MOLECULAR STUDIES ON THE ENZYMES OF FOLATE BIOSYNTHESIS IN THE MALARIA PARASITE *PLASMODIUM*

A thesis submitted to
THE UNIVERSITY OF MANCHESTER INSTITUTE OF SCIENCE AND TECHNOLOGY
for the degree of
Doctor of Philosophy

by

James Alexander Lowther

FEBRUARY 2002

Department of Biomolecular Sciences
UMIST

Declaration

No portion of this thesis has been submitted in support of an application for another degree or qualification of this or any other university, or other institution of learning.

All work described in this thesis was carried out in accordance with COSHH regulations and guidelines.

ABSTRACT

The disease malaria, caused by protozoan parasites of the genus *Plasmodium*, and spread by mosquitoes, still represents one of the major causes of mortality and morbidity in the developing world. As yet, there is no vaccine against the disease available, and parasites resistant to the commonly used antimalarial drugs are increasingly common. As a result, there is a concerted worldwide search for new drugs and drug targets in the malaria parasite. One area of parasite biochemistry already successfully targeted is the biosynthesis and metabolism of folate cofactors, which act as single-carbon unit donors and acceptors in a number of cellular processes, including the biosynthesis of several nucleotides and amino acids.

Pf gch-1, the gene apparently encoding GTP cyclohydrolase I (GTCase), which catalyses the first step in folate biosynthesis, has been isolated from the genome of the human malaria parasite *P.falciparum*. Here, investigations into expression of this gene in a number of heterologous systems are presented. *Pf gch-1*, and fragments thereof, have been cloned into two different vectors designed for expression of genes in *Escherichia coli*. Transformation of these constructs into *E.coli* cells has resulted in the expression and purification of a number of fusion proteins incorporating C-terminal portions of the putative *P.falciparum* GTCase, suitable for use in biochemical analyses of the enzyme function. Inoculation of a chicken with one of these fusion proteins has resulted in the isolation of antibodies which show encouraging levels of specificity and sensitivity against the inoculating antigen; these antibodies should prove invaluable in further studies of the parasite GTCase in, for example, intracellular localisation studies and purification of the native protein from parasite cells.

Previous attempts to use this gene to rescue yeast cells deficient in GTCase have failed for reasons believed to involve truncation of the mRNA transcripts produced by the gene, as a result of recognition of certain sequence motifs within the gene by the yeast's cellular mRNA processing machinery. Here the removal of some of these motifs from the gene by site-directed mutagenesis is described. Unfortunately these modifications proved unable to allow either full or partial complementation of the yeast deficiency by clones of *Pf gch-1*, and C-terminal fragments of the same, in two different yeast expression vectors. This result prompted consideration of other host cell types for expression of *Pf gch-1*. Preliminary investigations into expression of this gene in human embryonic kidney cells are described here.

Analysis of the expression of *Pf gch-1* mRNA through the course of the intraerythrocytic (blood stage) cycle of the malaria parasite has indicated that expression is highest in the mature trophozoite phase of the cycle, and lowest in the schizont stage. This is broadly consistent with previous results regarding variation of GTCase activity and DNA synthesis in the intraerythrocytic cycle.

We have exploited the rapidly accumulating body of information provided by the genome sequencing projects of several *Plasmodium* spp. to identify homologues of *Pf gch-1* in the rodent parasite *P.yoelii* and the simian parasite *P.knowlesi*. Analysis of the sequences of the polypeptides encoded by these genes has revealed that, while they are very closely similar in a C-terminal 'core' region (this portion of the gene is highly conserved between widely different taxa), these genes are markedly and, in some regions, unusually divergent. The sequencing of *Pf gch-1* homologues from two separate subspecies of *P.yoelii* has enabled interesting comparisons of the population structures of this species and *P.falciparum* to be drawn; whilst no differences in the coding sequence of *Pf gch-1* were found in geographically diverse isolates of *P.falciparum*, a considerable number of length polymorphisms and point mutations were found between the two *P.yoelii* sequences. This lends support to the supposition, made elsewhere, that all extant strains of *P.falciparum* derive from a relatively recent common ancestor.

Finally, described within are attempts to isolate the gene encoding dihydroneopterin aldolase (DHNA), another enzyme of folate biosynthesis, from the genome of *P.falciparum*.

ACKNOWLEDGEMENTS

I would firstly like to thank my supervisor John Hyde for giving me the opportunity to work in the UMIST Molecular Parasitology research group, as well as for his guidance and encouragement throughout the project. I am also particularly grateful to all the other members of the parasitology group, both past and present; Martin Read, Wang Ping, Wang Qi, Niroshini Nirmalan, Barbara Verrall, Enrique Salcedo, Paul Sims, Tanya Aspinall, Damian Marlee, Chung-Shinn Lee, Andrew Blagborough and Karen Read for their kind and invaluable help and support. My acknowledgements also go to all the staff and students of the Dept. of Biomolecular Sciences, UMIST, particularly Duncan Smith of the Leukaemia Research Fund laboratory, who carried out the trypsinisation and MALDI-TOF mass spectroscopy of GST-Cgch on my behalf, Dr. Stuart Wilson and Louise Roaden who lent their expertise on animal cell expression systems and carried out the transfections of human cells respectively, and Dr. Simon Hubbard and Dr. Khushwant Sidhu for their assistance in bioinformatics.

Finally I would like to offer my eternal gratitude to all my friends and family, whose love and companionship has sustained me through the duration of this project, especially Carol, to whom I dedicate this thesis.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

Å	Angstrom(s)
aa	amino acid(s)
AMPS	ammonium peroxodisulphate
AP	alkaline phosphatase
BCIP	5-bromo-4-chloro-3-indolyl phosphate
BH_4	tetrahydrobiopterin
bp	base pairs
cDNA	complementary DNA
CH_2FH_4	5,10-methylene tetrahydrofolate
CH_3FH_4	5-methyl tetrahydrofolate
Ci	Curie(s)
cpm	counts per minute
CSP	circumsporozoite protein
CSPD	disodium 3-(4-methoxyspiro{1,2-dioxetane-3,2'-(5'-chloro)tricyclodecan}-4-yl) phenyl phosphate

C_t	Threshold cycle
Da	Dalton(s)
dATP	deoxyadenosine 5'-triphosphate
dCTP	deoxycytidine 5'-triphosphate
DDT	dichlorodiphenyl trichloroethane
DEPC	diethyl pyrocarbonate
dGTP	deoxyguanosine 5'-triphosphate
DHFR	dihydrofolate reductase
DHFS	dihydrofolate synthetase
DHNA	dihydroneopterin aldolase
DHNTE	dihydroneopterin triphosphate epimerase
DHPS	dihydropteroate synthase
DIG	digoxigenin
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxynucleoside 5'-triphosphate
dTMP	deoxythymidine 5'-monophosphate
dTTP	deoxythymidine 5'-triphosphate
dUMP	deoxyuridine 5'-monophosphate
EDTA	ethylenediamine tetraacetic acid
EST	expressed sequence tag
FH_2	dihydrofolate
FH_4	tetrahydrofolate
FPGS	folyl polyglutamate synthase
$-Glu_5$	pentaglutamyl (of folates)
$-Glu_n$	polyglutamyl (of folates)
GST	glutathione S-transferase
GTCase	GTP cyclohydrolase I (see note at end of section)
GTP	guanosine 5'-triphosphate
Hr	hour(s)
IgG	immunoglobulin G
IgY	immunoglobulin Y
IPTG	isopropyl ß-D-thiogalactopyranoside

Kb	kilobasepair(s) (of DNA); kilobase(s) (of RNA)
kDa	kilodalton(s)
k_I	Michaelis inhibition constant
LB	Luria-Bertani medium
LSU	large subunit
MALDI-TOF	matrix-assisted laser desorption/ionisation time of flight
Mb	megabasepair(s)
MBP	maltose binding protein
min	minute(s)
MOPS	3-[N-morpholino]propanesulphonic acid
mRNA	messenger RNA
MS	methionine synthase
MTHFR	5,10-methylene tetrahydrofolate reductase
m/z	mass/charge ratio
NBT	nitro blue tetrazolium
nt	nucleotide(s)
OD_{600}	optical density at 600nm
ORF	open reading frame
pABA	*para*-aminobenzoic acid
pABG	*para*-aminobenzoylglutamate
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
Pf	pertaining to *Plasmodium falciparum*
PfTRAP	*Pf* thrombospondin related adhesion protein
pI	isoelectric point
PPPK	6-hydroxymethyl-dihydropterin pyrophosphokinase
pRBC	parasitised RBC
psi	pounds per square inch pressure
PTPS	6-pyruvoyl tetrahydropterin synthase
RBC	red blood cell
RNA	ribonucleic acid
RNase	ribonuclease

rpm	revolutions per minute
rRNA	ribosomal RNA
RT-PCR	reverse transcriptase PCR
sdH_2O	sterile distilled water
SDS	sodium dodecyl sulphate
SHMT	serine hydroxymethyl transferase
Spp.	species (as plural)
SSC	standard sodium citrate
SSU	small subunit
TBE	Tris/boric acid/EDTA buffer
TE	Tris/EDTA buffer
TEMED	tetramethylethylene diamine
T-fold	tunnelling fold
T_m	melting temperature
tRNA	transfer RNA
TS	thymidylate synthase
UTR	untranslated region
UV	ultraviolet
v/v	volume per volume
w/	with
WHO	World Health Organisation
w/v	weight per volume
YEPD	yeast extract/peptone/dextrose medium
YNBD	yeast nitrogen base/dextrose medium

Note: GTP cyclohydrolase I (GTCase) is often referred to in the literature by other abbreviations, e.g. GTP-CH-1, CYH etc.

CHAPTER 1

INTRODUCTION

1.1 : Malaria

Malaria is one of the world's oldest recorded diseases, with references to the illness in Sumerian and Chinese literature dating back as far as 3000 BC (Desowitz, 1991). The disease has played a major part in human history. The courses of countless battles, from the defeat of Athens by the Spartans at Syracuse in 413 BC to the Vietnam War in modern times, have been affected by battlefield malaria casualties. Today, malaria remains the world's most important tropical disease. It is endemic in more than 100 countries and territories, mostly in tropical and subtropical regions of the globe (Figure 1.1), and threatens around 40% of the world's population. Subsaharan Africa accounts for the vast majority (>90%) of the 300-500 million clinical cases per annum with 60% of the remaining cases concentrated in 6 countries; India, Brazil, Colombia, Sri Lanka, Vietnam and the Solomon Islands. Of the 1.5-2.7 million deaths per annum

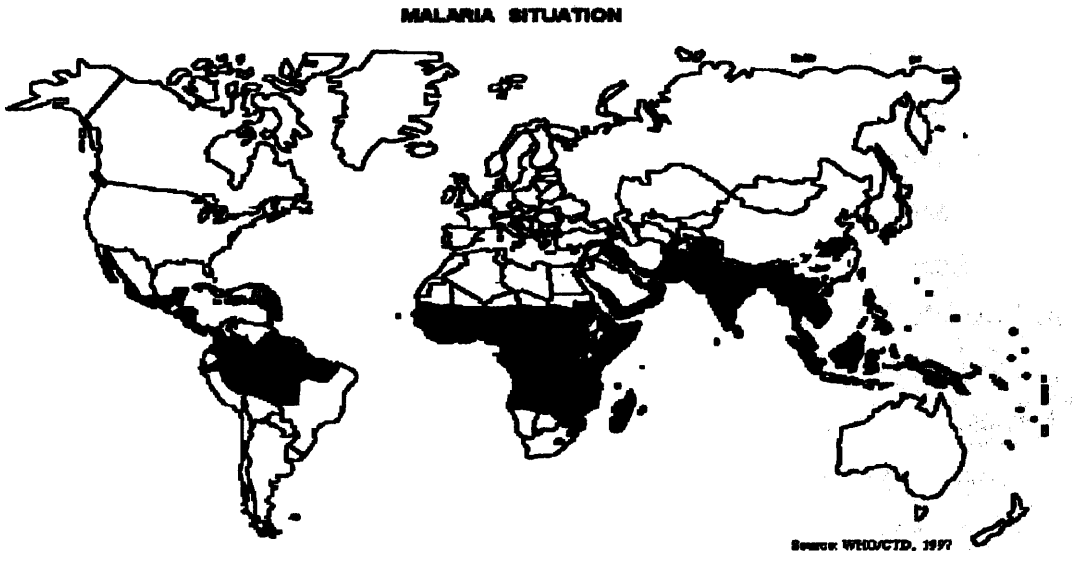


Figure 1.1 The global distribution of malaria
Reproduced from the website of the WHO, 1998.

from the disease, the majority are found amongst African children of under 5 years in age (Butler *et al.*, 1997).

Although malaria has been eradicated from Europe and North America, the World Health Organisation was forced to abandon its stated aim of global eradication in 1969, in the face of emerging problems with drug and insecticide resistance.

1.2 : The malaria parasite

The causative agents of malaria are dixenous protozoan parasites of the genus *Plasmodium* (order Eucoccidiida, class Sporozoea, phylum Apicomplexa). Of about 170 *Plasmodium* species described, which infect birds, mammals and reptiles in a species-specific manner, 4 are transmitted to humans:

P.falciparum, P.vivax, P.malariae and *P.ovale.*

P.vivax, P.malariae and *P.ovale* are often grouped together in the subgenus *Plasmodium* with many other primate malaria parasites. *P.falciparum* however is normally placed in the subgenus *Laverania*, which it shares only with *P.reichenowi*, a parasite of chimpanzees. These 2 species appear to be most closely related to avian parasites including *P.gallinaceum* (Escalante *et al.*, 1998).

The distribution of each of the four human malaria parasites is cosmopolitan, however local distribution varies greatly, and *P.ovale* is predominantly found in tropical West Africa. *P.falciparum* and *P.vivax* are together responsible for over 90% of infections. *P.falciparum* is responsible for the most serious form of the disease and most of the fatalities, although *P.vivax* causes severe and debilitating attacks.

The ability to propagate *P.falciparum* continuously in culture was established by Trager & Jensen in 1976. As a result, this is and has been the most widely studied of the human malaria parasites. However, the host specificity of *Plasmodium* spp. means that animal models for study of the human malaria parasites are extremely limited. Much work has therefore also been concentrated on the rodent malaria parasites, including *P.chabaudi, P.vinckei, P.berghei* and *P.yoelii.*

Malaria parasites are transmitted to the vertebrate host principally by female mosquitoes of the genus *Anopheles*. Around 70 *Anopheles* spp. are able to transmit parasites to humans, with species occupying different geographical ranges and habitat niches. Mosquitoes of the *A.gambiae* complex are responsible for much of the transmission of *P.falciparum* in tropical Africa (Collins & Besansky, 1994).

1.3 : Life cycle of the malaria parasite

The life cycle of all species of malaria parasite consists of an asexual replicative phase in the vertebrate host and a sexual stage in the mosquito vector (Figure 1.2).

1.3.1 : Life cycle in the vertebrate host

Several tens to several hundreds of *Plasmodium* sporozoites are injected into the bloodstream of the host when an infected mosquito takes a blood meal. The sporozoites then migrate to the liver where they penetrate the hepatocytes. Here the parasite nucleus undergoes a multiplication process called schizogony generating up to 30,000 merozoites. These leave the liver and invade the red blood cells (RBCs). In the case of two of the human malaria parasites, *P.vivax* and *P.ovale,* hypnozoite forms of the parasite may remain quiescent in the liver for prolonged periods before undergoing secondary schizogony. This phenomenon is responsible for the recurrence of the disease in previously infected individuals (Cogswell, 1992).

Having invaded the RBC, the parasite occupies a membranous compartment known as the parasitophorous vacuole. Here it develops into the trophozoite, which feeds on the host cell cytoplasm. Early trophozoite forms are known as ring-stage parasites, owing to the characteristic 'signet-ring' appearance of their nuclei when stained and viewed under a microscope. Mature trophozoites undergo another round of schizogony, yielding a multinucleate schizont, which subsequently is divided into 8-32 new merozoites, released when the RBC membrane ruptures. These merozoites then invade further uninfected RBCs, initiating a fresh cycle of feeding and schizogony, or alternatively develop within the RBC into male and female gametocytes.

1.3.2 : Life cycle in the mosquito host

A mosquito taking a blood meal from a host at the right stage of infection will ingest a mixture of merozoites and gametocytes. The latter mature into gametes within the mosquito gut. The microgamete (male) penetrates the macrogamete (female) to produce a fertilised zygote. This cell differentiates to form an oökinete, and undergoes a round of meiotic division to restore the haploid genome. If the mosquito ingests gametocytes of more than one genotype, either from separate hosts or a single multiply infected individual, recombination of parasite alleles may occur.

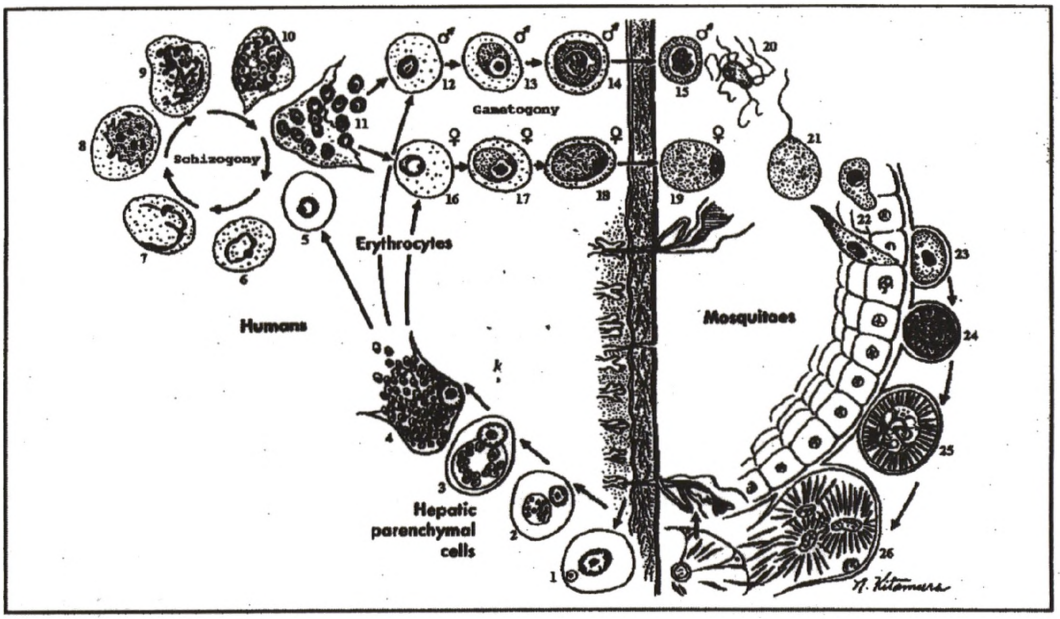


Figure 1.2 The life cycle of the malaria parasite
1-4, asexual cycle in liver; 5-11, asexual cycle in RBCs; 12-15, microgametocyte development; 16-19, macrogametocyte development; 20, production of microgametes; 21, fertilisation of macrogamete; 22, oökinete penetrates wall of mosquito stomach; 23-25, production of sporozoites within oöcyst; 26, release of sporozoites, most entering salivary glands of the mosquito. Reproduced from Garcia & Bruckner (1993).

The oökinete embeds in the wall of the midgut and develops into an oöcyst, in which thousands of sporozoites are produced by a division process analogous to schizogony, called sporogony. These are released into the haemocoel of the mosquito, from where they can migrate to the salivary glands.

1.4 : Clinical symptoms of the disease

An infection with malaria parasites gives rise within two weeks to periodic bouts of chills, fevers and sweats. The period of the fever is very regular, 72 hours in *P.malariae* infections, 48 hours in the cases of the other three species, and the fever coincides with the synchronised liberation of merozoites from the RBCs (1.3.1).

Lysis of RBCs can lead to anaemia, while the release of inflammatory cytokines, for example tumour necrosis factor (TNF), by the host immune system may also play a role in the pathology of the disease. The increased pathogenicity of *P.falciparum* with respect to the other species is due partly to the high parasitaemias (percentage of parasitised RBCs (pRBCs)) possible with this species, up to 100 times greater than with the others (Bruce-Chwatt, 1985), as well as complications arising from pRBC cytoadherence, a set of phenomena only observed in *P.falciparum*. The adherence of

pRBCs to other cells is mediated through the appearance on or at the surface of the pRBC of parasite encoded proteins including knob-associated histidine-rich protein (KAHRP) and *P.falciparum* erythrocyte membrane protein 1 (PfEMP1) (reviewed by Cooke *et al.*, 2000). The pRBC can bind to other pRBCs (auto-agglutination), non-parasitised RBCs (rosetting) or endothelial cells in the microvasculature (sequestration). Sequestration of or blockages caused by auto-agglutinated or rosetted pRBCs in the capillaries of the brain (cerebral malaria) or other organs can cause localised hypoxia and organ dysfunction, and ultimately death.

In areas of endemic malaria transmission, survival to the age of 5 normally guarantees some degree of naturally acquired partial immunity to malaria, whereby the severity of symptoms associated with malaria infection decreases. A number of host genetic factors are also associated to some degree with immunity from the disease, the best known being the 'sickle-cell' allele of haemoglobin S, which causes sickle cell anaemia in homozygous individuals, but confers partial immunity to malaria in both homozygous and heterozygous individuals.

1.5 : Control of malaria

Currently implemented strategies for control of the disease concentrate on either control of the insect vector or chemotherapy. A great deal of effort has also been expended in the development of experimental antimalarial vaccines.

