protein concentration**. *Biotechnology and Bioengineering*. **85**, 601-609

Vedadi, M., Lew, J., Artz, J., *et al*. (2007) **Genome-scale protein expression and structural biology of Plasmodium falciparum and related Apicomplexan organisms**, *Molecular and Biochemical Parasitology 151*, 100-110.

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

Wang, P., Wang, Q., Yang, Y., Coward, J. K., Nzila, A., Sims, P. F. G. and Hyde, J. E. (2010) **Characterisation of the bifunctional dihydrofolate synthase-folylpolyglutamate synthase from Plasmodium falciparum; a potential novel target for antimalarial antifolate inhibition**. *Mol Biochem Parasitol*. **172**, 41-51

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

Wang, H. S., Goffreda, M., and Leustek, T. (1993) **Characteristics of an Hsp70 Homolog Localized in Higher-Plant Chloroplasts That Is Similar to Dnak, the Hsp70 of Prokaryotes**, *Plant Physiology 102*, 843-850.

Wattanarangsan, J., Chusacultanachai, S., Yuvaniyama, J., Kamchonwongpaisan, S. and Yuthavong, Y. (2003) **Effect of N-terminal truncation of Plasmodium falciparum dihydrofolate reductase on dihydrofolate reductase and thymidylate synthase activity**. *Molecular and Biochemical Parasitology*. **126**, 97-102

Waitumbi, J. N., Anyona, S. B., Hunja, C. W., *et al*. (2009) **Impact of RTS,S/AS02(A) and RTS,S/AS01(B) on Genotypes of P. falciparum in Adults Participating in a Malaria Vaccine Clinical Trial**. *Plos One*. **4**

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

Wells, T. N. and Poll, E. M. (2010) **When is enough enough? The need for a robust pipeline of high-quality antimalarials**. *Discov Med*. **9**, 389-398

Witter, K., Cahill, D. J., Werner, T., *et al*. (1996) **Molecular cloning of a cDNA coding for GTP cyclohydrolase I from Dictyostelium discoideum**. *Biochemical Journal*. **319**, 27-32

Wolin, S. L., and Walter, P. (1988) **Ribosome Pausing and Stacking During Translation of a Eukaryotic Messenger-Rna**, *Embo Journal 7*, 3559-3569.

Yonath, A. (1992) **Structural studies on a ribonucleoprotein organelle: The ribosome**, *Nucleic Acids and Molecular Biology*, 246-269.

Young, J. C., Hoogenraad, N. J., and Hartl, F. U. (2003) **Molecular chaperones Hsp90 and Hsp70 deliver preproteins to the mitochondrial import receptor Tom70**, *Cell 112*, 41-50.

Yuthavong, Y. (2002) **Basis for antifolate action and resistance in malaria**. *Microbes and Infection*. **4**, 175-182.

Zou, L. L., Miles, A. P., Wang, J. and Stowers, A. W. (2003) **Expression of malaria transmission-blocking vaccine antigen Pfs25 in Pichia pastoris for use in human clinical trials.** *Vaccine*. 21, 1650-165