1.5.1 : Vector control

Attempts to limit transmission of malaria to humans from the mosquito have taken a variety of forms. Insect numbers have been reduced through the spraying of insecticides, most notably DDT and its successors such as dieldrin, and the use of bednets impregnated with a pyrethroid insecticide (usually permethrin) in order to protect individuals from mosquito bites is also widespread, and has been shown to substantially decrease morbidity and mortality from malaria (Greenwood, 1997). However problems with insecticide resistance and environmental considerations have reduced the efficacy of this approach.

The drainage of marshy habitats favoured as mosquito breeding sites or the release of larvivorous fish into such areas has successfully reduced insect numbers in many areas, particularly in Europe. The success of this approach is very much dependent on the habitat requirements of the target vector species however.

Genetic manipulation of mosquito populations through the release of sterile male insects, or insects unable to support parasite growth has also been considered (Collins and Besansky, 1994). It may also be desirable to introduce foreign genes, for example encoding products capable of attacking the insect stage parasite, into wild mosquito populations. The development of a protocol for stable transformation of germlines of *Anopheles stephensi* using the *Minos* transposable element from the Mediterranean fruitfly *Ceraitis capitata* (Catteruccia *et al,* 2000) will hopefully facilitate this approach.

1.5.2 : Chemotherapy

There are several classes of antimalarial drugs in general use, targeting a variety of stages of the parasite life cycle. Drugs can be used either as curative or preventative (prophylactic) measures.

4-methanolquinolines

These drugs, effective against the blood stages of the parasite, are derivatives of quinine, extracted from the bark of the Cinchona tree, which has been used in the treatment of malaria for centuries. Quinine remains the drug of choice for the treatment of severe, multiply resistant, *falciparum* malaria. Newer derivatives of quinine include mefloquine and halofantrine. Mefloquine is a highly potent long-acting compound used extensively to treat multiple drug-resistant parasites in Southeast Asia and also as a prophylactic. This drug can have severe side effects however, and is not recommended for people with a history of epilepsy or psychiatric disorders.

4-aminoquinolines

Chloroquine, the principal antimalarial in this class, is a blood stage drug, generally believed to inhibit the polymerisation of toxic haem moieties produced by the breakdown of haemoglobin (reviewed by Ginsburg *et al.*, 1999) (1.7.2). Chloroquine is cheap, highly effective against sensitive parasites and produces few side effects. As a result, it has historically been the most extensively used synthetic antimalarial agent. Resistance to this drug in *P.falciparum* is now widespread however (1.5.3), limiting its role in the treatment of infections of this species. Chloroquine resistance in *P.vivax* is still comparatively localised however.

Napthoquinones

Atovaquone is the major drug in this class. Its mode of action involves inhibition of the electron transport chain in the mitochondrion. Atovaquone is normally given in combination with proguanil, a pro-drug that is metabolised into the antifolate cycloguanil. The marked synergy between these drugs is not believed to be related to the antifolate properties of cycloguanil; rather, proguanil itself is thought to potentiate the action of atovaquone (Srivastava & Vaidya, 1999).

Antifolates

Antifolate drugs target enzymes involved in the biosynthesis of folates (1.8.1). Sulphonamides (including sulfadoxine) and sulphones act as inhibitors of dihydropteroate synthase (DHPS). Pyrimethamine and cycloguanil (administered as the pro-drug proguanil) are known to selectively inhibit plasmodial dihydrofolate reductase (DHFR). Recent evidence suggests, however, that pyrimethamine has a second mode of action, involved in the utilisation of exogenous folate (Wang *et al.*, 1999) (1.8.2).

Antifolates are used in combinations such as Fansidar, (sulfadoxine and pyrimethamine), where the two drugs exhibit marked synergy. Fansidar is currently used as a first-line response to chloroquine-resistant parasites in Africa.

Artemisinin and derivatives

The sesquiterpene lactone artemisinin is the active ingredient in the traditional chinese medicine qingaosu, derived from the plant *Artemisia annua* (Klayman, 1985). The drug and its synthetic derivatives, including artemether and artesunate, which have only been introduced into clinical use comparatively recently, are the most rapidly acting of all antimalarial agents, and are active against a broad spectrum of parasite stages (Alin and Bjorkman, 1994). These drugs act to interfere with parasite membranes, disrupt protein synthesis and effect changes in organelle morphology.

1.5.3 : Drug resistance

The appearance of parasite strains resistant to the commonly used antimalarials has represented a major problem for malaria control in the second half of the 20th century. Resistance to chloroquine was first observed in the 1950s and chloroquine resistant parasites are now found throughout the transmission range, particularly in Africa (Wernsdorfer and Payne, 1991). Chloroquine resistance in *P.falciparum* is

strongly linked to mutations in the gene encoding *P.falciparum* chloroquine resistance transporter (PfCRT; Fidock *et al.,* 2000) a transmembrane protein localised to the membrane of the parasite digestive vacuole (see 1.7.2) possibly involved in the transport of chloroquine across this membrane. Resistance to chloroquine and other quinoline drugs (as well as resistance to artemisinin) is also possibly mediated by mutations in *Pfmdr1*, the gene encoding p-glycoprotein homologue 1 (Pgh-1; Reed *et al.*, 2000), a transmembrane pump homologous to human P-glycoprotein, responsible for the multi-drug resistant phenotype of some tumour cells.

Resistance to the antifolate drugs is now spreading from foci in Southeast Asia and South America. Pyrimethamine resistance has long been associated with the accumulation of missense point mutations in the bifunctional parasite *dhfr-ts* gene, corresponding to changes at amino acid positions 51, 59, 108 and 164 (reviewed by Hyde, 1990), the DHFR enzyme being the site of action of this drug (1.5.3). The correlation between these mutations and pyrimethamine resistance has been confirmed by biochemical analysis of recombinant mutant enzymes (Sirawaraporn *et al.,* 1997). Similarly, missense mutations in the bifunctional parasite *pppk-dhps* gene, corresponding to changes at amino acid positions 436, 437, 540, 581 and 613, have been implicated in sulphonamide resistance (Triglia *et al.*, 1998). Also, strains which are wild type for *dhps* but which are able to utilise exogenous folate (1.8.2) are naturally resistant to sulphonamides (Wang *et al.,* 1997). Interestingly, this sulphonamide resistance appears to be ameliorated by the addition of sublethal doses of pyrimethamine, even in parasites which are resistant to pyrimethamine alone (Wang *et al.,* 1999).

1.5.4 : Vaccine development

The development of an antimalarial vaccine has been hampered by several fundamental problems, including the complex nature of the parasite life cycle, and the high degree of antigenic diversity between parasite strains (White & Ho, 1992). Nevertheless, it has long been established that irradiated *Plasmodium* sporozoites are capable of inducing protective immunity against malaria in humans (Clyde *et al.,* 1973), and although isolation of sufficient numbers of these cells for large-scale vaccine production is in practical terms impossible, this observation provides support for the belief than an effective malaria vaccine is a realistic prospect. Much of the research into vaccine candidates has focussed on producing recombinant protein or synthetic peptides

incorporating cell surface antigens which have been found to play important roles in mediating protective immunity, through both B and T cell responses, in areas of endemic malaria transmission (1.4). These antigens have included circumsporozoite protein (CSP), from the preerythrocytic (liver) stages of the organism (Stoute *et al.*, 1997), plus merozoite surface protein 1 (MSP-1; Blackman *et al.*, 1994) and apical membrane antigen 1 (AMA-1; Crewther *et al.*, 1996) from the blood stages of the life cycle. Many such antigens have proved to be highly polymorphic; invariable regions of these proteins, for example the multiple NANP repeat in CSP, are therefore often incorporated into such vaccines. Many modern vaccines incorporate antigens from several stages of the parasite life cycle. The first multicomponent vaccine of this type was SPf66, consisting of a synthetic polypeptide incorporating three peptides from blood stage antigens and one peptide from the CSP (Patarroyo *et al.*, 1987), while more recently NYVAC-Pf7, a highly attenuated vaccinia virus with 7 *P.falciparum* genes encoding proteins from all parts of the life cycle has shown some efficacy in field trials (Ockenhouse *et al.*, 1998).

Much research effort has also been expended on the design and testing of DNA based vaccine systems, whereby immunisation is attempted through inoculation with plasmid DNA carrying parasite antigen genes, for example, inoculation of rodents with a DNA vaccine encoding the *P.yoelii* CSP showed some success in inducing an immune response (Sedegah *et al.*, 1994). Increasingly many researchers are modifying this approach to incorporate heterologous prime-boost regimens. Schneider *et al.* (2001) have demonstrated that vaccinating chimpanzees using plasmid DNA expressing the preerythrocytic antigens thrombospondin related adhesion protein (PfTRAP) and liver stage specific antigen-1 (PfLSA-1) of *P. falciparum*, followed 6 weeks later by recombinant modified vaccinia virus expressing PfTRAP produces a stronger immune response than when DNA is used for both prime and boost inoculations.

1.6 : Molecular biology of the malaria parasite

P.falciparum possesses a haploid genome of $3x10^7$ bp (Lanzer *et al.*, 1993), divided between 14 chromosomes ranging from 600 kb to 3.5 Mb (Wellems *et al.*, 1987). Marked differences in chromosome size and morphology are noted between different strains of the species however. The parasite also possesses two extrachromosomal DNAs; a concatenated 6kb fragment with mitochondrial type features, encoding three mRNAs, plus large subunit (LSU) and small subunit (SSU)

rRNAs (Gray *et al.*, 1999), and a 35kb circular element, similar to those found in plastids, encoding 23 mRNAs (mostly for ribosomal proteins), LSU and SSU rRNAs and 25 tRNA species (Wilson *et al.*, 1996).

All three *P.falciparum* genomes are highly A+T rich. The nuclear genome has an A+T content of 82%; 74% in coding regions and 86% in flanking regions (Weber, 1987). This A+T richness is shared with several other *Plasmodium* spp. including *P.yoelii* and *P.gallinaceum*, most primate-specific parasites show a far less marked A+T bias however; for example, the *P.vivax* genome has an A+T content of around 55%. The mRNA synonymous strands of *P.falciparum* genes are purine rich (Hyde & Sims, 1987). This factor, combined with the A+T bias of the genome, results in highly biased codon usage. The majority of parasite genes have none or only a few introns (57% of putative ORFs on chromosome 2 apparently contain no introns, for example (Gardner *et al.*, 1998)), and those that are present are generally short, averaging 200bp.

P.falciparum is one of a number of parasitic species to be currently the subject of a genome sequencing project. Complete sequences of chromosome 2 (Gardner *et al.*, 1998) and chromosome 3 (Bowman *et al.*, 1999) have already been published and sequences of all the other chromosomes are nearing completion. The information generated by this initiative should revolutionise the study of the organism. Already, novel biochemical pathways, and hence drug targets, in the parasite have been postulated. For example, inhibitors of 1-deoxy-D-xylulose 5-phosphate (DOXP) isoreductase, an enzyme involved in one such pathway, the nonmevalonate pathway of isoprenoid biosynthesis, have been shown to suppress *P.falciparum* growth *in vitro* (Jomaa *et al.*, 1999). It is hoped that in the future, genome-wide analysis of stage-specific mRNA transcription and protein expression through post-genomic technologies such as DNA microarrays and 2D gel electrophoresis/mass spectroscopy proteome analysis will reveal stage specific antigens and T-cell epitopes which can be used to further the development of antimalarial vaccines (1.5.4, reviewed by Carucci, 2000). Five other *Plasmodium* spp. are also the subject of smaller-scale genome sequencing projects; *P.vivax*, the rodent parasites *P.chabaudi, P.berghei* and *P.yoelii*, and *P.knowlesi*, a simian parasite, naturally infecting macaques (*Macaca* spp.) and langurs (*Presbytis* spp.).

1.7 : Biochemistry of the malaria parasite

1.7.1 : Energy production

Blood stages of *P.falciparum* lack carbohydrate reserves, utilising blood glucose as the primary energy source. The rate of glycolysis in pRBCs is markedly increased (Roth *et al.*, 1982). However, the parasite does not have a complete citric acid cycle (Opperdoes, 1995). Energy production is therefore almost entirely anaerobic, with lactic acid as the end product. Lactic-acidosis, caused by lactic acid accumulation, is an important and potentially fatal complication of severe malaria (Krishna *et al.*, 1994), and build-up of this chemical is also a major irritation to experimenters working with *in vitro* cultures of *P.falciparum* (Zolg *et al.*, 1984).

1.7.2 : Protein production

P.falciparum possesses a limited ability to synthesise certain amino acids, including alanine, aspartic acid and methionine. It is also able to take up other amino acids, such as isoleucine, preformed from the host plasma. These supplies appear to be insufficient however, and therefore the degradation of haemoglobin following endocytosis of RBC cytoplasm provides the parasite with most of its amino acid requirements (Rosenthal & Meshnick, 1996). Various proteases in the digestive vacuole of *P.falciparum* have been implicated in this process (Luker *et al.*, 1996; Gamboa de Dominguez & Rosenthal, 1996). Two closely related aspartic proteases, plasmepsins I and II, are responsible for initial breakdown of haemoglobin into large peptide fragments and the release of haem, while the cysteine protease falcipain subsequently breaks these peptides down into oligopeptides of around 8 amino acids in length, which are thought to be exported to the parasite cytoplasm and converted there into single amino acids by exopeptidases (reviewed by Francis *et al.*, 1997). Potentially toxic haem moieties produced by haemoglobin breakdown are polymerised to haemozoin (Dorn *et al.*, 1995). Interference in this process is believed to be the major mode of action of the quinoline family of antimalarial drugs (1.5.2).

1.7.3 : Nucleic acid production

Malaria parasites require massive quantities of purines and pyrimidines for nucleic acid synthesis during the exponential multiplication phases of their life cycles

(1.3). These requirements are generally met by purine salvage and *de novo* pyrimidine synthesis.

The major precursors for purine conversion are adenosine, inosine and hypoxanthine, hypoxanthine being the most favoured *in vitro*. These species (predominantly adenosine) are salvaged from the RBC cytoplasm.

The precursors of *de novo* pyrimidine biosynthesis are carbamyl phosphate and aspartate. These are converted to dUMP in a series of reactions catalysed by a set of parasite enzymes, one of which, dihydroorotate synthase, has been proposed as the target of a number of antimalarial drugs including atovaquone (1.5.2). The conversion of dUMP to dTMP, catalysed by thymidylate synthase (TS), involves the transfer of a $-CH_2$ group and requires the presence of folate cofactors (1.8) the synthesis and metabolism of which is another major drug target (1.5.2).

1.8 : Folate metabolism

Reduced folate cofactors serve as donors and acceptors of one-carbon units in a variety of cellular processes, including the biosynthesis of purines, dTMP, glycine, serine, methionine, pantothenate and N-formylmethionyl tRNA. In the malaria parasite, folate cofactors have been implicated in the production of dTMP and amino acids only.

1.8.1 : Folate biosynthesis

The folate requirements of mammals are met by the presence in the diet of the vitamin folic acid, particularly rich sources of which include leafy green vegetables and liver. Folic acid is reduced via dihydrofolate (FH_2) to tetrahydrofolate (FH_4), the active cofactor, by mammalian DHFR.

Many microorganisms and plants possess a folate biosynthetic capability however, whereby FH_2 is formed from the precursors GTP, *para*-aminobenzoic acid (pABA) and glutamate.

The dihydrofolate biosynthetic pathway (Figure 1.3) was elucidated through the study of cell extracts and purified enzymes from a variety of bacterial species, particularly *E.coli* (reviewed by Brown & Williamson, 1987).

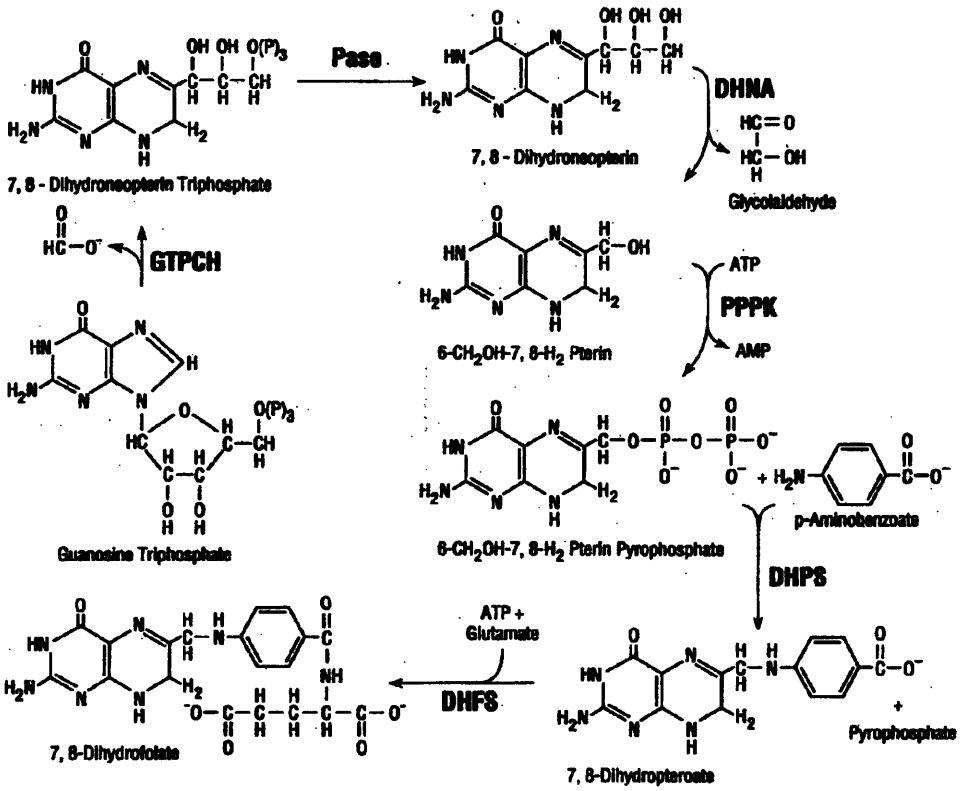


Figure 1.3 The dihydrofolate biosynthetic pathway
GTPCH, GTP cyclohydrolase I; Pase, unidentified phosphatase. Adapted from Lacks, (1995).

GTP is converted to dihydroneopterin triphosphate by GTP cyclohydrolase I (GTCase)(1.9) yielding formic acid as a byproduct. Dihydroneopterin triphosphate is then dephosphorylated to dihydroneopterin by an as yet unidentified mechanism, although it has been demonstrated that divalent cations are able to catalyse the removal of pyrophosphate from dihydroneopterin triphosphate, yielding dihydroneopterin monophosphate (De Saizieu *et al.,* 1995). Dihydroneopterin is cleaved by dihydroneopterin aldolase (DHNA)(1.10) to yield 6-hydroxymethyl-dihydropterin. Following addition of a pyrophosphate group by 6-hydroxymethyl-dihydropterin pyrophosphokinase (PPPK), the pteridine group is conjugated with pABA by DHPS to form dihydropteroate. Dihydrofolate synthetase (DHFS) then joins dihydropteroate to glutamate, to yield FH_2.

In many organisms, the chief forms of intracellular folate present are polyglutamated derivatives of FH_4; in the intracellular 'folate pool' of *P.falciparum* the principal folate present is 5-methyl tetrahydrofolate pentaglutamate (CH_3FH_4-Glu_5,

Krungkrai *et al.,* 1989). FH_2 is reduced to FH_4 by DHFR. Additional glutamate residues are then added irreversibly by folyl polyglutamate synthase (FPGS) activity. It is thought that most enzymes that utilise folate cofactors bind polyglutamated forms preferentially.

The existence of this pathway for the *de novo* synthesis of FH_2 has been confirmed in the malaria parasite (Fairlamb, 1989) and genes encoding DHFR (Bzik *et al.,* 1987), PPPK and DHPS (Brooks *et al.,* 1994; Triglia & Cowman, 1994) and GTCase, DHFS and FPGS (Lee C.S. *et al.*, 2001) in *P.falciparum* have been isolated. DHFR activity is carried on one module of a protein that also includes a module carrying TS activity, involved in folate utilisation (1.8.3). This bifunctional DHFR-TS organisation is also seen in other protozoan parasites, including *Toxoplasma gondii* (Roos, 1993), *Leishmania major* (Beverley *et al.,* 1986) and *Trypanosoma cruzi* (Reche *et al.,* 1994), and higher plants (Cossins *et al.,* 1997). PPPK and DHPS activities are also carried on separate modules of a single polypeptide, a feature also noted in *T. gondii* (Pashley *et al.,* 1997), higher plants (Cossins *et al.,* 1997), the fungal pathogen *Pneumocystis carinii*, (Volpe *et al.,* 1993), and *S.cerevisiae* (gene *FOL1*, http:// genome-www.stanford.edu/Saccharomyces). In the latter two cases, separate modules of the polypeptide also carry DHNA activity. The closely related DHFS and FPGS activities in *P.falciparum* are provided by a single bifunctional protein (Salcedo *et al.,* 2001). This is the only known example of a bifunctional DHFS-FPGS enzyme amongst eukaryotes, although many bacteria carry such enzymes, including *E.coli* (Bognar *et al.,* 1987).

1.8.2 : Folate salvage in the malaria parasite

In contrast to many bacterial species where folate biosynthesis is obligatory, many strains of malaria parasite demonstrate an ability to utilise exogenous folate (Milhous *et al.,* 1985; Watkins *et al.,* 1985; Krungkrai *et al.,* 1989). Inhibition of *in vitro* parasite growth by sulfadoxine can be alleviated in certain strains by addition of folic acid to the growth medium. This has been termed the 'folate effect' (Wang *et al.,* 1997).

The mechanism of folate salvage in the parasite is unclear however. Plasmodial DHFR is not thought to be able to reduce folic acid to FH_2 (Walter, 1991); it has therefore been postulated that the mechanism of folic acid utilisation involves the cleavage of this species into pterin-aldehyde and *para*-aminobenzoylglutamate (pABG),

which are suggested to act as substrates for PPPK and DHPS respectively (the latter reaction yielding FH_2 rather than dihydropteroate, the usual product of DHPS catalysis; Ferone, 1973) (Figure 1.4).

In addition to the utilisation of exogenous folic acid, malaria parasites *in vitro* are able to utilise 5-methyl tetrahydrofolate (CH_3FH_4) present in the RBC plasma (Asawamahasakda & Yuthavong, 1993). This acts as a cofactor for methionine synthase (MS), generating FH_4.

1.8.3 : Folate utilisation in the malaria parasite

FH_4 or, more normally, a polyglutamated derivative of this species (FH_4-Glu_n) is converted to 5,10-methylene tetrahydrofolate or a derivative (CH_2FH_4(-Glu_n)) through the addition of a single carbon unit from serine by the enzyme serine hydroxymethyl transferase (SHMT), generating glycine as a byproduct. CH_2FH_4(-Glu_n) can then be used as a one-carbon donor in the conversion of dUMP to dTMP by TS. Alternatively, CH_2FH_4(-Glu_n) may be further reduced by 5,10-methylene tetrahydrofolate reductase (MTHFR) to CH_3FH_4(-Glu_n) which serves as a cofactor in the conversion of homocysteine to methionine by MS (Asawamahasakda & Yuthavong, 1993). In the latter case FH_4(-Glu_n) is regenerated, while in the former TS catalysed conversion FH_2(-Glu_n) is generated. This is then reduced to FH_4(-Glu_n) by DHFR.

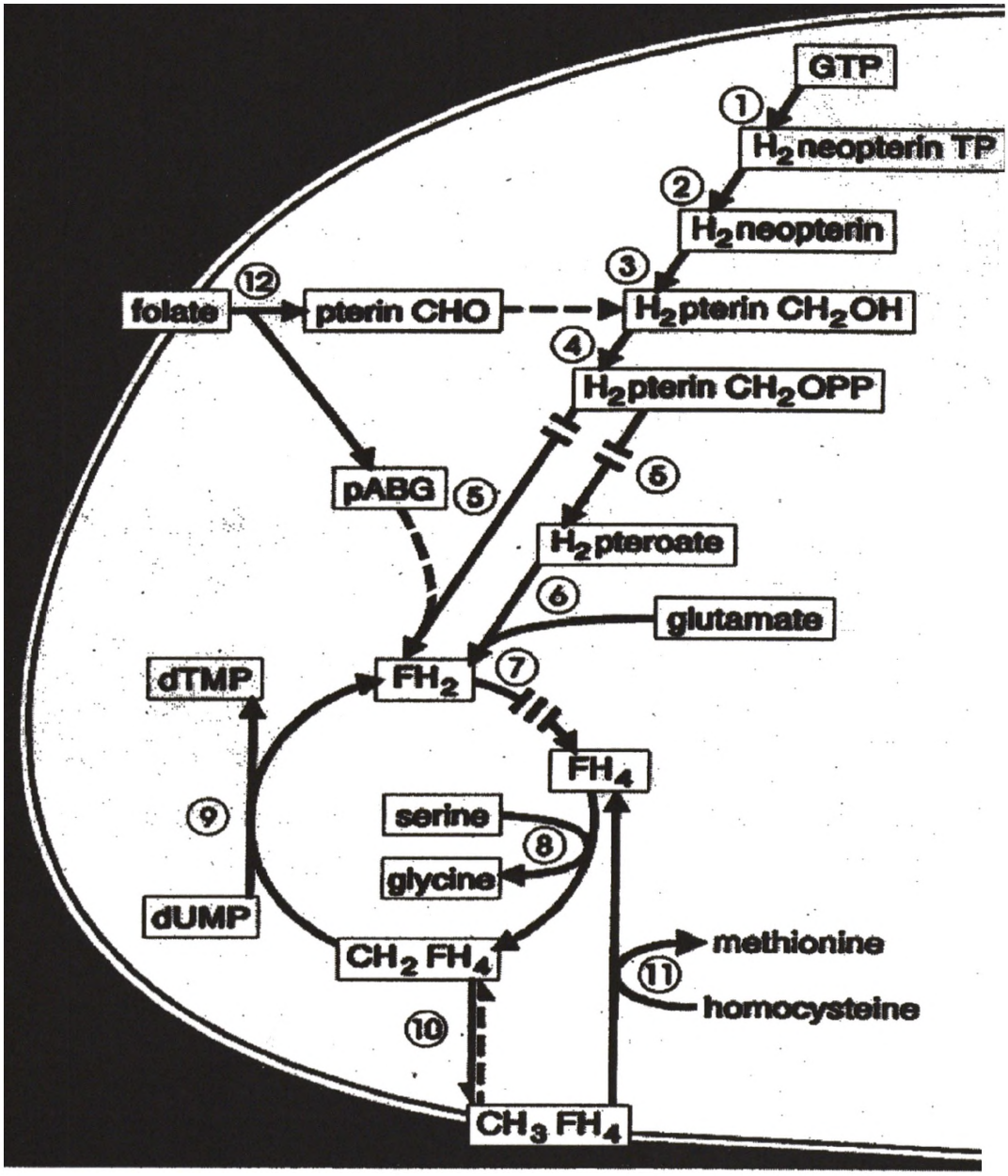


Figure 1.4 Metabolism, biosynthesis and salvage of folates in the malaria parasite Enzymes involved in folate metabolism: 1, GTCase; 2, unidentified phosphatase; 3, DHNA; 4, PPPK; 5, DHPS; 6, DHFS; 7, DHFR; 8, SHMT; 9, TS; 10, MTHFR; 11, MS; 12, pteridine 6-methylhydrolase. (-----) Proposed pathways; (-//-) inhibited by sulfones and sulfonamides; (-///-) inhibited by pyrimethamine and proguanil. Adapted from Walter (1991).

1.9 : GTP cyclohydrolase I

GTCase is responsible for the conversion of GTP to dihydroneopterin triphosphate, yielding formate (see Figure 1.3). Although GTCase activity has been identified in every organism studied, the physiological significance of its activity varies widely between species. In animals and some other higher organisms, dihydroneopterin triphosphate is converted by two further enzymes, 6-pyruvoyl tetrahydropterin synthase (PTPS) and sepiapterin reductase, to tetrahydrobiopterin (BH_4). In humans this molecule serves as a cofactor for nitric oxide synthase, catalysing the production of nitric oxide, which acts in a variety of roles, including as a vasorelaxant, a cytotoxic agent and a neuronal messenger (Griffith & Stuehr, 1995). BH_4 also acts as a cofactor for the phenylalanine, tyrosine and tryptophan amino acid hydroxylases (Nichol *et al.*, 1985). These enzymes are responsible for phenylalanine degradation and biosynthesis of catecholamine and serotonine neurotransmitters, and the shortage of BH_4 resulting from defects in human GTCase can cause hyperphenylalaninaemia (Niederwieser *et al.*, 1984) and a variety of nervous disorders, including dopa-responsive dystonia (Ichinose *et al.*, 1994).

In reptiles, amphibians and insects, GTCase is also implicated in the synthesis of various pteridine pigments, including eye pigments in the fruit fly *Drosophila melanogaster,* and pigments in butterfly wings (reviewed by Pfleiderer, 1994).

In plants and most microorganisms, including *P.falciparum*, GTCase catalyses the first step in the biosynthesis of folate cofactors (1.8.1).

The conversion catalysed by GTCase is highly complex, involving the hydrolytic opening of the imidazole ring of GTP, the removal of the C-8 carbon as formate, the Amadori-type rearrangement of the ribose ring and finally formation of the pteridine ring through the recyclisation of the intermediate species.

It was originally postulated therefore that the reaction was catalysed by several enzymes, or a multi-enzyme complex. Later research (Yim & Brown, 1976) suggested, however, that the conversion was catalysed in *E.coli* by a complex of eight identical polypeptides with a single, equivalent, active site per subunit.

Subsequent analysis of GTCase purified from *E.coli* confirmed that the enzyme consists of an array of identical subunits, although there are in fact ten, rather than eight, polypeptides in the structure (mammalian GTCases are also found as homodecamers (Auerbach & Nar, 1997; Steinmetz *et al.*, 1998), however the GTCase of *Bacillus subtilis* consists of a homooctamer (DeSaizieu *et al.*, 1995)). Five subunits associate to

form a pentameric torus, and two tori sit face to face to form the 250kDa homodecameric complex, which has dimensions of 65 x 100Å (Meining *et al.,* 1995) (Figure 1.5).

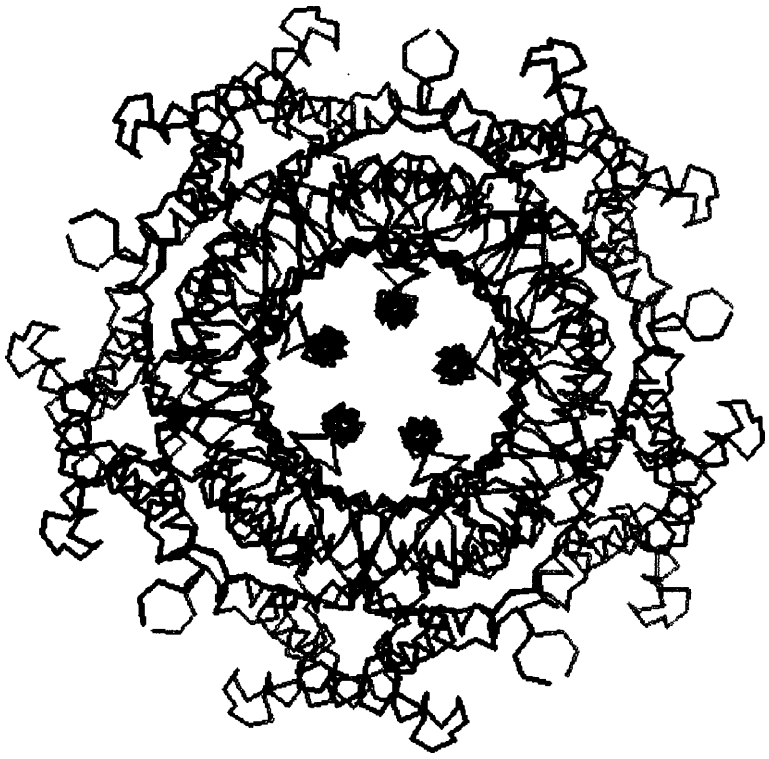


Figure 1.5 Top view of the decamer of *E.coli* GTP cyclohydrolase I
The five subunits of the upper torus are coloured red, green, blue, yellow, and cyan respectively. The lower torus is underlaid in grey (image produced using RasMol software).

An individual subunit folds into an α+β structure with a predominantly helical N-terminal region. The compact C-terminal domain consists of a four stranded β sheet flanked on either side by α-helices (Nar *et al.,* 1995a)(Figure 1.6). GTCases are included alongside DHNA, PTPS and dihydroneopterin triphosphate epimerase (DHNTE), all of which are pterin synthesis enzymes, in addition to urate oxidase in a recently described family of proteins which possess so-called T-fold (tunnelling fold) domains, similar to the *E.coli* GTCase C-terminal domain (Colloc'h *et al.,* 1997; Colloc'h *et al.,* 2000).

Figure 1.6 Two views of an *E.coli* GTP cyclohydrolase I monomer
α-helix and β-sheet regions are shown in red and yellow respectively (images produced using RasMol software).

The active site of the enzyme is found at the interface of 3 adjacent subunits, with important active site residues found mainly in loop regions of the C-terminal domains of the subunits (Nar *et al.*, 1995b). The major part of the active site cavity is formed by three sequentially distant loop segments of one subunit encompassing residues 109-113, 150-153 and 179-181. The residues involved in the binding of the pyrimidine portion of the purine ring of the substrate analogue dGTP (Ile132, Gln151 and Glu152) are found at the bottom of the active site cavity, while a cluster of basic residues around the pocket entrance binds to the triphosphate chain (Figure 1.7).

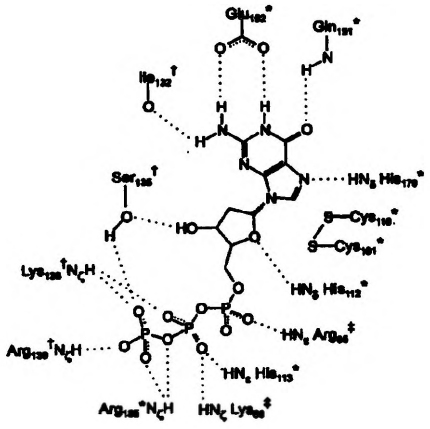


Figure 1.7 Interactions between active site residues and the substrate analogue dGTP in *E.coli* GTP cyclohydrolase I Residues are labelled with asterisk, cruciform and cross of Lorraine symbols in order to indicate their origin in 3 different monomers of GTCase. The presence of the disulphide bond between Cys110 and Cys181 has subsequently been disproved (reproduced from Nar *et al.*, 1995b).

The precise mechanism of the GTCase reaction is still a matter of question, however, it has been demonstrated that, in *E.coli*, enzymatic conversion of GTP to intermediate **4** (as shown in Figure 1.8) is fully reversible (Bracher *et al.*, 1999), an essential zinc ion in the active site, co-ordinated by His113, Cys110 and Cys181 (the latter two residues shown erroneously to form a disulphide bond in Figure 1.7), is involved in the nucleophilic attack of C-8 by a hydroxyl ion (reaction **A**, Auerbach *et al.*, 2000), and the rate of formate production (reactions **A-D**) exceeds the rate of dihydroneopterin triphosphate formation by an order of magnitude (Schramek *et al.*, 2001). It has also been suggested that the ring closure and dehydration of the Amadori product (reactions **H-J**) could occur independently of the enzyme (Bracher *et al.*, 2001).

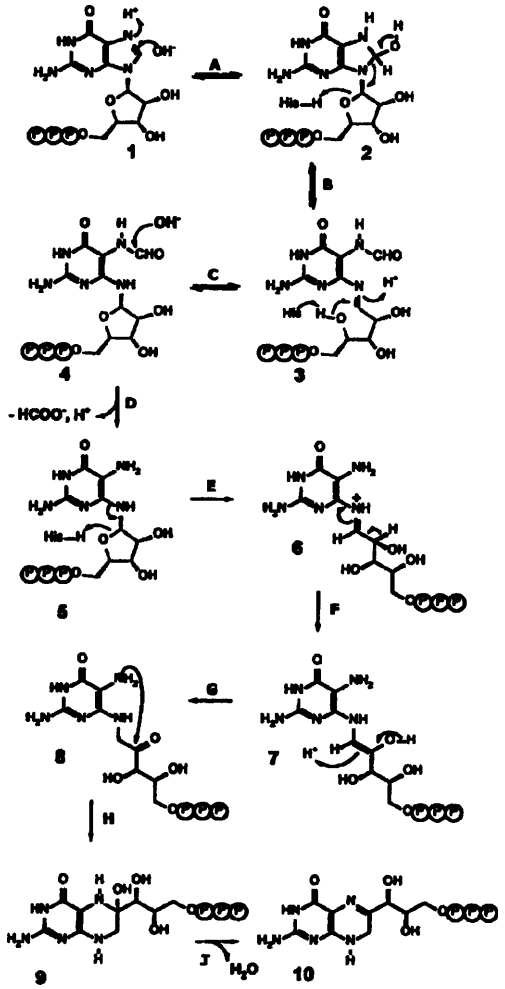


Figure 1.8 Hypothetical reaction mechanism of GTP cyclohydrolase I
Species **1** is GTP, species **10** is dihydroneopterin triphosphate.(Adapted from Bracher *et al.*, 2001).

The first GTCase gene sequences were published in 1991, following the isolation of the gene from *E.coli* (Katzenmeier *et al.,* 1991) and the cDNA from a rat liver library (Hatakeyama *et al.,* 1991). Subsequently more than 30 complete or partial GTCase sequences have been released. The polypeptides encoded by these genes show a high degree of homology between species in a 'core' C-terminal region (corresponding to the C-terminal domain in *E.coli*) and most active site residues (as marked in Figure 1.7) are fully conserved across all species, the exceptions being Lys68 and Ile132, which show considerable variation between species and the residues Lys136, Arg139 and Arg185, which are changed to Ala, Lys and Gln residues respectively in the bacterium *Helicobacter pylori* only (Appendix III).

1.9.1 : GTCase in *P.falciparum*

GTCase activity has been demonstrated in several *Plasmodium* spp. by Krungkrai *et al.,*(1985). The enzyme from *P.falciparum* shows some differences from those in other non-plasmodial species, most notably in its marked susceptibility to inhibition by N^7-methylguanosine (k_i for the malarial enzyme = 4.2μM, for the human enzyme =15.0μM). As a result, N^7-methylguanosine is able to suppress parasite growth *in vitro.* Enzyme activity was claimed to be independent of metal ions, however, the GTCase from *E.coli* has been shown to require a zinc ion bound in the active site (Auerbach *et al.,* 2000), and given the high degree of conservation of active site residues between species, including *P.falciparum* (see below), it is possible that a requirement for zinc in the parasite enzyme was overlooked.

Pf gch-1, a gene encoding an apparent GTCase in *P.falciparum,* has been isolated from both K1 (Thai isolate) genomic and Tak9/96 (Thai isolate) cDNA libraries (Lee C.S., 1999; Lee C.S. *et al.,* 2001; Appendix II). The gene, with an uninterrupted open reading frame of 1170bp, is carried in a single copy on chromosome 12. The 389 amino acid polypeptide encoded by the gene has a calculated M_r of around 46kDa. This lends credence to the assumption that the parasite enzyme is active as a multimer (given the multimeric structures observed in GTCases in other species) as the estimated M_r of the malarial enzyme is approximately 300kDa (Krungkrai *et al.,* 1985). The polypeptide shows a high degree of homology to GTCases from other species in its C-terminal 'core' region (Table 1.1) and all the principal active site residues are conserved. Structure predictions of the conserved region produce a pattern highly similar to that of

the corresponding C-terminal domain of the *E.coli* enzyme. The N-terminal region of the *P.falciparum* GTCase is unusually long however, and the full length protein is considerably longer than that of other species (cf. 221aa in *E.coli,* 243aa in *Saccharomyces cerevisiae* and 250aa in human GTCases). The N-terminal region in the *P.falciparum* enzyme contains several putative phosphorylation sites and a calcium-binding loop motif and may play a role in the regulation of the enzyme's activity.

Organism	Amino acid range	Identity (%)	Similarity (%)
Pf	262-367	n/a	n/a
Cj	60-165	48	66
Ce	98-203	46	69
Gg	112-217	45	70
Sce	117-222	45	67
Mm	117-222	44	71
Hs	126-231	43	71
Dd	106-211	43	70
Dm	148-253	41	68
Sco	75-182	40	68
Bs	63-169	38	63
Hp	56-161	37	63
Sp	60-165	36	65
Ec	96-201	36	57
Hi	94-199	33	61

Table 1.1 Similarity and identity of the core region of *P.falciparum* GTCase to corresponding sequences from other organisms The core region is defined as in Nar *et al.*, (1995b). Identity and Similarity scores were calculated using GeneDoc software. Pf, *Plasmodium falciparum*; Cj, *Campylobacter jejuni*; Ce, *Caenorhabditis elegans*; Gg, *Gallus gallus*; Sce, *Saccharomyces cerevisiae*; Mm, *Mus musculus*; Hs, *Homo sapiens*; Dd, *Dictyostelium discoideum*; Dm, *Drosophila melanogaster*; Sco, *Streptomyces coelicolor*; Bs, *Bacillus subtilis*; Hp, *Helicobacter pylori*; Sp, *Streptococcus pneumoniae*; Ec, *Escherichia coli*; Hi, *Haemophilus influenzae.*

1.10 : Dihydroneopterin aldolase

DHNA is responsible for the conversion of dihydroneopterin to 6-hydroxymethyl-dihydropterin, yielding glycoaldehyde as a byproduct (see Figure 1.3), a reaction that has been described as analogous to that catalysed by rabbit muscle aldolase (Mathis & Brown, 1970). DHNA participates in the folate biosynthetic pathway (1.8.1), and has thus far been identified exclusively from organisms with a folate biosynthetic capability. A non-folate related role has been suggested for the enzyme in the photosynthetic cyanobacterium *Synechocystis* sp. PCC 6803 however (Lee S.W. *et al,*

1999). In this organism, in addition to its status as an intermediate in the synthesis of folate, 6-hydroxymethyl-dihydropterin, the product of DHNA catalysis, acts as a precursor of the pteridine glycoside cyanopterin. This pteridine has no known function, although it is produced by *Synechocystis* in considerable quantities.

DHNA activity was first identified through studies of cell free extracts from several bacterial species. The enzyme was originally purified from *E.coli* and estimated to have a M_r of 100kDa based upon native PAGE and gel-filtration analysis (Jones & Brown, 1967; Mathis & Brown, 1970). A multimeric structure for the active enzyme was first suggested by Lopez & Lacks (1993) who worked on the DHNA of *Streptococcus pneumoniae*. Crystallographic studies on the enzyme purified from *Staphylococcus aureus* have revealed that DHNA in this species consists of a 110 kDa homooctamer, with dimensions of 65 x 70Å (Hennig *et al.*, 1998). The subunits are arranged in a similar fashion to the monomers of *E.coli* GTCase (1.9), with four subunits forming a tetrameric torus, and two tori sitting face to face to form the complete complex (Figure 1.9).

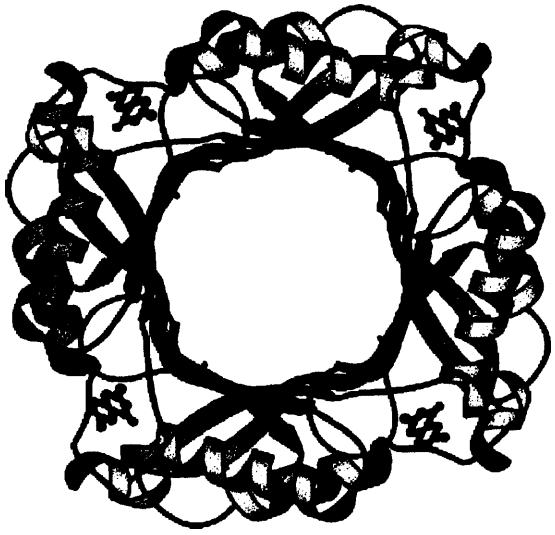


Figure 1.9 Top view of one tetramer of *S.aureus* dihydroneopterin aldolase complexed with the substrate analogue dihydrobiopterin One subunit is coloured yellow. The complete DHNA consists of two of these tetramers sat face to face (adapted from Colloc'h *et. al.*, 2000).

An individual subunit folds into an α+β structure, similar to the *E.coli* GTCase C-terminal domain, consisting of a four stranded β sheet flanked on one side by a

number of α-helices (Hennig *et al.*, 1998)(Figure 1.10). *S.aureus* DHNA is included alongside GTCases in the family of proteins which possess T-fold domains (1.9, Colloc'h *et al.*, 1997; Colloc'h *et al.*, 2000).

Figure 1.10 The *S.aureus* dihydroneopterin aldolase monomer
α-helix and β-sheet regions are shown in red and yellow respectively (image produced using RasMol software).

The active site of the enzyme is found at the interface of two adjacent subunits (Hennig *et al.*, 1998), an observation that supports the claim that multimeric assembly is required for enzyme activity in *S. pneumoniae* (Lopez & Lacks, 1993), and its morphology is highly similar to the active sites of other pterin-binding enzymes, for example that of the *E.coli* DHFR (Bystroff *et al.*, 1990, Figure 1.11).

The first gene sequence shown to encode DHNA was the *sulD* gene from the sulphonamide resistance operon of *S.pneumoniae* (Lopez & Lacks, 1993). In this species DHNA activity is carried on the N-terminal domain of a modular bifunctional protein which also carries PPPK activity. This result enabled several previously published gene sequences to be retrospectively identified as DHNA encoding genes, including *ORF1* from the folate biosynthetic operon of *B.subtilis* (Slock *et al.*, 1990), and the FasB encoding portion of the multifunctional *FAS* (folic acid synthesis) gene from *P.carinii*, (Volpe *et al.*, 1992, Volpe *et al.*, 1993). The FasC and FasD domains of this gene have been demonstrated to encode PPPK and DHPS respectively, and subsequently the FasA domain has been shown to consist of a polypeptide with 23%

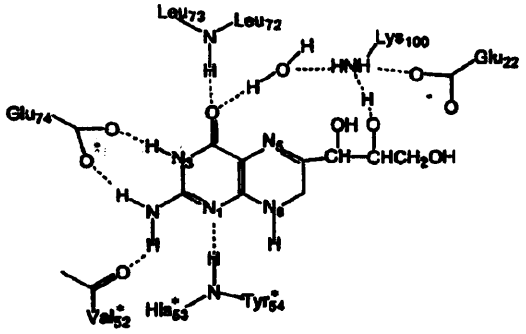


Figure 1.11 Interactions between active site residues and the substrate analogue dihydrobiopterin in *S.aureus* dihydroneopterin aldolase Residues labelled with asterisks originate in a different monomer of DHNA from those which are unmarked (adapted from Deng *et al.*, 2000).

sequence identity to the FasB (DHNA) domain (Volpe *et al.*, 1995). The FasA domain has been shown to be essential to the DHNA activity of the multifunctional protein, indicating either that FasA and FasB both serve as subunits of DHNA, or that one domain is essential for the function of the other.

At the time of writing, DHNA genes have been identified, almost exclusively on the basis of sequence homology, in more than 30 organisms. The prokaryotic genes in this number all encode apparently single-domained monofunctional proteins, leaving *S.pneumoniae* as the only prokaryotic species where DHNA activity resides on a multifunctional protein identified as yet. Interestingly however, *E.coli* carries genes encoding both DHNA and another protein, DHNTE, which shares 30% sequence identity with the *E.coli* DHNA (Haußmann *et al.*, 1997). This enzyme, which catalyses the epimerisation of dihydroneopterin triphosphate to dihydromonapterin, a pteridine compound of unknown function, shows enormous similarity to DHNA in terms both of its mechanism (Haußmann *et al.*, 1998) and its monomeric and multimeric structures (Ploom *et al.*, 1998), and it is included within the DHNA archetype in the family of proteins including T-fold domains (Colloc'h *et al.*, 2000). A putative DHNTE encoding gene has also been identified in *Pseudomonas aeruginosa* (gene folX; http://www.pseudomonas.com).

The related yeast species *S.cerevisiae* and *Schizosaccharomyces pombe* carry *Fas* type genes homologous to that of *S.pneumoniae*, whereby two DHNA-like domains (FasA and FasB) are found as part of a multifunctional polypeptide with PPPK and DHPS domains, indicating a possible fungal origin for this type of gene organisation

(gene *FOL1*, http://genome-www.stanford.edu/Saccharomyces; gene SPBC1734.03, http://www.sanger.ac.uk/Projects/S_pombe).

The flowering plant *Arabidopsis thaliana* carries two genes on separate chromosomes encoding DHNA-like polypeptides, which show 65% sequence identity (genes At3g11750 and At5g62980; http://www.tigr.org). The reason for this apparent duplication is unknown as yet.

In contrast to GTCase (1.9), the polypeptides encoded by DHNA genes show a comparatively low level of interspecies homology, and only three residues, equivalent to Gly17, Glu22 and Lys100 in *S.aureus* (the latter pair of which are found in the active site and strongly implicated in catalysis (Figure 1.11; Hennig *et al.*, 1998)), are completely conserved between all complete prokaryotic DHNA and FasB type sequences, although other active site residues for example Tyr54, Leu72, Leu73 and Glu74 (Figure 1.11), are largely conserved or semi-conserved (Appendix IV).

As yet, no sequence encoding DHNA has been identified from *P.falciparum,* or any other malaria parasite.

1.11 : Aims of the project

A well developed understanding of the function of all the enzymes involved in folate biosynthesis and metabolism (1.8) in the human malaria parasite *P.falciparum* is of vital importance, both in terms of elucidating the precise mode of action of the currently used antifolate drugs (1.5.2) and also with a view to identifying novel drug targets. In this project we aimed to exploit our knowledge of the sequence of *Pf gch-1* (1.9.1), the parasite gene believed to encode GTCase, which catalyses the first step in folate biosynthesis, in order to come to a fuller understanding of the biochemistry and physiology of this enzyme in *P.falciparum*. In the first instance we aimed to achieve heterologous expression of *Pf gch-1* in the bacterium *E.coli,* with a view to purifying expressed protein and assaying it for catalytic activity, in order to unambiguously identify the gene product as the parasite GTCase, and further investigate the particular biochemical properties of this species. We also aimed to use protein purified in this way in order to raise antibodies against the native parasite GTCase; such antibodies could then be used in a variety of further analyses, for example as an affinity ligand in purification of the native enzyme from parasite cells, or in order to investigate the cellular localisation of the enzyme through *in situ* immunolabelling studies.

Previous research has demonstrated that *S.cerevisiae* cells in which *FOL2*, the gene encoding GTCase, has been 'knocked out' show a recessive conditional lethality as they are able to propagate in the presence of folinic acid (5-formyl tetrahydrofolate)(Nardese *et al.*, 1996). Such cells therefore represent an ideal system in which to attempt expression of putative GTCase genes in order to unambiguously ascribe such identity to these genes through functional complementation. Early complementation tests of *Pf gch-1* in *S.cerevisiae FOL2* knockout cells had failed, for reasons believed to involve premature truncation of the mRNA transcripts of the parasite gene transcripts (Lee, 1999). We therefore aimed to circumvent these problems using a variety of approaches, including site-directed mutagenesis of *Pf gch-1* in order to achieve functional complementation. We also aimed to monitor stage-specific expression of *Pf gch-1 in vivo* using sequence-specific reverse transcriptase (RT) PCR.

Finally, we aimed to exploit the rapidly expanding base of knowledge provided by the malaria parasite genome sequence projects (1.6) in order to isolate the gene encoding the folate biosynthetic enzyme DHNA in *P.falciparum*, with a view to further analysis of this enzyme.

CHAPTER 2

MATERIALS AND METHODS

DNA, RNA and crude extract of *P.falciparum* strain K1 (Thai isolate) was provided by Martin Read of the UMIST Molecular Parasitology research group. RNA (stage-specific and otherwise) from *P.falciparum* strain HB3 (Honduran isolate) was provided by Niroshini Nirmalan of the UMIST Molecular Parasitology research group. *P.yoelii nigeriensis* lethal strain DNA was a gift from Dr. David Baker, Dept. of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine. Lysates of mouse RBCs infected with *P.y.yoelii* 17XL were kindly provided by Dr. Andrew Taylor-Robinson, Antimicrobial Research Centre, University of Leeds.

2.1 : Solutions

2.1.1 : TE buffer

10 mM Tris-HCl
1 mM disodium-EDTA
adjusted to pH 8.0

2.1.2 : 10 x PBS buffer

1.4 M NaCl
27 mM KCl
101 mM Na_2HPO_4
18 mM KH_2PO_4
adjusted to pH 7.3

2.1.3 : Lithium acetate solution

0.1 M lithium acetate
10 mM Tris-HCl (pH 8.0)
1 mM disodium-EDTA

DNA electrophoresis solutions

2.1.4 : 1 x TBE buffer

87 mM Tris-HCl

87 mM boric acid

2 mM disodium-EDTA

adjusted to pH 8.3

2.1.5 : 10 x Agarose gel loading buffer

10% (w/v) Ficoll 400

0.25% bromophenol blue

in 1 x TBE buffer (2.1.4)

RNA electrophoresis solutions

2.1.6 : 10 x MOPS buffer

0.2 M MOPS

90 mM sodium acetate

10 mM disodium-EDTA

adjusted to pH 7.0

2.1.7 : Formaldehyde gel loading buffer

50% (v/v) de-ionised formamide (2.1.21)

1 x MOPS buffer (2.1.6)

1.85 M formaldehyde

10% (v/v) glycerol

0.25% (w/v) bromophenol blue

Protein electrophoresis solutions

2.1.8 : 4 x SDS-PAGE separating gel buffer

1.5 M Tris-HCl (pH 8.8)

0.4% (w/v) SDS

2.1.9 : 4 x SDS-PAGE stacking gel buffer

0.5 M Tris-HCl (pH 6.8)

0.4% (w/v) SDS

2.1.10 : 5 x SDS-PAGE running buffer

15 g/l Tris base

72 g/l glycine

5 g/l SDS

2.1.11 : '1 x' SDS-PAGE reducing/loading buffer

62.5 mM Tris-HCl (pH 6.8)

10% (w/v) glycerol

2% (w/v) SDS

5% (v/v) β-mercaptoethanol

0.05% (w/v) bromophenol blue

2.1.12 : 10 x SDS-PAGE reducing buffer

0.1 M Tris-HCl (pH 7.5)

10 mM disodium-EDTA

10% (w/v) SDS

50% (v/v) β-mercaptoethanol

2.1.13 : 5 x PAGE loading buffer

50% (v/v) glycerol

0.25% (w/v) bromophenol blue

2.1.14 : Non-denaturing PAGE running/separating gel buffer

20 mM Tris-HCl (pH 8.9)

20 mM glycine

2.1.15 : Non-denaturing PAGE stacking gel buffer

20 mM Tris-HCl (pH 9.4)

20 mM glycine

2.1.16 : Coomassie blue stain solution

40% (v/v) methanol

10% (v/v) acetic acid

0.1% (w/v) Coomassie blue

2.1.17 : Destain solution

40% (v/v) methanol

10% (v/v) acetic acid

Hybridisation solutions

2.1.18 : 20 x SSC

3 M NaCl

0.3 M sodium citrate

adjusted to pH 7.0

2.1.19 : Maleic acid buffer (pH 7.5)

0.1 M maleic acid

0.15 M NaCl

2.1.20 : DIG washing solution

maleic acid buffer (2.1.19)

0.3% (v/v) Tween 20

2.1.21 : DIG blocking solution

maleic acid buffer (2.1.19)

1% (w/v) DIG blocking reagent (Boehringer Mannheim)

2.1.22 : CSPD detection buffer

100 mM Tris-HCl (pH 9.5)

100 mM NaCl

2.1.23 : (Pre)Hybridisation solution

50% (v/v) de-ionised formamide (2.1.25)

4 x SSC (2.1.18)

5 x Denhardt's solution (2.1.24)

0.1% (w/v) SDS

2.1.24 : 50 x Denhardt's solution

1% (w/v) Ficoll 400

1% (w/v) polyvinylpyrrolidine

1% (w/v) bovine serum albumin

2.1.25 : De-ionised formamide

5g of Amberlite monobed resin was added to 50 ml of formamide, stirred for 30 min at room temperature and filtered through Whatman No.1 filter paper.

Protein purification and proteolysis buffers

2.1.26 : Glutathione elution buffer

50 mM Tris-HCl (pH 8.0)

10 mM glutathione (reduced)

2.1.27 : Maltose column buffer

20 mM Tris-HCl (pH 7.4)

0.2 M NaCl

10 mM β-mercaptoethanol

1 mM disodium-EDTA

2.1.28 : Maltose elution buffer

10 mM maltose in maltose column buffer (2.1.27)

2.1.29 : GST cleavage buffer

50 mM Tris-HCl (pH 7.0)

150 mM NaCl

1 mM disodium-EDTA

1 mM dithiothreitol

2.1.30 : Genenase reaction buffer

20 mM Tris-HCl (pH 8.0)

0.2 M NaCl

Western blotting solutions

2.1.31 : Western transfer solution

25 mM Tris base

192 mM glycine

0.1% (w/v) SDS

20% (v/v) methanol

2.1.32 : Western washing buffer

1 x PBS (2.1.2)

0.3% (v/v) Tween 20

2.1.33 : Western blocking buffer

1 x PBS (2.1.2)

0.3% (v/v) Tween 20

10% (w/v) non-fat dry milk

2.1.34 : AP buffer

100 mM Tris-HCl (pH 9.5)

100 mM NaCl

5 mM $MgCl_2$

2.2 : Microbiological media

All media were sterilised by autoclaving for 20 min at 15 psi on liquid cycle

Bacteriological media

2.2.1 : LB medium

10 g/l Bactotryptone

5 g/l yeast extract

10 g/l NaCl

adjusted to pH 7.0

supplemented with 16 g/l bactoagar for a solid medium

2.2.2 : SOC medium

20 g/l Bactotryptone

5 g/l yeast extract

0.5 g/l NaCl

2.5 mM KCl

10 mM $MgCl_2$

20 mM glucose

adjusted to pH 7.0

Yeast media

2.2.3 : YEPD medium

10 g/l Bacto-Yeast extract

20 g/l Bacto-peptone

2% (w/v) glucose

2.2.4 : YNBD medium

6.7 g/l Bacto-yeast nitrogen base (without amino acids)

2% (w/v) glucose

supplemented with 16 g/l bactoagar for a solid medium

synthetic media consisted of YNBD plus dropout mix (-met, -ura) (2.2.5), supplemented with folinic acid, uracil (2.2.5) and geneticin (2.2.6) as appropriate.

Nutritional supplements and antibiotics

2.2.5 : Nutritional supplements

nutritional supplements were added to the media where required to the following final concentrations:-

uracil : 20 μg/ml

folinic acid : 50 μg/ml

dropout mix (-met, -ura) : 0.2% (w/v)

dropout mix (-met, -ura) was prepared by thoroughly mixing 4g of leucine, 2g of each of alanine, arginine, asparagine, aspartic acid, cysteine, glutamine, glutamic acid, glycine, histidine, inositol, isoleucine, lysine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine and valine, 0.5g of adenine and 0.2g of pABA.

2.2.6 : Antibiotics

antibiotics were added to the media where required to the following final concentrations:-

carbenicillin (to select for amp^R genotype) : 50 μg/ml

tetracycline : 12.5 μg/ml

geneticin (to select for kan^R genotype) : 200 μg/ml

2.3 : Standard nucleic acid manipulation techniques

2.3.1 : Isolation of DNA from reaction mixes (affinity column method)

DNA was purified from PCR mixes, restriction digests and other enzymatic reaction mixes using the QIAquick PCR Purification Kit (QIAGEN) or the CONCERT™

Rapid PCR Purification System (GibcoBRL) according to the manufacturers' protocols. Purified DNA was eluted in TE buffer (2.1.1) or 10 mM Tris-HCl (pH 8.0).

2.3.2: Isolation of DNA by size-exclusion chromatography

The reaction mix containing the DNA of interest was made up to 50µl with TE buffer (2.1.1), and applied to a prepared Microspin G25 column (Amersham) sat in a microfuge tube. The column was spun for 2 min at 735 x g. An additional 50µl TE was then applied to the column, which was then spun for a further 2 min at 735 x g. The DNA was contained in the combined eluates.

2.3.3 : Small scale preparation of plasmid DNA

Plasmid DNA was purified from 1 to 5 ml of an *E.coli* overnight culture by an alkaline lysis method using the QIAprep Spin Miniprep Kit (QIAGEN) or the CONCERT™ Rapid Plasmid Miniprep System (GibcoBRL) according to the manufacturers' protocols. Purified DNA was eluted in TE buffer (2.1.1) or 10 mM Tris-HCl (pH 8.0).

2.3.4 : Extraction of total RNA from yeast

Total RNA was purified from 2ml of an *S.cerevisiae* mid-log phase culture using the RNeasy Mini Kit (QIAGEN) or, where explicitly stated, from 20ml of the same using the RNeasy Midi kit (QIAGEN) according to the manufacturer's protocols. In both cases the yeast cells were incubated post-harvest for 30min in a solution containing 60U/ml of lyticase (Sigma) in order to disrupt the cell wall and produce spheroplasts. Following cell lysis with a highly denaturing buffer, RNA was purified on an affinity column and eluted in RNase free H_2O. Where possible, all additional materials and equipment were treated with 0.1% diethyl pyrocarbonate (DEPC), incubated overnight and then autoclaved at 15psi for 20min on liquid cycle in order to inactivate RNases.

2.3.5 : Purification of DNA from agarose gel slices

Gel slices were excised from the agarose slab using a clean scalpel blade. DNA was purified from gel slices of less than 400mg using the QIAquick Gel Extraction Kit (QIAGEN) or the CONCERT™ Rapid Gel Extraction System (GibcoBRL) according to

the manufacturers' protocols. Purified DNA was eluted in TE buffer (2.1.1) or 10 mM Tris-HCl (pH 8.0).

2.3.6 : Determination of DNA concentration

In order to roughly determine the DNA concentration of an unknown sample, an aliquot of the sample was subjected to agarose gel electrophoresis (2.6.1) alongside known quantities of a bacteriophage λ genomic DNA restriction digest. The brightness of the band produced by the unknown sample on the transilluminator was compared to the control bands, and the DNA concentration was extrapolated from this comparison.

2.3.7 : Drop dialysis

This method was used to remove residual salts, which might inhibit DNA ligation and other sensitive procedures, from DNA preparations. Samples of DNA in volumes of 20-100 µl were applied to the shiny surface of a 0.025 µm filter (Millipore) floating on 10mM Tris-HCl and allowed to dialyse for 30-60 min at room temperature.

2.4 : Standard enzyme treatments

2.4.1 : Restriction endonuclease digestion

DNA was digested by incubation with the appropriate enzyme at 37°C for 2 hrs or more. Enzymes were supplied by Boehringer Mannheim or Stratagene and used with the recommended buffer. 2-3 units of enzyme per µg of DNA were used and the final volume was adjusted with sterile distilled H_2O such that the DNA concentration was no greater than 50ng/µl. For digestions with two enzymes, the digestion was performed simultaneously where the recommended buffer systems were compatible. Otherwise DNA was purified (2.3.1) following restriction with one enzyme, and subsequently digested with the second enzyme.

2.4.2 : Alkaline phosphatase treatment

To prevent recircularisation of digested vector DNA, arctic shrimp alkaline phosphatase (supplied by Boehringer Mannheim) was used to remove the 5' phosphate groups from the ends of the DNA molecules. Vector DNA was incubated with the

enzyme at 37°C for 10 min in 1 x reaction buffer. At least 1 Unit of enzyme was added for every picomole of DNA molecules. The enzyme was then completely inactivated by heating at 60°C for 15 min.

2.4.3 : Ligation

T4 DNA ligase was supplied by Boehringer Mannheim. Ligation reactions were carried out in 1 x ligase buffer (as detailed in the manufacturer's protocol) in 20µl total volume (the volume was adjusted using sterile distilled de-ionised H_2O). 5 ng of vector DNA (2.5.1) was added, while insert DNA (2.5.1) was added to give insert:vector molar ratios of 6:1 and 3:1. 0.2 Weiss units of ligase was added. Ligation controls containing single-cut, unphosphatased plasmid both with and without ligase, and prepared vector with ligase but no insert were also prepared. The reaction mixtures were incubated at 15°C overnight and terminated by incubation at 70°C for 15min.

2.4.4 : T4 polynucleotide kinase reaction

Where 5' phosphorylated oligonucleotides were required for site-directed mutagenesis (2.9) 500 picomoles of oligonucleotide were incubated with 20 Units T4 polynucleotide kinase (supplied by New England Biolabs) in 100µl 1 x reaction buffer at 37°C for 30-60 min. The enzyme was subsequently deactivated by incubation at 75°C for 10 min.

2.5 : Cloning and transformation

2.5.1 : Preparation of vector and insert DNA for cloning

Vector and insert source DNA was subjected to restriction endonuclease digestion (2.4.1) for 2 hrs with the appropriate enzyme(s) to generate DNA molecules with the necessary compatible sticky ends. Where necessary, vector DNA was treated with alkaline phosphatase (2.4.2) to prevent later recircularisation. Vector and insert DNA was purified as in 2.3.1.

All DNA solutions were drop dialysed (2.3.7), and the DNA concentrations were assessed (2.3.6) before use in ligation reactions (2.4.3).

2.5.2 : Preparation of competent cells by $CaCl_2$ methodology

A fresh overnight culture of *E.coli* was used to inoculate 50 ml of LB (2.2.1) supplemented with the appropriate antibiotic (2.2.6) as necessary. The cells were grown at 37°C to an OD_{600} of 0.7 to 0.8. The cells were then cooled on ice for 10 min before being centrifuged at 3,000 rpm for 5 min in pre-cooled centrifuge tubes. After decanting the supernatant, the cells were resuspended in 25 ml ice-cold 0.1M $CaCl_2$. The cells were then left on ice for 30 min before being spun down again as before. Finally the cells were resuspended in 2.5 ml ice-cold 0.1M $CaCl_2$, and used within the next few hours.

2.5.3 : Transformation using $CaCl_2$ treated competent cells

100µl aliquots of competent cells (2.5.2) were mixed with 20 ng of uncut plasmid DNA and left on ice for 5 min. Aliquots of the transformed cells were then spread onto agar plates (containing the appropriate antibiotic) which had been prewarmed to 37°C and incubated at 37°C overnight.

2.5.4 : Preparation of electrocompetent cells

A fresh overnight culture of *E.coli* was diluted 100-fold into 1 litre of LB (2.2.1) and grown at 37°C with vigorous shaking to an OD_{600} of 0.5 to 0.8 (corresponding to mid-log phase). The cells were chilled on ice for 15min and centrifuged in a cold rotor at 4,000g for 15 min. After discarding the supernatant, the pellets were resuspended in 1 litre of ice-cold water and harvested again. A further wash with 1 litre of ice-cold water was performed, and the cells were then resuspended in 20ml of ice-cold 10% (v/v) glycerol. Following a repeat centrifugation the cells were resuspended in a final volume of 3 ml ice-cold 10% glycerol. The cells were subsequently frozen in 40µl aliquots on dry ice and stored at -80°C. The cells are viable for six months under these conditions.

2.5.5 : Electroporation

Electroporation was performed using the Biorad Gene Pulser. Capacitance, resistance and voltage were set at 25 µF, 200 Ω and 1.80 kV respectively. A 40µl aliquot of electrocompetent cells (2.5.4) was thawed and placed on ice. The cell suspension was then mixed with 1µl ligation mix (2.4.3) or 250 pg uncut plasmid DNA in low salt buffer and incubated briefly on ice. The mix was then pulsed in an ice-cold

cuvette. The cells were immediately resuspended in 1 ml of SOC (2.2.2) that had been pre-warmed to 37°C, and transferred to a fresh 1.5 ml microcentrifuge tube. Phenotypic expression was allowed by shaking gently at 37°C for 1hr. Aliquots of the transformed cells were then spread onto agar plates containing the appropriate antibiotic and incubated at 37°C overnight.

2.5.6 Preparation of competent yeast cells

A fresh overnight culture of *S.cerevisiae* was used to inoculate 50 ml of YEPD (2.2.3). The cells were grown at 30°C to an OD_{600} of 0.5 to 0.8. The cells were then centrifuged at 4,000 rpm for 5 min at 4°C. After decanting the supernatant, the cells were resuspended in 5ml lithium acetate solution (2.1.3). The cells were then spun down again as before, resuspended in 0.5ml lithium acetate solution and used within the next few hours.

2.5.7 : Transformation of yeast cells

100µl aliquots of competent cells (2.5.6) were mixed with 1µg of plasmid DNA and left at room temperature for 5 min. 280µl of 50% (w/v) PEG4000 in lithium acetate solution (2.1.3) was then added to each aliquot, and mixed by inversion. Following incubation at 30°C for 45min, 43µl of DMSO was added to each transformation mixture and mixed by inversion to give a final DMSO concentration of approximately 10%. Each mixture was then incubated at 42°C for 5min. The cells were then spun in a microcentrifuge for 5sec, washed in sdH_2O, resuspended in 1ml YEPD media (2.2.3) and incubated with gentle shaking at 30°C overnight. The cells were finally spun in a microcentrifuge for 5sec, washed several times in sdH_2O, resuspended in 200µl sdH_2O, spread onto uracil depleted synthetic media plates and incubated at 30°C overnight.

2.5.8 : Analysis of *E.coli* transformants

The uptake of plasmids was selected by the conferred resistance to an antibiotic e.g. ampicillin/carbenicillin. Potential recombinants were subjected to further analysis. Colonies were streaked onto LB agar (2.2.1) containing carbenicillin (2.2.6) with individual sterile toothpicks (the plates were subsequently incubated at 37°C overnight and stored at 4°C until required again). 50 µl aliquots of 10 mM Tris-HCl or TE buffer

(2.1.1) were then inoculated with 4 or 5 toothpicks each. The resulting mixed cell suspensions were incubated in a boiling water bath for 10-15 min to lyse the cells. The cell lysates were cooled on ice, and 10 µl of each was used as template in a 25 µl 'primary' PCR (2.7.1) with primers chosen either to anneal only to insert specific sequences or to anneal on either side of the vector's multiple cloning site. The PCR products were resolved and visualised by agarose gel electrophoresis (2.7.1). Groups of transformants including one or more recombinant were identified by the presence of an amplification product of the expected size in the case of insert-specific PCR, or the presence of a product significantly larger than that produced when non-recombinant plasmid is used as the template in the case of vector-specific PCR. Each member of the 'positive' group of colonies was then used to inoculate an individual aliquot of Tris-HCl or TE (2.1.1), lysed as before and subjected to a 'secondary PCR', either insert- or vector-specific. In this way, recombinant colonies could be isolated. Clonal plasmid DNA was then purified (2.3.4) and subjected to restriction endonuclease digestion (2.4.1) with the appropriate enzyme(s) and visualised by agarose gel electrophoresis (2.6.1) in order to verify the presence of an insert of the expected size. Plasmid DNA was then sequenced (2.8) to check for misincorporations and frame-shifts.

2.5.9 : Analysis of *S.cerevisiae* transformants

The uptake of plasmids was selected for by the conferred ability to grow on uracil depleted solid synthetic media (2.2.4). The presence of the plasmid was further confirmed if desired by yeast colony PCR. Colonies were picked with individual sterile toothpicks and inoculated into 50 µl aliquots of TE buffer (2.1.1). The cell suspensions were then incubated in a boiling water bath for 10-15 min to lyse the cells. The cell lysates were cooled on ice, and 10 µl of each was used as template in a 25 µl PCR (2.7.1) with primers chosen to anneal to plasmid specific sequences. The PCR products were resolved and visualised by agarose gel electrophoresis (2.7.1).

2.6 : Gel electrophoresis

2.6.1 : Agarose gel electrophoresis for DNA

Standard agarose slab gels were used at 0.7-1.0 % (w/v) to resolve DNA molecules ranging from 0.4 to 7 kb. Agarose was dissolved in 1 x TBE (2.1.4) to the desired concentration by heating the solution in a microwave oven. The melted agarose was allowed to cool to about 50°C, mixed with ethidium bromide (0.5 μg/ml final concentration) and poured into a gel mould with a comb positioned towards one end to form the wells. The set gel was placed in an electrophoresis tank and covered with 1 x TBE. Samples were loaded in 1 x agarose gel loading buffer (2.1.5) and electrophoresis conducted at around 6-7 V/cm. The DNA was visualised and photographed if necessary on a transilluminator at 254 nm.

2.6.2 : Formaldehyde denaturing RNA gel electrophoresis

All equipment and solutions were pre-treated to prevent contamination with RNases as described before (2.3.4). 1.3g of agarose was dissolved in 95ml of 1 x MOPS buffer (2.1.6) by heating the solution in a microwave oven. The melted agarose was allowed to cool to about 50°C, mixed thoroughly but gently with 5ml of 37% (w/v) formaldehyde solution and poured into a gel mould with a comb positioned towards one end to form the wells. Samples of total RNA were mixed with one or more volumes of formaldehyde gel loading buffer (2.1.7), incubated at 65°C for 15min, and subsequently kept on ice until required. The set gel was placed in an electrophoresis tank and covered with 1 x MOPS buffer. After the samples (to which ethidium bromide had been added to a final concentration of 50μg/ml) had been loaded, electrophoresis was conducted at around 5 V/cm, for a sufficient time to let the dye migrate 75% of the length of the gel. The DNA was visualised and photographed alongside a ruler on a transilluminator at 254 nm.

2.6.3 : SDS denaturing polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was utilised to separate polypeptides according to Laemmli (1970). Owing to the nature of the equipment used, gels were produced in duplicate.

Glass plates, previously treated with dichlorodimethylsilane, were washed with hot water and detergent, dried, washed with ethanol and dried again. These plates, separated by 0.75mm plastic spacers, were then assembled on a Bio-Rad mini-protean II casting stand.

40µl of a 10% (w/v) AMPS solution was then added to 10ml of 12% polyacrylamide separating gel mix (40% (v/v) acrylamide/bisacrylamide stock solution (30% (w/v), Severn Biotech or Bio-Rad), 1 x SDS-PAGE separating gel buffer (2.1.8), 1% (v/v) TEMED) and this mixture was quickly poured into the plate assemblies to an approximate height of 5cm. Ethanol was then layered over the top of the gel mix to exclude oxygen and aid polymerisation. After the gels had set, the ethanol was poured off and the top of the gels were washed with distilled water.

35µl of 10% (w/v) AMPS was then added to 5ml of 4% polyacrylamide stacking gel mix (13.3% (v/v) acrylamide/bisacrylamide stock solution, 1 x SDS-PAGE stacking gel buffer (2.1.9), 1% (v/v) TEMED) and this mixture was quickly poured onto the set separating gels up to the top of the plate assemblies. A plastic comb was then inserted into each stacking gel to create a set of wells. The gels were allowed to set for at least 90min, then the combs removed. The gel assemblies were then removed from the casting stand, positioned on a Bio-Rad mini-protean II electrophoresis frame and placed in an electrophoresis tank. 1 x SDS-PAGE running buffer (2.1.10) was poured into the tank, and any remaining air bubbles were flushed from the wells.

Protein samples were mixed with 10 x SDS-PAGE reducing buffer (2.1.12) and 5 x PAGE loading buffer (2.1.13) to final concentrations of 1 x each, or alternatively one or more volume of '1 x' SDS-PAGE reducing/loading buffer (2.1.11) was added to the sample. Cell pellets were resuspended in '1 x' SDS-PAGE reducing/loading buffer. All samples were then boiled for 6min before being loaded into the wells using microcapillary pipette tips. Electrophoresis was conducted at 120V for approximately 2hr. Subsequently, if staining of the gel for total protein was required, the separating gel was stained for 1hr with Coomassie blue stain solution (2.1.16), and destained in destain solution (2.1.17) to allow the polypeptides to be visualised. If greater sensitivity was required, the separating gel was stained overnight in a 4:1 mix of Brilliant Blue G – colloidal solution (Sigma): methanol and destained for approximately 1min in a 25% (v/v) methanol, 10% (v/v) acetic acid solution then fixed in a 25% methanol solution to

allow the polypeptides to be visualised (except where indicated, the former staining technique was used.)

2.6.4 : Non-denaturing polyacrylamide gel electrophoresis

Non-denaturing PAGE was carried out as SDS-PAGE (2.6.3) with the following exceptions. SDS-PAGE stacking gel and separating gel buffers were replaced with non-denaturing PAGE stacking gel and running/separating gel buffers (2.1.15, 2.1.14) respectively. SDS-PAGE running buffer was also replaced with non-denaturing PAGE running/separating gel buffer (2.1.14). All samples were prepared in 1 x PAGE loading buffer (2.1.13) only, and were not boiled prior to electrophoresis.

2.7 : Polymerase chain reaction

2.7.1 Standard PCR

The composition of a generalised PCR consisted of 200 μM dATP, dCTP, dGTP and dTTP, 1.5 mM $MgCl_2$, 3 ng or 0.3 pmol/μl of each primer, a very small amount of template DNA (for example 1 pg/μl of plasmid or 3ng/μl genomic DNA) and 0.025 Units/μl of DNA polymerase (supplied by Boehringer Mannheim) in 1 x polymerase buffer. For PCRs where high yield of amplification product was the greatest consideration, *Taq* DNA polymerase was used. However, where accurate replication of the template sequence was paramount (for example in generating insert DNA for cloning) *Pwo* DNA polymerase was used.

The generalised parameters of a PCR using primers which annealed to A+T rich parasite DNA sequences (1.6) were as follows: -

30 cycles of;

Denaturation: 94°C x 30 sec

Annealing: 50°C x 30 sec

Extension: 70°C x 3 min

Where non-parasite sequence was used as the target for primer annealing, a higher annealing temperature could be used, e.g. 60°C.

Amplification products were resolved and visualised using agarose gel electrophoresis (2.6.1).

2.7.2 : Reverse transcriptase (RT) PCR

RT-PCR was used to indirectly amplify DNA fragments from RNA. The first step in this technique is the synthesis of 1st strand cDNA from the RNA template, using the 1st Strand cDNA Synthesis Kit for RT-PCR (AMV) (manufactured by Roche). The following were added to a microfuge tube on ice and made up to 20µl with sdH_2O; 2µl 10x reaction buffer, 4µl 25mM $MgCl_2$, 2µl dNTP mix, 2µl of either oligo-$p(dT)_{15}$ primer or sequence-specific primer at 10pmol/µl, 50U of RNase inhibitor, 20U of avian myeloblastosis virus (AMV) reverse transcriptase and 3µl total parasite RNA (approximately 1µg).

The reaction was incubated at 25°C for 10min, then 42°C for 60min. Finally the reverse transcriptase was inactivated by incubation at 99°C for 5min. Subsequently, this reaction mix was used as template in a standard PCR (2.7.1), usually 7µl in a 100µl reaction volume.

2.8 : DNA Sequencing

The ABI PRISM® BigDye™ Terminator Cycle Sequencing Ready Reaction Kit (Perkin Elmer) was utilised for fluorescent cycle sequencing of plasmid DNA and PCR products. 10 µl reaction mixes were prepared with 1.6 pmol of oligonucleotide, 4µl of Ready Reaction Mix and either 100-200 ng of plasmid DNA or 5-10 ng of PCR product. These were subjected to thermal cycling as described in the manufacturer's protocol. The extension products were then purified by ethanol/sodium acetate precipitation, resuspended in 2 µl loading buffer and denatured by heating at 95°C for 2 min, before loading into an ABI 373A DNA sequencer (Perkin Elmer). Sequence data was viewed using Chromas software.

2.9 : Site-directed mutagenesis

The Chameleon™ Double-Stranded Site-Directed Mutagenesis Kit (Stratagene) was used to introduce specific point mutations into the nucleotide sequence of a plasmid template (Figure 2.1). The target plasmid is heat denatured and two (or more) oligonucleotide primers are simultaneously annealed to one of the strands. The selection primer changes one nonessential unique restriction site to a new restriction site. The mutagenesis, or mutagenic primer(s) encodes a specific mutation in a defined sequence

of the target plasmid. The primers are extended around the plasmid using a nucleotide and DNA polymerase/DNA ligase mix. The newly synthesised strand contains the desired mutations and the new restriction site. The DNA is then subjected to digestion with the enzyme corresponding to the unique restriction site present in the unmutated parental plasmid (the selection enzyme). Parental plasmid DNA is thus linearised while parent/mutant hybrid plasmid DNA remains circularised.

The hybrid DNA is then transformed (2.5.3) into XL*mutS*, a repair deficient strain of *E.coli* (Appendix I). This strain is unable to distinguish between the parental and newly synthesised strands in the hybrid plasmid and arbitrarily selects one strand for degradation, synthesising the complement to the undegraded strand. Half the transformants now carry the mutant plasmid, while the other half carry the parental plasmid. The transformed bacteria are grown in a liquid culture overnight and plasmid DNA is subsequently isolated (2.3.4). The DNA is then subjected to a second digestion with the selection enzyme to linearise the parental half of the plasmid mix, before being transformed into a standard *E.coli* cell line, for example, XL1-Blue (Appendix I). Aliquots of the transformation mix are then spread onto LB agar (2.2.1) containing the appropriate antibiotic (2.2.6) and incubated at 37°C.

Prospective mutant colonies were streaked onto LB with individual sterile toothpicks (the plates were subsequently incubated at 37°C overnight and stored at 4°C until required again). The toothpicks were then used to inoculate 50 µl aliquots of 10 mM Tris-HCl. The resulting cell suspensions were incubated in a boiling water bath for 10-15 min to lyse the cells. The cell lysates were cooled on ice, and 10 µl of each was used as template in a 25 µl PCR (2.7.1) with primers chosen to amplify the mutation target sequence. The amplification products were purified (2.3.1) and sequenced (2.8). Following identification of clones which had incorporated the desired mutations, clonal plasmid DNA was purified (2.3.4) and sequenced directly to confirm that incorporation had occurred and to check for extraneous mutations.

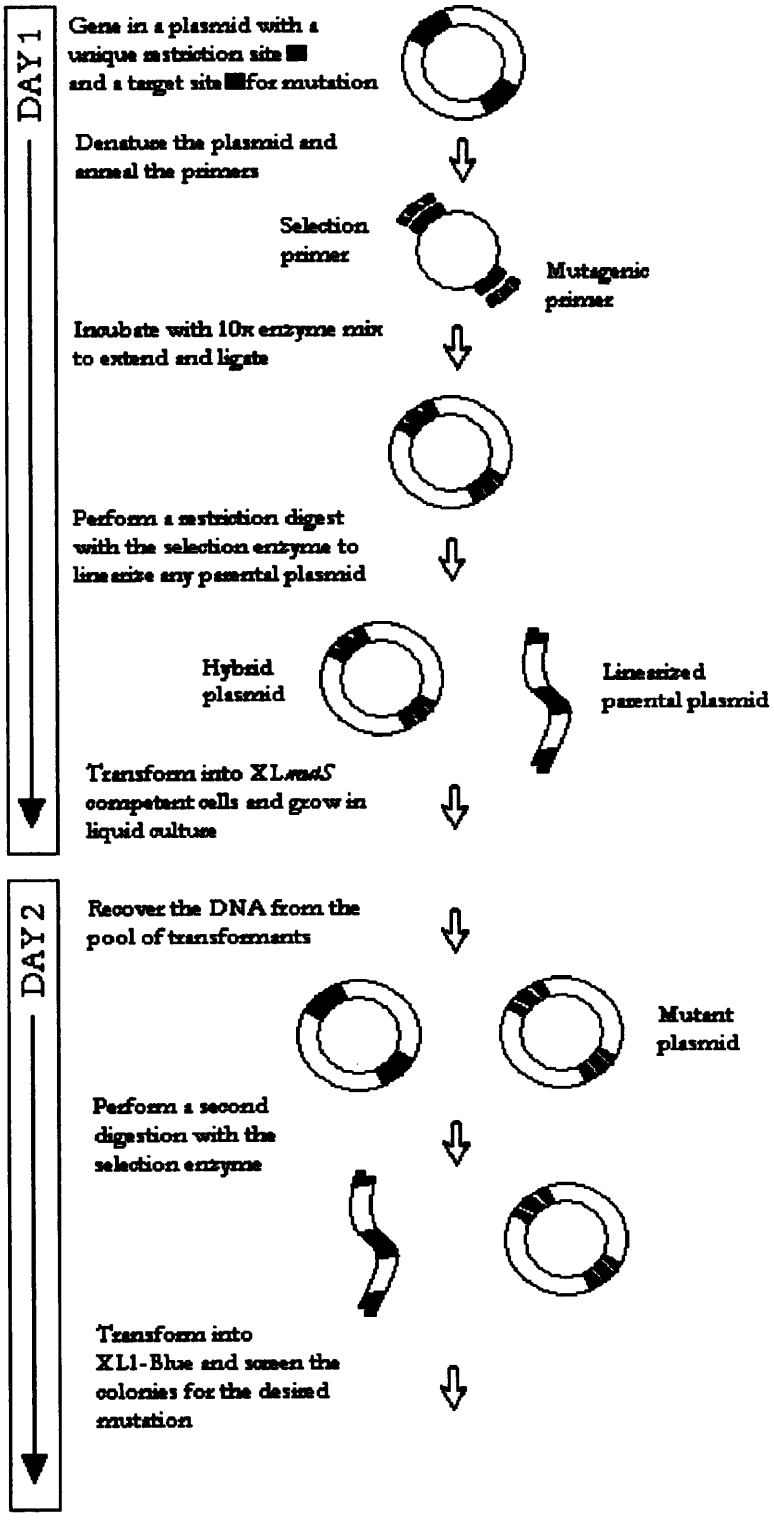


Figure 2.1 Overview of the Chameleon site-directed mutagenesis protocol
Adapted from the manufacturer's protocol (Stratagene).

2.10 Northern analysis

2.10.1 : Northern blot

Northern transfer of RNA from a formaldehyde denaturing gel (2.4.2) to a hybridisation membrane was performed using the upward capillary technique. All the required apparatus and materials were pre-treated to ensure RNase inactivation as described in 2.3.4. A prepared gel (2.4.2) was washed for 2 x 20min with 10 x SSC (2.1.18). A transfer assembly was then prepared as follows; a soaked foam sponge was placed in a 20cm x 30cm tray filled to a depth of approx. 1cm with 10 x SSC. 3 layers of soaked blotting paper were then placed on top of the sponge and the gel was placed on top of the paper. A piece of Hybond N+ nylon membrane (Amersham Pharmacia Biotech) cut to the exact size of the gel but with one corner cut off in order to facilitate later orientation, which had been briefly pre-wetted, was then placed on top of the gel. A further 3 layers of soaked blotting paper were placed on top of the membrane. A stack of paper towels of approximately 12cm in height was then added, and finally a glass plate with a 300g weight positioned in the centre was placed on top of the entire assembly. The gel was left to blot for at least 12hr, then the membrane was extricated and the RNA immobilised onto the surface by crosslinking in a UV Stratalinker 1800 (Stratagene).

2.10.2 : Synthesis of DIG-labelled RNA probe

Preparation of digoxygenin-UTP (DIG) labelled RNA probe was performed using the DIG RNA Labeling Kit (SP6/T7) (supplied by Boehringer Mannheim). This technique requires a plasmid clone of the target gene in **opposite** orientation to either an SP6 or T7 RNA polymerase promoter sequence. Following linearisation of the plasmid by restriction (2.4.1) at a remote site, antisense RNA incorporating the DIG-labelled UTP is synthesised using the correct polymerase.

The probe was tested for label incorporation by comparison with control labelled RNA. Serial dilutions of both the probe and the control RNA were prepared, and 1µl of each dilution spotted onto a piece of Hybond N+ nylon membrane (Amersham). After immobilisation of the RNA on the membrane by crosslinking in a UV Stratalinker 1800 (Stratagene), detection of DIG label was carried out using anti-DIG-AP conjugate as in 2.10.5, except that the membrane was finally incubated in 20ml AP buffer (2.1.34) containing the colorimetric reagents NBT and BCIP (Promega, 132µl and 66µl

respectively), rather than detection buffer containing CSPD. The RNA spots could thus be visualised directly as in immunoprobing of Western blots (2.12.2).

2.10.3 : Synthesis of radiolabelled DNA probe

Preparation of labelled DNA was performed using random primed synthesis. The following were added to a microfuge tube on ice and made up to 20µl with sdH_2O; 25 ng denatured template DNA, 3µl dNTP mixture (prepared by making a 1+1+1 mixture of 0.5mM dATP, dCTP and dGTP each), 2µl Hexanucleotide mix (Roche), 5µl (50uCi) [α^{32}P] dTTP, (3000 Ci/mmol, aqueous solution, Amersham) and 1µl (2U) labelling grade Klenow enzyme (Roche) ([α^{32}P] dTTP was used where parasite DNA sequence was used as the template, however, it was more convenient to use [α^{32}P] dCTP, and the corresponding dNTP mix, in other cases.)

The reaction mix was mixed well and incubated at 37°C for 30min. The reaction was then terminated through incubation at 65°C for 10min. Labelled probe was purified in 100µl volume from the reaction by size exclusion chromatography (2.3.2) and then assayed for activity by counting total radiation counts per minute (cpm) on a 1409 DSA liquid scintillation counter (Wallac). Probes with activities of greater than $1x10^8$ cpm/µg template DNA were deemed to be useful for northern analysis.

2.10.4 : Hybridisation with DIG-labelled probe and post hybridisation washes

As the probe in this protocol is RNA-based, all the required apparatus and materials were pre-treated to ensure RNase inactivation as described in 2.3.4. Hybridisations were performed in glass hybridisation tubes in a Techne hybridisation oven. Membranes were prehybridised in DIG EasyHyb solution (DIG EasyHyb granules (Boehringer Mannheim) dissolved in RNase free water) at 68°C for at least 1hr. Probes (2.10.2) were denatured by heating for 10min in a boiling water bath, diluted in DIG EasyHyb solution to a concentration of 50-100ng/ml and added to the membrane after the prehybridisation solution had been discarded. The probe was allowed to hybridise with the membrane at 68°C overnight then discarded. The membrane was then washed twice at room temperature with 15ml of a 2 x SSC (2.1.18)/0.1%SDS solution and twice at 68°C with a 0.1 x SSC/0.1%SDS solution.

2.10.5 : Chemiluminescence detection

After the post hybridisation washes (2.10.4) the membrane was rinsed briefly in DIG washing solution (2.1.20) and then incubated at room temperature with 20ml of DIG blocking solution (2.1.21) for 30min. The blocking solution was discarded and anti-DIG-AP conjugate (Boehringer Mannheim) at a dilution of 1:10,000 in 20ml DIG blocking solution was added to the membrane and incubated for a further 30min. The membrane was subsequently washed twice for 15min each with 50ml of DIG washing solution. After rinsing with 10ml of CSPD detection buffer (2.1.22) the membrane was immersed in the same buffer containing 0.25 mM CSPD (Boehringer Mannheim) for 10-15 min. The membrane was then sealed in Saran wrap and incubated at 37°C for 15min, before being exposed at room temperature to X-ray film in an autoradiography cassette for 1 to 16hr.

2.10.6 : Hybridisation with radiolabelled probe and post hybridisation washes

As the probe in this protocol is DNA-based, and the RNA on the membrane is immobilised by UV, special measures to safeguard against RNA degradation were not undertaken. Hybridisations were performed in glass hybridisation tubes in a Techne hybridisation oven. Membranes were prehybridised in (pre)hybridisation solution (2.1.23) at 37°C for at least 1hr. Probes (2.10.3) were added to 15ml of (pre)hybridisation solution and denatured by heating for 10min in a boiling water bath, before being added to the membrane after the prehybridisation solution had been discarded. The probe was allowed to hybridise with the membrane at 37°C overnight then discarded. The membrane was subsequently washed four times with 15ml of a 2 x SSC (2.1.18)/0.1%SDS solution and four times with a 0.1 x SSC/0.1%SDS solution, all at 37°C. The membrane was then sealed in Saran wrap and exposed at -80°C to X-ray film in an autoradiography cassette for 1 to 7 days.

2.11 : Fusion protein purification and proteolysis

2.11.1 : Bacterial cell lysis

Cells harvested from 50ml of induced culture (2.11.2, 2.11.3) were resuspended in 1ml of 1 x PBS (2.1.2). *E.coli* strain BL21 cells (Appendix I) were then lysed by

sonication using an MSE Soniprep 150 sonicator. The cell suspension was subjected to 10 bursts of 10sec each (separated by 20sec intervals) at an approximate amplitude of 20μm using the micro-probe attachment of the sonicator. Successful lysis was apparent visually by a change in the colour and quality of the suspension from a milky pale yellow to a more translucent bronze/straw colour. MSD 2658 cells proved more refractory to lysis and were treated with up to 3 cycles of sonication as described above, interspersed with rapid freezing in a dry ice/ethanol bath followed by incubation at 37°C to rapidly rethaw the cell suspension. Following lysis, the cell sonicates were centrifuged at 10,000rpm for 10min at 4°C to pellet cell debris and other insoluble material. The soluble fraction was then decanted into a fresh tube.

2.11.2 : Purification of GST fusion proteins by glutathione affinity chromatography

A fresh overnight culture of an expression strain of *E.coli* (BL21 or MSD 2658, Appendix I) transformed with the pGEX-6P-2 (Appendix VI) derived plasmid of interest was used to inoculate 50 ml of LB (2.2.1) supplemented with carbenicillin (2.2.6). The cells were grown at 37°C or lower to log phase (2-4hr) then expression of the GST fusion protein was induced by the addition of 1mM IPTG. The cells were then harvested 4hr after induction by being centrifuged at 3,000 rpm for 5 min at 4°C and lysed by sonication, or sonication/freeze/thaw cycles, as appropriate (2.11.1).

A chromatography column with a glutathione sepharose 4B resin (Amersham Pharmacia Biotech) matrix, with a bed volume of e.g. 250μl per 0.5ml of sonicate, was then prepared as described in the manufacturer's protocol. The sonicate was added to the capped column and the whole was incubated for 30min at room temperature with gentle agitation in order to allow the fusion protein in the sonicate to bind to the matrix. The column was then drained and washed three times with 10 bed volumes of 1 x PBS (2.1.2). In order to elute the fusion protein, 1 bed volume of glutathione elution buffer (2.1.26) was added to the column, incubated at room temperature for 10min and then collected. This elution step was then repeated a further two times, and the three eluates pooled together.

2.11.3 : Purification of MBP fusion proteins by maltose affinity chromatography

A fresh overnight culture of an expression strain of *E.coli* (BL21 or MSD 2658, Appendix I) transformed with the pMAL-c2-G (Appendix VI) derived plasmid of interest was used to inoculate 50 ml of LB (2.2.1) supplemented with carbenicillin (2.2.6). The cells were grown at 37°C or lower to log phase (2-4hr) then expression of the MBP fusion protein was induced by the addition of 1mM IPTG. The cells were then harvested 4hr after induction by being centrifuged at 3,000 rpm for 5 min at 4°C and lysed by sonication, or sonication/freeze/thaw cycles, as appropriate (2.11.1).

A chromatography column with an amylose resin (New England BioLabs) matrix, with a total resin volume of e.g. 500µl per 0.5ml of sonicate, was then prepared at 4°C as described in the manufacturer's protocol. The sonicate was passed over the column to allow the fusion protein to bind to the matrix. The column was then washed three times with 4 resin volumes of maltose column buffer (2.1.27). In order to elute the fusion protein, 0.5 resin volumes of maltose elution buffer (2.1.28) were added to the column, incubated for 5min and then collected. This elution step was then repeated a further three times, and the four eluates pooled together.

2.11.4 : Dialysis of protein samples

Purified protein samples were dialysed, in order to change buffer conditions or remove low molecular weight compounds, in aliquots of up to 3ml using 10,000 molecular weight cut-off Slide-A-Lyzer dialysis cassettes (Pierce), according to the manufacturer's protocol. Cassettes were suspended at room temperature into 1000-2000ml of the constantly stirred buffer of choice for 3 hours or more, with one replacement of buffer at the half way point.

2.11.5 : Proteolysis of GST fusion proteins

Purified GST fusion proteins (2.11.2) were dialysed (2.11.4) against 1 x PBS (2.1.2) in order to remove glutathione. The dialysed protein was then added to a glutathione sepharose 4B chromatography column with a bed volume of e.g. 250µl per 0.75ml of purified protein sample, prepared as described in the manufacturer's protocol (Amersham Pharmacia Biotech) and incubated for 30min at room temperature with gentle agitation in order to allow the fusion protein to bind to the matrix. The column

was then drained and washed at 4°C with 10 bed volumes of GST cleavage buffer (2.1.29). One bed volume of PreScission protease (Amersham Pharmacia Biotech) at a concentration of 160U/ml in GST cleavage buffer was then added to the capped column and the whole was incubated for 4hr at 4°C with gentle agitation in order to allow proteolysis of the fusion protein to occur. The column was then drained and the eluate, ostensibly containing the polypeptide of interest, was retained. In order to elute the GST domain and the (GST-linked) protease, the column was re-equilibrated with 1 x PBS (2.1.2) then treated with glutathione elution buffer (2.1.26) as in 2.11.2.

2.11.6 : Proteolysis of MBP fusion proteins

Purified MBP fusion proteins (2.11.3) were dialysed (2.11.4) against genenase reaction buffer (2.1.30). Genenase I (New England BioLabs) was then added to an aliquot of dialysed protein to a final concentration of 10μg/ml. This mixture was incubated for 4-8hr at room temperature in order to allow proteolysis of the fusion protein to occur. The reaction mixture was then added to an amylose resin chromatography column, with a resin volume of e.g. 500μl per 1ml of purified protein sample, prepared as described in the manufacturer's protocol (New England BioLabs), at 4°C. The column was drained and the eluate, ostensibly containing the polypeptide of interest, was retained. In order to elute the MBP domain the column was washed several times with maltose column buffer (2.1.27) then treated with maltose elution buffer (2.1.28) as in 2.11.3.

2.12 : Western analysis

2.12.1 : Western blot

Proteins from unstained PAGE gels (2.6.3, 2.6.4) were transferred to nitrocellulose membrane (Schleicher and Schuell) using the Mini Trans-Blot blotting cell (Bio-Rad). The clear side of a transfer cassette was submerged in western transfer solution (2.1.31). A foam sponge was placed onto this, and a soaked piece of blotting paper onto the sponge. A piece of nitrocellulose membrane, cut to slightly larger than the separating gel, was then placed on top of the paper, so that it was positioned entirely in line with the grid of holes in the cassette wall. The separating gel, which had been

equilibrated for 30min in western transfer solution post-electrophoresis, was placed over the membrane, and finally another piece of soaked blotting paper and foam sponge were added on top. The cassette was then closed around the entire sandwich, and slotted into a blotting frame in a standard electrophoresis tank (Bio-Rad), so that the clear wall of the cassette faced the positive electrode of the frame. Western transfer solution was poured into the tank (which contained an ice-pack to prevent overheating), and transfer was conducted at 200mA for 2hr. Subsequently, the membrane was extricated from the sandwich, blotted dry and stored at room temperature, or frozen in western transfer solution until needed.

2.12.2 : Immunoprobing of blotted proteins

A membrane onto which proteins had been western blotted (2.12.1) was incubated in western blocking buffer (2.1.33) overnight at room temperature. The buffer was then discarded and replaced with the primary antibody of choice (specific to the antigen/epitope of interest) diluted appropriately (usually 1:500 – 1:10,000) in 20ml western blocking buffer. After incubation for at least 1hr at room temperature, the antibody mix was discarded and the membrane rinsed twice with western washing buffer (2.1.32) then washed for 2 x 15min in 20ml western washing buffer at room temperature. The washing buffer was then replaced with alkaline phosphatase (AP) conjugated secondary antibody (specific against the antibody class of the primary antibody, for example, if the primary antibody is mouse IgG, the secondary antibody will be e.g. rabbit anti-mouse IgG-AP conjugate) diluted appropriately (usually 1:5,000 – 1:10,000) in 20ml western blocking buffer. After incubation for at least 1hr at room temperature, the antibody mix was discarded and the membrane rinsed and washed with western washing buffer as before. The membrane was then equilibrated in AP buffer (2.1.34) for 5 min, then incubated with 20ml AP buffer containing the colorimetric reagents NBT and BCIP (Promega, 132µl and 66µl respectively). Reactive areas of the blot, theoretically corresponding to the location of the antigen/epitope of interest, turned purple within 5min-1hr, depending on the antibodies used and the quantity of antigen present. The appearance of this purple coloration at particular loci is referred to in the text as 'lighting up'. The colour reaction was stopped by thoroughly rinsing the membrane in sdH_2O.

2.12.3 : Direct staining of blotted proteins

Two reagents, ProtoGold (British BioCell International) and Ponceau S were used to stain membranes onto which proteins had been western blotted (2.12.1) for total protein. For ProtoGold staining the membrane was washed for 2 x 5min in 1 x PBS (2.1.2), then incubated in western washing solution (2.1.32) at room temperature for 30min. It was then washed for 3 x 5min in western washing solution and 1 x 5min in sdH_2O. The membrane was finally incubated in the proprietary ProtoGold solution for 5min-1hr. The staining was stopped by thoroughly rinsing the membrane in sdH_2O. For Ponceau S staining the membrane was washed for 5min in 5% (v/v) acetic acid then incubated for 20sec in a 0.1% (w/v) solution of Ponceau S in the same. Excess stain was washed away using sdH_2O, and the membrane was then blotted dry.

2.13 : Immunological techniques

2.13.1 : Immunisation of experimental mice with antigen extracted from nitrocellulose fragments

An aliquot of antigen, dialysed against sdH_2O to remove salts, was frozen in a dry ice/ethanol bath and lyophilised in a refrigerated vacuum drier. The lyophilised protein was then resuspended in '1 x' SDS-PAGE reducing/loading buffer (2.1.11), loaded onto an SDS-PAGE gel (2.6.3) alongside an aliquot of pre-stained molecular weight markers and subjected to electrophoresis. Protein was then transferred from the gel (unstained) to nitro-cellulose membrane by western blot (2.12.1), and the membrane was stained for total protein with Ponceau S (2.12.3). Stained bands containing the protein of interest were then excised from the membrane using a sterile scalpel blade and placed in a 1.5ml Eppendorf tube. DMSO was added to the membrane fragments to give an approximate antigen concentration of 400-500 μg/ml and the whole was vortexed to dissolve the nitro-cellulose matrix.

The DMSO/antigen/nitro-cellulose mixture was subsequently delivered to the laboratories of Eurogentec (Herstal, Belgium). Here it was mixed with an equal volume of Freund's complete adjuvant (Freund, 1947). Experimental mice were then inoculated with 100μl each of the antigen preparation (equivalent to 40-50μg antigen) through intramuscular injection. The injections were repeated at 14 days and 28 days after the

initial inoculation. 350µl of serum was bled from each mouse 10 days after the third injection.

2.13.2 : Depletion of mouse serum for antibodies against *E.coli* and nitrocellulose

100µl of serum was added to 900µl of sonicate (2.11.1) from an untransformed *E.coli* strain and shaken gently at room temperature in the presence of a disc of nitro-cellulose for 1hr (with several changes of nitro-cellulose disc).

2.13.3 : Immunisation of experimental chickens with lyophilised antigen

An aliquot of antigen, dialysed against 1 x PBS (2.1.2), was frozen in a dry ice/ethanol bath then lyophilised in a refrigerated vacuum drier. The lyophilised antigen was subsequently delivered to the laboratories of Eurogentec (Herstal, Belgium). Here it was resuspended in 50% (v/v) Freund's adjuvant. Experimental chickens were then injected with an aliquot of the antigen preparation equivalent to 40-50µg antigen. The injections were repeated at 14 days and 28 days after the initial inoculation. Eggs were collected daily from 5 days after the third injection.

2.13.4 : Partial purification of IgY from chicken egg yolk

Chicken IgY was purified from egg yolk according to Polson *et al.* (1985). Yolk was homogenised in approximately 4 volumes of sterile 1 x PBS (2.1.2) containing 4.375% (w/v) PEG6000 (to give a final PEG concentration of 3.5%) using a hand homogeniser, then incubated for 20min at room temperature. The homogenate was then centrifuged at 6,000rpm for 30min at room temperature, and the supernatant was decanted into a fresh tube. This centrifugation was repeated until as much insoluble material as possible had been removed from the homogenate. Approximately 0.22 volumes of 50% (w/v) PEG6000 in PBS was then added to the clarified supernatant to give a final PEG concentration of 12% (w/v). This preparation was then centrifuged as before and the supernatant discarded. The precipitate was resuspended in 5 yolk volumes of 12% (w/v) PEG6000, then centrifuged again as before, with the supernatant being discarded. This resuspension/centrifugation step was repeated and finally the precipitate, containing partially purified IgY, was resuspended in 1 yolk volume of sterile 1 x PBS. This was stored at 4°C for up to 1 month or in aliquots at -20°C for longer periods.

CHAPTER 3

EXPRESSION OF THE GTP CYCLOHYDROLASE I GENE FROM *PLASMODIUM FALCIPARUM* IN *ESCHERICHIA COLI*

3.1 : Introduction

The bacterium *Escherichia coli* is the most widely used host for expression of cloned genes. It is fast growing and exceptionally easy to maintain in culture, while an extensive range of techniques for biochemical and genetic manipulation of the organism have been perfected. Expression of foreign genes can often proceed at very high levels. For these reasons, *E.coli* was chosen as a convenient system for expression of the *Pf gch-1* gene, and fragments thereof, with a view to producing significant levels of functional protein, as well as using expressed gene product to raise antibodies for possible cellular localisation studies. The expression vector of choice was initially pET-15b (Novagen) which produces gene product fused to an N-terminal hexahistidine tag, facilitating purification on an affinity chromatography column containing nickel particles. However, when attempts to produce clones using this vector proved unsuccessful, it was resolved to attempt to clone the gene into the pGEX-6P-2 expression vector (Appendix VI), which expresses gene product as a fusion with a cleavable N-terminal glutathione S-transferase (GST) tag.

3.2 : Production of clones for expression of GST fusion proteins

Four oligonucleotide primers, gtpgex1, gtpgex2, gtpgex3 and gtpgex4 (Appendix V) were designed to produce *Pf gch-1* derived DNA fragments with terminal *Eco* RI and *Not* I sites suitable for unidirectional cloning into pGEX-6P-2.

The 3 inserts produced (Figure 3.1A) were as follows:-

an approximately 1200bp product amplified by gtpgex1 and gtpgex4, corresponding to the full length *Pf gch-1* gene, hereby known as insert gch, an approximately 550bp product amplified by gtpgex1 and gtpgex3, corresponding to the N-terminal half of the gene product (Ngch), and an approximately 600bp product amplified by gtpgex2 and gtpgex4, corresponding to the C-terminal half of the gene

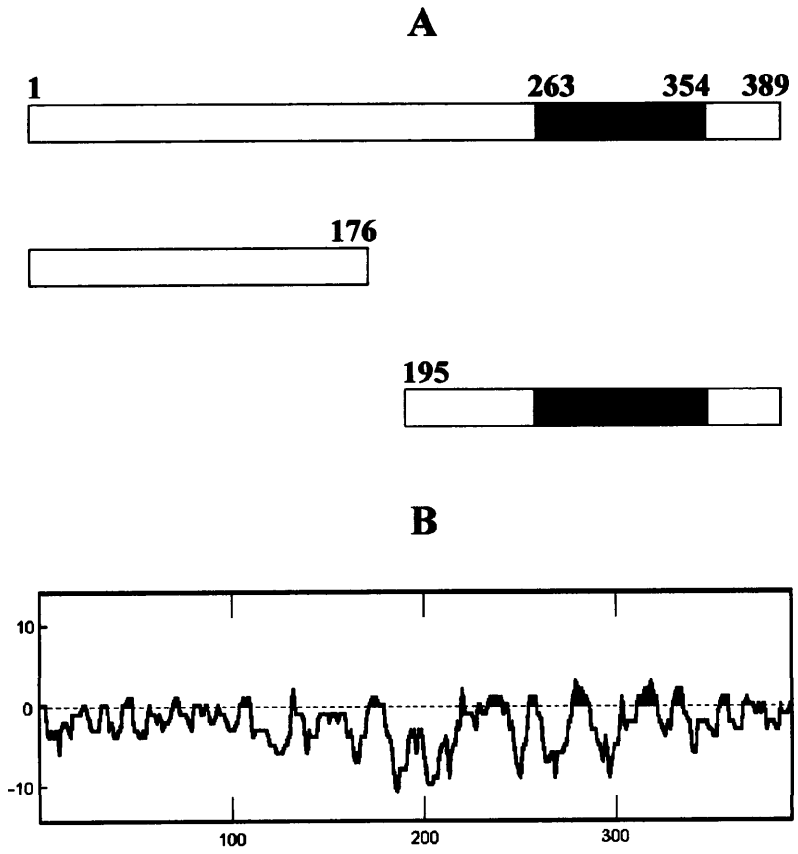


Figure 3.1 A) *Pf gch-1* derived inserts for use with pGEX-6P-2
Upper, gch; middle, Ngch; lower, Cgch. Black box indicates the conserved 'core' region of the gene. Numbers indicate corresponding amino acids in polypeptide sequence.
B) Antigenicity of *P.falciparum* GTCase
X axis, amino acid number; Y axis, measure of antigenicity. 'Positive' peaks are shaded in red. (Theoretical plot, generated using AnTheProt software).

product (Cgch.) The two shorter inserts were chosen due to their potential to act as antigens in the production of antibodies specific to *Pf* GTCase; Ngch as this part of the gene is absent or poorly conserved in other species (Appendix III), Cgch on the basis of its comparatively high antigenic potential (Figure 3.1B)

PCRs (2.7.1) were set up using a clone of *Pf gch-1* cDNA in pBluescript (pBS; Stratagene), and the three different primer combinations. The amplification products were visualised on an agarose gel (2.6.1) and purified from the reaction mixes (2.3.1), then digested with the cloning enzymes (2.4.1), as was the vector. All three constructs were subsequently produced by cloning (2.5.1)(see Figure 3.2 for the isolation of pGEX-Cgch clones).

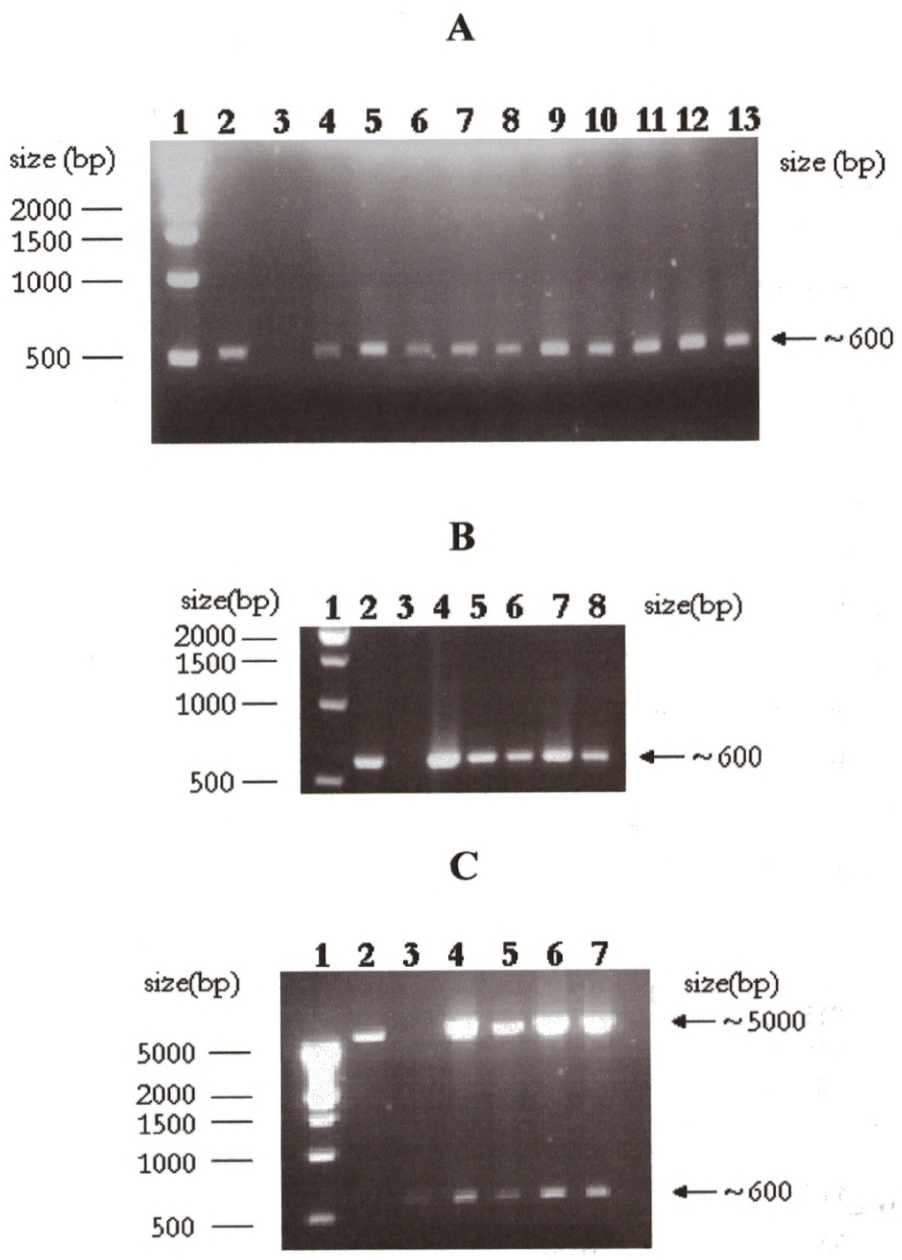


Figure 3.2 Isolation and identification of pGEX-Cgch clones

A) Primary PCR mixes (colonies screened with *Pf gch-1* specific primers); lane numbers: 1, DNA marker XVII (Boehringer Mannheim); 2, positive control PCR (clone of *Pf gch-1* cDNA in pBS used as template); 3, negative control PCR (colony transformed with vector only used as template); 4-13, putative pGEX-Cgch clones 1-50 in batches of 5.

B) Secondary PCR mixes; lane numbers: 1, DNA marker XVII; 2, positive control PCR; 3, negative control PCR; 4-8, putative pGEX-Cgch clones 6-10.

C) *Eco* RI/*Not* I digests of pGEX-Cgch clones; lane numbers: 1, DNA marker XVII; 2, linearised pGEX-6P-2; 3, gtpgex2/gtpgex4 PCR product; 4-7, clones 6-9.

3.3 : Expression of the pGEX-6P-2 derived constructs

Each of the 3 constructs and pGEX-6P-2 were transformed into cells of *E.coli* strain BL21 (Appendix I) using $CaCl_2$ methodology (2.5.2, 2.5.3). 50ml log-phase cultures of each transformant were induced at 37°C to produce fusion protein through addition of 1mM IPTG. The cells were harvested by centrifugation approximately 4 hours after induction (see 2.11.2). Small quantities of cells (harvested from 1.5ml of induced culture) were resuspended in 250µl '1 x' SDS-PAGE reducing/loading buffer (2.1.11), boiled to lyse and subjected to SDS-PAGE (2.6.3). The gel was subsequently stained for total protein with Coomassie Blue (2.1.16)(Figure 3.3).

The cells transformed with pGEX-6P-2 alone produced large quantities of a protein of around 31 kDa in size, almost certainly corresponding to the GST fusion domain (26 kDa.) Although no obvious fusion protein was visible in the lysate of cells transformed with pGEX-gch, the lysates of the cells transformed with pGEX-Ngch and pGEX-Cgch showed possible fusion proteins of approximately 34 kDa and 56 kDa respectively. In the case of pGEX-Ngch this visible product was almost certainly not a full length GST-Ngch fusion, rather a truncation of the same, as indicated by the relatively small difference in running distance between this species and the native GST

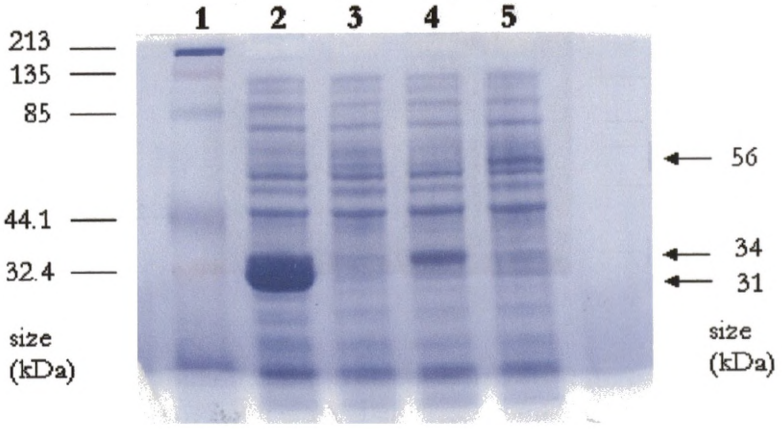


Figure 3.3 Crude lysates of induced BL21 pGEX clones
Lane numbers: 1, Kaleidoscope molecular weight markers (Bio-Rad); 2, BL21 pGEX-6P-2 vector only cells; 3, BL21 pGEX-gch cells; 4, BL21 pGEX-Ngch cells; 5, BL21 pGEX-Cgch cells. n.b. the volume of each pGEX expression-related protein sample subjected to SDS-PAGE is equivalent to between 50 and 100µl of original induced culture unless otherwise stated.

produced by pGEX-6P-2. In the case of pGEX-Cgch however, the visible product ran to approximately the distance expected of a full length GST-Cgch fusion (49 kDa).

3.4 : Purification of GST & GST-gch fusion proteins

In order to test the glutathione-affinity protein purification protocol (2.11.2) it was decided to carry out an initial trial purification of GST alone. The remainder of the BL21 pGEX-6P-2 cells which had been harvested following induction (3.3) were resuspended in 1ml PBS (2.1.2) and subsequently lysed by sonication (2.11.1). This lysate was separated into soluble and insoluble fractions by centrifugation. Small quantities of each fraction were subjected to SDS-PAGE (2.6.3, equivalent results shown in Figure 3.4). The majority of native GST was found to be present in the soluble fraction. It was therefore resolved to purify the GST from the soluble fraction using glutathione affinity chromatography (2.11.2). This resulted in purification of the vast majority of soluble GST, with very little undesirable co-purification (Figure 3.4).

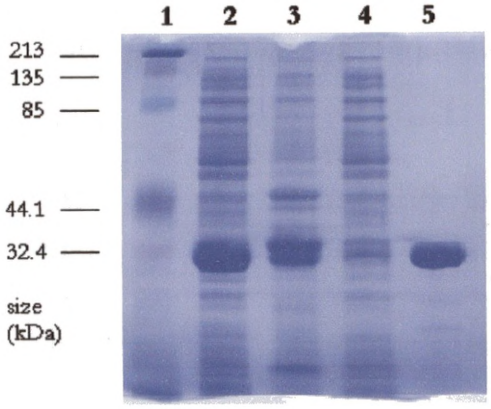


Figure 3.4 Affinity purification of GST
Lane numbers: 1, Kaleidoscope molecular weight markers; 2, BL21 pGEX-6P-2 soluble fraction; 3, BL21 pGEX-6P-2 insoluble fraction; 4, BL21 pGEX-6P-2 pre-wash eluate; 5, BL21 pGEX-6P-2 post-column eluate.

The remainder of the induced BL21 cells transformed with the pGEX-gch constructs were resuspended in 1ml PBS each and subsequently lysed by sonication. The lysates were separated into soluble and insoluble fractions by centrifugation as before. Small quantities of each fraction were subjected to SDS-PAGE (Figure 3.5)

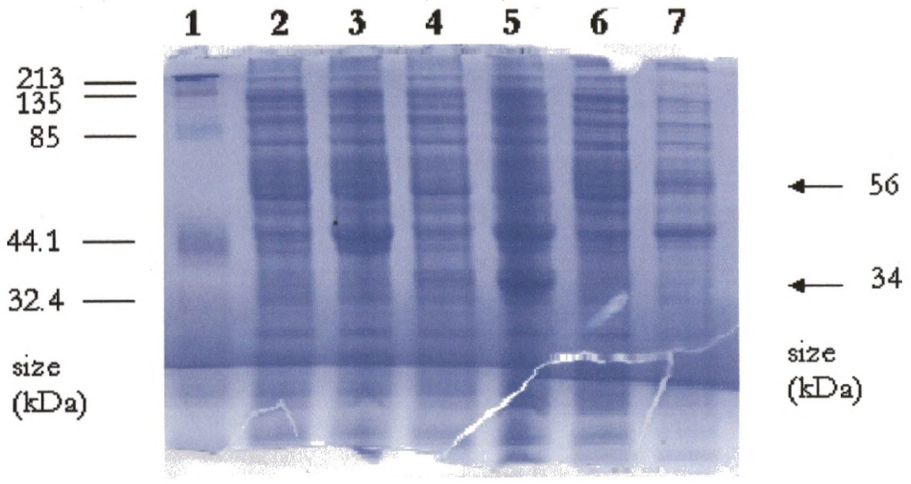


Figure 3.5 Solubility of proteins in pGEX-gch transformants grown at 37°C
Lane numbers: 1, Kaleidoscope molecular weight markers; 2, BL21 pGEX-gch soluble fraction; 3, BL21 pGEX-gch insoluble fraction; 4, BL21 pGEX-Ngch soluble fraction; 5, BL21 pGEX-Ngch insoluble fraction; 6, BL21 pGEX-Cgch soluble fraction; 7, BL21 pGEX-Cgch insoluble fraction.

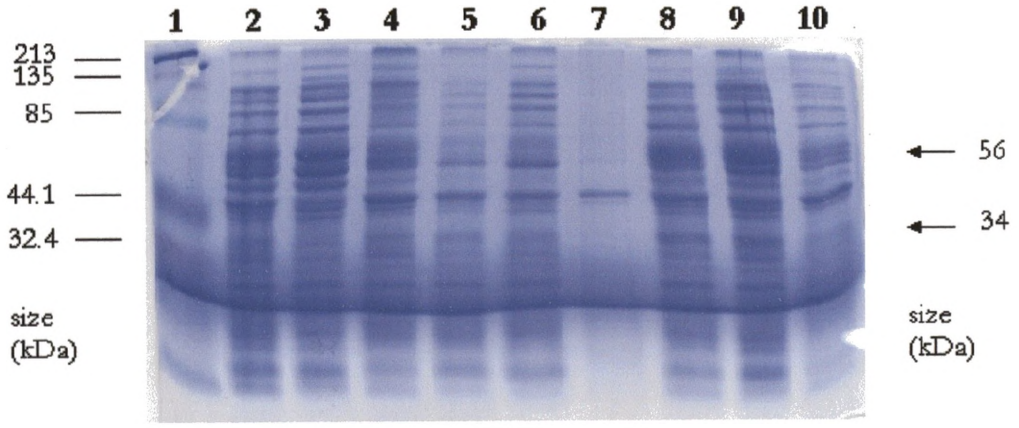


Figure 3.6 Solubility of proteins in pGEX-gch transformants grown at 30°C
Lane numbers: 1, Kaleidoscope molecular weight markers; 2, BL21 pGEX-gch total protein; 3, BL21 pGEX-gch soluble fraction; 4, BL21 pGEX-gch insoluble fraction; 5, BL21 pGEX-Ngch total protein; 6, BL21 pGEX-Ngch soluble fraction; 7, BL21 pGEX-Ngch insoluble fraction; 8, BL21 pGEX-Cgch total protein; 9, BL21 pGEX-Cgch soluble fraction; 10, BL21 pGEX-Cgch insoluble fraction.

As significant quantities of protein, particularly the putative fusion proteins, were present in the insoluble fraction, it was decided to repeat the induction at 30°C, with the intention of increasing protein solubility. The BL21 cells were induced at 30°C,

then harvested, resuspended and lysed, and small aliquots subjected to SDS-PAGE as before (Figure 3.6). A general improvement in solubility was noted. The BL21 cells transformed with pGEX-Ngch showed a notably low yield of protein in this case however, probably due to unusually low numbers of cells in the induced culture. The induction of this clone was repeated therefore, and normal cell growth and protein yields were observed (results not shown).

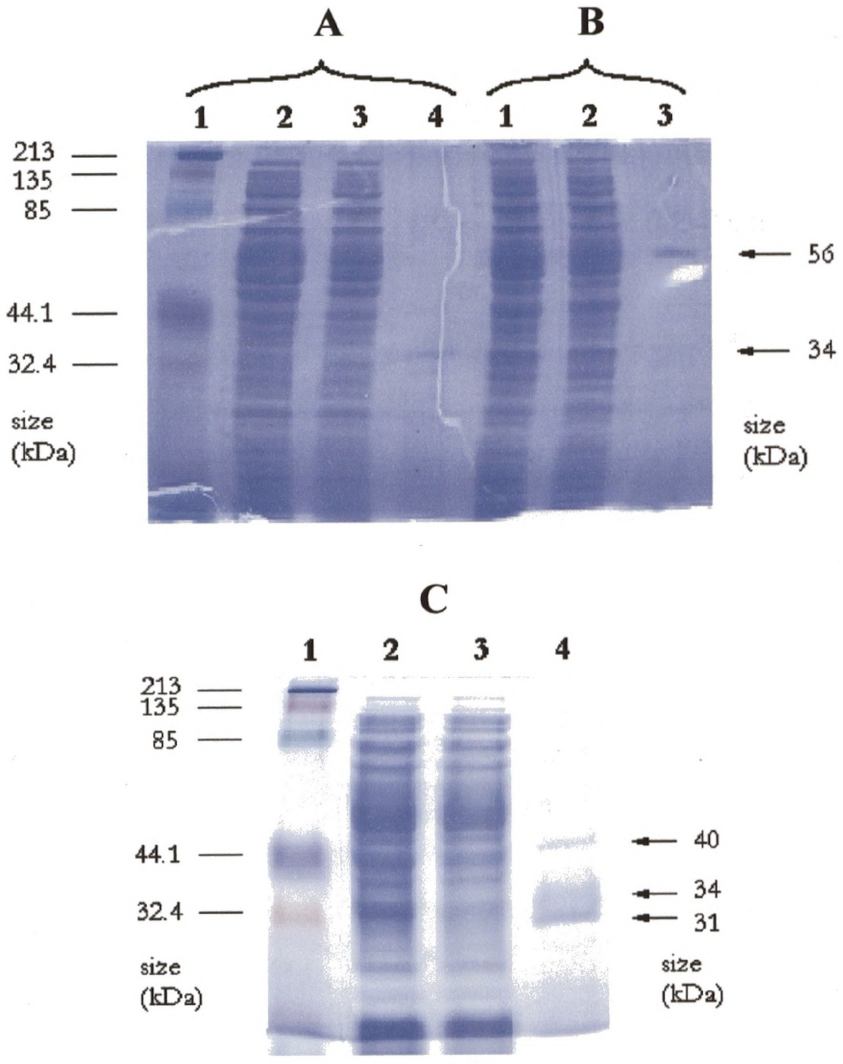


Figure 3.7 Purification of putative GST-gch fusion proteins
A) GST-gch; Lane numbers: 1, Kaleidoscope molecular weight markers; 2, BL21 pGEX-gch soluble fraction; 3, BL21 pGEX-gch pre-wash eluate; 4, BL21 pGEX-gch post-column eluate. **B) GST-Cgch;** 1, BL21 pGEX-Cgch soluble fraction; 2, BL21 pGEX-Cgch pre-wash eluate; 3, BL21 pGEX-Cgch post-column eluate. **C) GST-Ngch** 1, Kaleidoscope molecular weight markers; 2, BL21 pGEX-Ngch soluble fraction; 3, BL21 pGEX-Ngch pre-wash eluate; 4, BL21 pGEX-Ngch post-column eluate.

500µl of each of the 3 soluble fractions was subjected to affinity column purification (2.11.2, Figure 3.7). In the case of the BL21 pGEX-gch soluble fraction (Figure 3.7A) a polypeptide of around 34kDa was observed in the post-column eluate. In the BL21 pGEX-Ngch post-column eluate (Figure 3.7C) a polypeptide of approximately 34 kDa was present as anticipated, along with two additional polypeptides of around 31 kDa and 40 kDa. In the BL21 pGEX-Cgch post-column eluate (Figure 3.7B) an approximately 56 kDa polypeptide was present as anticipated.

To more accurately gauge the sizes of the purified polypeptides, small aliquots were subjected to SDS-PAGE alongside Precision Protein Standards (Bio-Rad), which are explicitly recommended for polypeptide molecular weight comparisons (Figure 3.8).

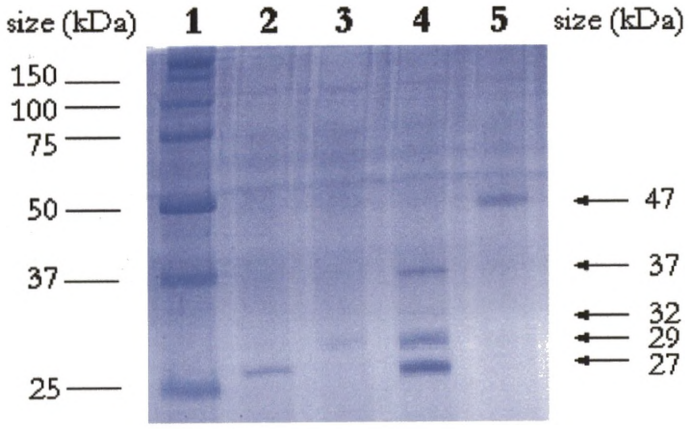


Figure 3.8 Gel for putative GST-gch fusion protein size calculation
Lane numbers; 1, Precision Protein Standards; 2, GST; 3, BL21 pGEX-gch post-column eluate; 4, BL21 pGEX-Ngch post-column eluate; 5, BL21 pGEX-Cgch post-column eluate.

The calculated sizes of GST and each of the putative fusion proteins is shown in Table 3.1. As anticipated, GST runs very close to its expected length of 26 kDa, while the polypeptides visible in the post-column eluates of the BL21 pGEX-gch and pGEX-Ngch cells (including an additional polypeptide of around 32 kDa in the post-column eluate of the BL21 pGEX-Ngch cells, revealed as a result of the increased clarity of the gel) appear to be significantly, if not massively, smaller than the respective theoretical fusion proteins. Encouragingly however, the discrepancy between the expected (49 kDa) and recorded (47 kDa) sizes of the theoretical and putative GST-Cgch fusions respectively is small, lending weight to the theory that the observed species is the full length expression product expected from the pGEX-Cgch construct.

Fusion protein	Expected molecular weight	Recorded molecular weight(s)
GST	26 kDa	27 kDa
GST-gch	72 kDa	29 kDa
GST-Ngch	47 kDa	27 kDa, 29 kDa, 32 kDa, 37 kDa
GST-Cgch	49 kDa	47 kDa

Table 3.1 Comparison of sizes of genuine and putative GST-gch fusion proteins
Theoretical molecular weights calculated using Seqaid II software. Putative fusion-protein sizes calculated using DNAfrag W software.

3.5 : Western blot analysis of the putative GST-gch fusion proteins

In order to confirm the origin of the polypeptides purified in 3.4, it was resolved to subject them to Western blot analysis (2.12) using an antibody specific to the GST domain produced by the pGEX vector. Proteins were transferred to nitro-cellulose membrane (2.12.1) from a polyacrylamide gel equivalent to Figure 3.9A. Following initial washes, the membrane was incubated first with goat anti-GST antibody, then with AP conjugated donkey anti-goat IgG antibody, and developed with the colorimetric agents NBT and BCIP (2.12.2). The pattern of bands observed after development is shown in Figure 3.9B.

The native GST produced by pGEX-6P-2 was seen to 'light up' in lane 2. Similarly, the 29 kDa putative fusion protein(s) lit up where present, in both the total protein and post-column eluate fractions of the BL21 pGEXgch and pGEX-Ngch cells alike. The larger 32 kDa and 37 kDa polypeptides from the pGEX-Ngch post-column eluate also lit up faintly (not visible in the reproduction), and faint bands were also seen at this position in lane 5, BL21 pGEX-Ngch total protein, as would be expected. The shortest polypeptide visible in the pGEX-Ngch post-column eluate, that of 27 kDa, lit up also. No equivalent band was seen in lane 5 however. Therefore, the presence of this band in the pGEX-Ngch post-column eluate was most likely due to contamination of the column in question with residual GST during the purification of this species from the soluble fraction of the BL21 pGEX-6P-2 cells, despite extensive washing of the

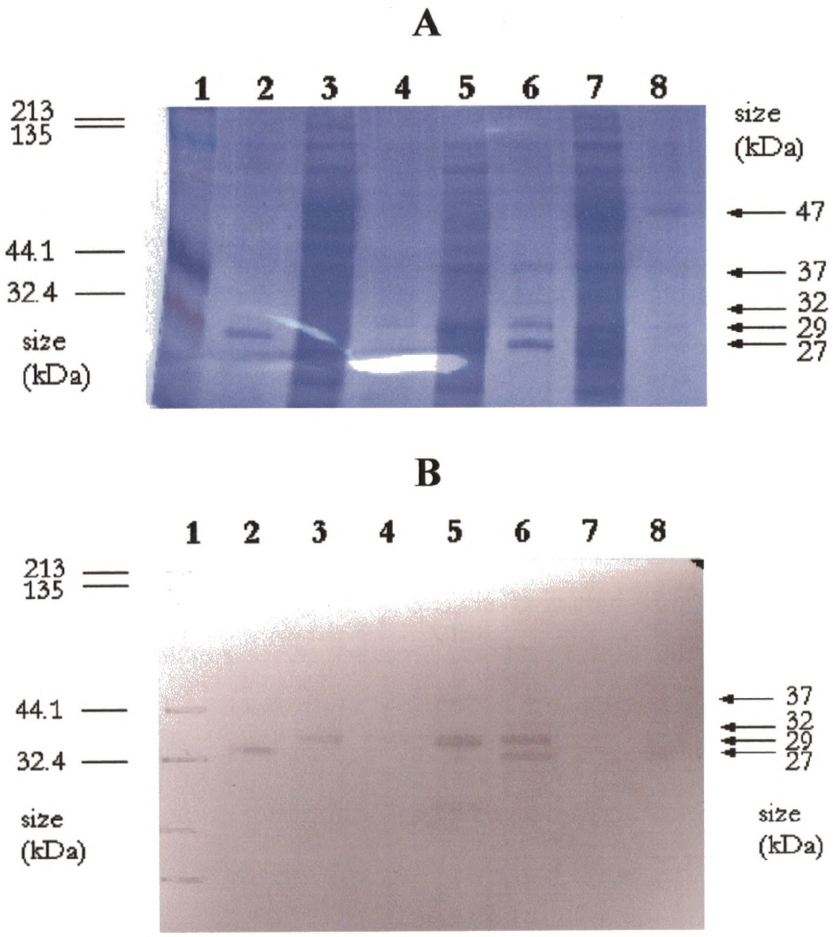


Figure 3.9 Western blot of putative GST-gch fusion proteins
A) Equivalent of original polyacrylamide gel; Lane numbers: 1, Kaleidoscope molecular weight marker; 2, GST (level attenuated to approx. 1/20 usual level;) 3, BL21 pGEX-gch soluble fraction; 4, BL21 pGEX-gch post-column eluate; 5, BL21 pGEX-Ngch soluble fraction; 6, BL21 pGEX-Ngch post-column eluate; 7, BL21 pGEX-Cgch soluble fraction; 8, BL21 pGEX-Cgch post-column eluate.
B) Developed membrane; Lane numbers: as A). n.b. positions of markers were added immediately post-transfer, also polypeptide sizes marked at the right of the figure correspond to those listed in Table 3.1.

column (as per the manufacturer's protocol) in order to regenerate the resin. Unexpectedly, the putative full-length GST-Cgch fusion protein present in the pGEX-Cgch post-column eluate (47 kDa) did not light up. This observation was repeated in further western blotting experiments using greatly enhanced quantities of the putative GST-Cgch (results not shown). This meant designation of this species as produced from the GST-Cgch open reading frame (ORF) present on the pGEX-Cgch construct was not possible by this means.

In an attempt to circumvent this problem, it was decided to subject the BL21 pGEX-Cgch post-column eluate to non-denaturing (2.6.4), rather than SDS (denaturing), PAGE before Western blotting; although the goat anti-GST antibody preparation used is explicitly recommended for use against denatured GST, the possibility remains that some antibodies against conformational epitopes will be present in the preparation. In addition, the simplicity of this alternative approach rendered it relatively attractive.

Proteins were transferred to nitro-cellulose membrane from a non-denaturing polyacrylamide gel equivalent to Figure 3.10A. The membrane was washed and incubated with the primary and secondary antibodies, then developed with NBT and BCIP. The pattern of bands observed after development is shown in Figure 3.10B.

Surprisingly, a comparatively faint but nonetheless clear band lit up from the BL21 pGEX-Cgch post-column eluate at a position on the membrane corresponding to the very upper edge of the gel, which is where the densest band of protein appeared (Figure 3.10A).

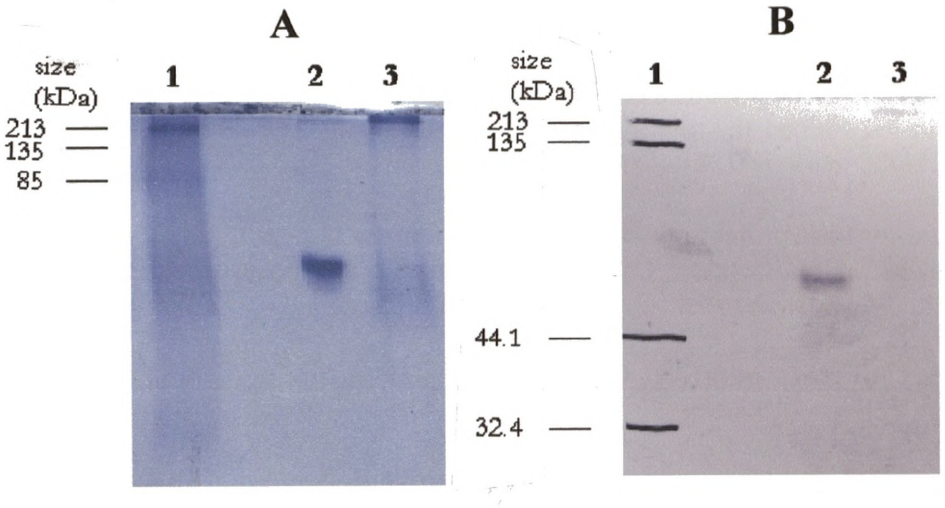


Figure 3.10 Western blot of putative GST-Cgch fusion protein transferred from a non-denaturing gel A) Equivalent of original polyacrylamide gel; Lane numbers: 1, Kaleidoscope molecular weight marker (used as a distance marker in this case;) 2, GST; 3, BL21 pGEX-Cgch post-column eluate (10-20 x quantity used in actual gel to aid visualisation).
B) Developed membrane; Lane numbers: as A). n.b. positions of markers were added immediately post-transfer.

It was suspected that the seemingly anomalous negative result produced when immunoprobing the putative GST-Cgch fusion protein following western blot from denaturing gels could be a result of absence of this species bound to the nitro-cellulose

membrane, due to either inefficient or over-efficient transfer. To test this hypothesis a gel identical to Figure 3.11A was prepared, and subsequently subjected to Western transfer. The gel was subsequently stained with Coomassie blue (2.1.16) for post-transfer total protein (Figure 3.11B), while the membrane was stained for total protein with ProtoGold (2.12.3)(Figure 3.11C). These results clearly showed that efficient transfer of the putative fusion protein from the gel to the membrane was occurring.

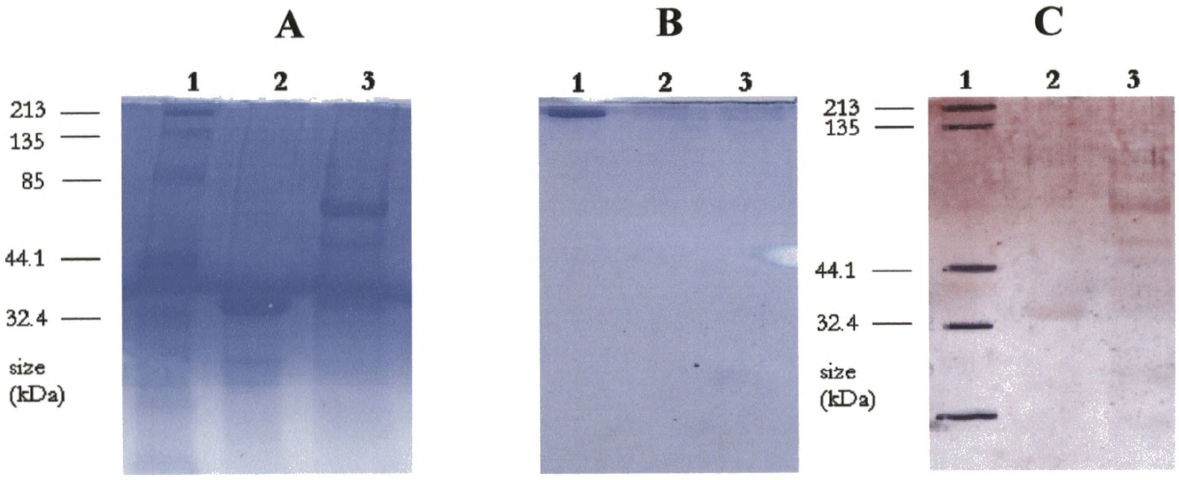


Figure 3.11 Transfer of putative GST-Cgch fusion protein from gel to membrane
A) Equivalent gel stained for total protein pre-transfer; Lane numbers: 1, Kaleidoscope molecular weight marker; 2, GST; 3, enhanced level BL21 pGEX-Cgch post-column eluate.
B) Gel stained for total protein post-transfer; Lane numbers: as A).
C) Blot stained for total protein; Lane numbers: as A). n.b. positions of markers were added immediately post-transfer.

3.6: Analysis of the putative GST-Cgch fusion protein using peptide-mass fingerprinting

In order to circumvent the problems in positively identifying the putative GST-Cgch fusion as a result of its failure to light up when probed with anti-GST antibodies (3.5), it was decided to subject the species to peptide-mass fingerprinting analysis, whereby the peptide fragments produced by a proteolytic digest of the protein are identified by matrix-assisted laser desorption/ionisation time of flight (MALDI-TOF) mass spectroscopy (Jonsson, 2001).

An SDS-PAGE gel and associated buffers (2.6.3) were prepared using de-ionised water and fresh, molecular biology grade reagents, in order to minimise the likelihood of extraneous protein contamination. An aliquot of the BL21 pGEX-Cgch

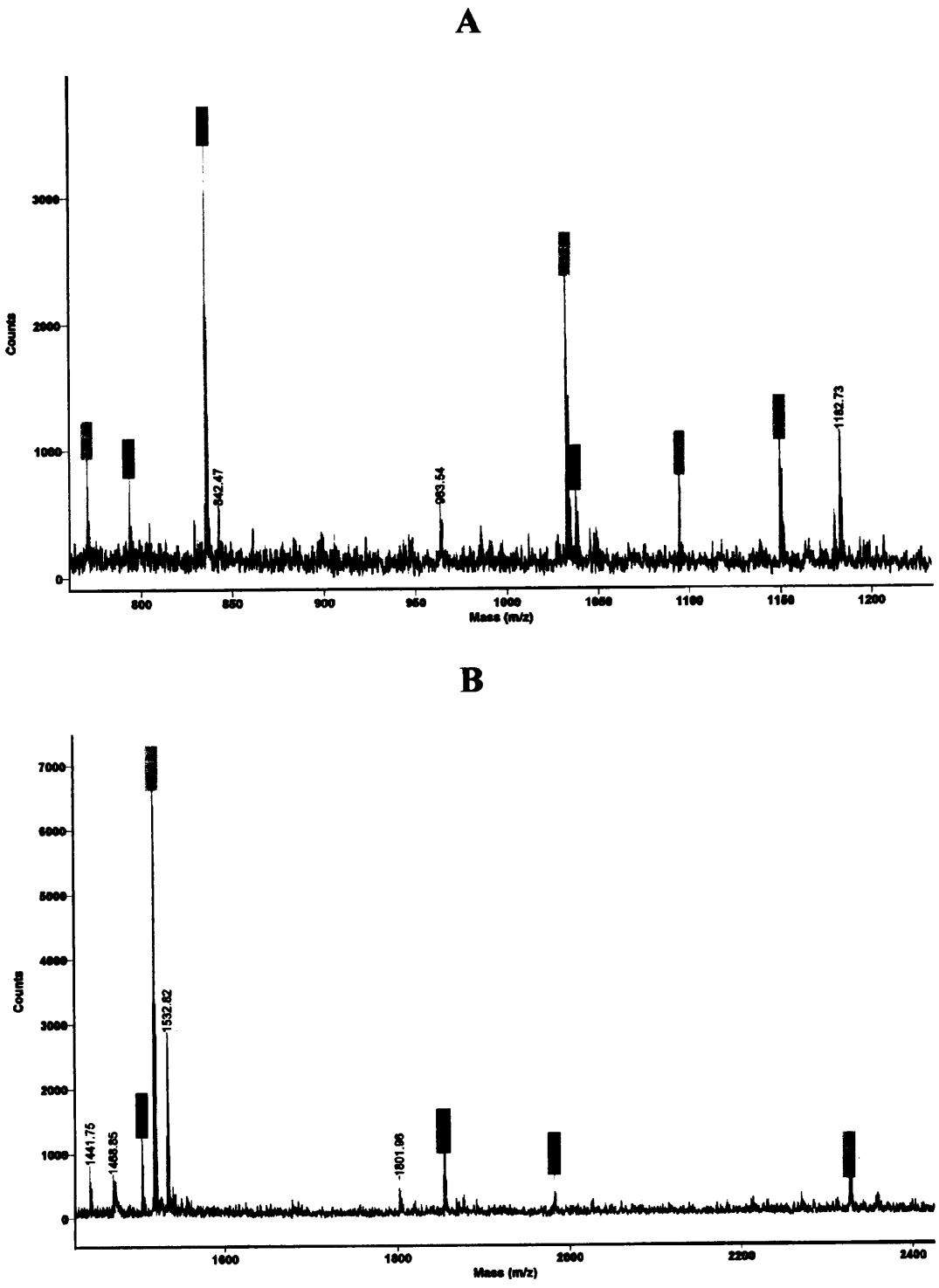


Figure 3.12 Mass spectrograms of a tryptic digest of the putative GST-Cgch fusion protein A; data in the range of 760 to 1230 Da, B; data in the range of 1400 to 2400 Da. Those peaks highlighted in yellow match theoretical digestion fragments from the GST domain and those in blue match fragments from the Cgch domain. The peak highlighted in red matches the fragment incorporating the junction of the two domains (see Table 3.2). n.b. digestion fragments are protonated by the MALDI-TOF procedure; a recorded mass of *x*Da therefore corresponds to a peptide mass of (*x-1*)Da.

post-column eluate (3.4) was subjected to electrophoresis and the gel stained with Coomassie blue solution (2.1.16), which had also been prepared as above. A fragment of gel containing the approximately 47 kDa polypeptide of interest was then excised with a sterile scalpel blade and treated with 100ng of the endopeptidase trypsin in order to cleave the polypeptide at all peptide bonds in which the carbonyl moiety was provided by a lysine or arginine residue. The peptide mixture was eluted from the polyacrylamide matrix and then subjected to MALDI-TOF mass spectroscopy at an accelerating voltage of 24,000V. The resulting mass spectrograms are shown in Figure 3.12.

•	2325.133	YIAWPLQGWQATFGGGDHPPK	2325.13
	2228.104	FELGLEFPNLPYYIDGDVK	
	2125.030	HHLLPFEGTCDIEYIPNK	
•	1980.939	ENPTVHSLNIDSSVENLN	1980.91
•	1852.994	SDLEVLFQGPLGSPGIPK	1852.97
	1829.919	LQLQEDLTNDICNALK	
•	1515.797	AEISMLEGAVLDIR	1515.84
•	1501.763	SNDIEEQIINISK	1501.80
	1232.659	VTGIHIYSLCK	
•	1148.633	LLLEYLEEK	1148.65
	1137.509	YEEHLYER	
•	1093.563	MSPILGYWK	1093.58
•	1036.632	YLKPLYIK	1036.60
•	1031.580	LTQSMAIIR	1031.60
	1025.576	IEAIPQIDK	
	955.437	HNMLGGCPK	
	945.481	TITYASYK	
	885.431	HLCINMR	
•	834.460	IVDVFSR	834.44
•	792.475	YIIGLSK	792.43
•	769.445	GLVQPTR	769.40

Table 3.2 Peptides of between 760 and 2400 Da generated by a theoretical digest of the GST-Cgch fusion protein Theoretical digest performed using the Masstools suite of programs (under development) provided privately by Dr. Simon Hubbard, UMIST. Those peptides highlighted in yellow derive from the GST domain and those in blue from the Cgch domain. The peptide highlighted in red crosses the junction of the two domains. Peptides marked with bullet points were positively matched to peaks in the mass spectrograms, whose (m/z-1) values are shown at the right (Figure 3.12).

12 out of 21 theoretical digestion fragments in the mass range over which data was collected were represented by peaks in the mass spectrogram (Table 3.2). These included not only the peptide incorporating the junction of the two domains of the GST-

Cgch fusion protein, but also the peptides derived from both termini of the polypeptide. In addition, only 7 out of the 19 peaks recorded by the mass detector had no counterpart in the set of theoretical peptides. Bioinformatic analysis indicated that the pattern of peaks produced was in excess of $1x10^{26}$ times more likely to have been produced by GST-Cgch than any real or putative *E.coli* protein (Table 3.3), the most likely source of any contamination. It is therefore reasonable to conclude that the candidate protein is indeed GST-Cgch, and that the Cgch part of the fusion continues to the natural stop codon.

Rank	Protein name/description	Peptide matches (x/19)	Protein length (aa)	Coverage (%)	Score
1	**GST-Cgch**	**12**	**430**	**32.6**	**1.000e+00**
2	dcp, dipeptidyl carboxypeptidase	3	681	4.6	5.270e-27
3	hypothetical protein	3	879	4.6	2.124e-27
4	cof, cof protein	2	276	8.3	5.510e-29
5	ybiK, hypothetical protein	2	321	5.9	4.664e-29
6	ydiR, putative electron transfer flavoprotein subunit	2	312	5.4	3.466e-29
7	dacA, penicillin-binding protein 5 precursor	2	403	5.7	1.625e-29
8	yadE, hypothetical protein	2	409	3.4	1.265e-29
9	ygeH, hypothetical protein	2	458	3.5	1.265e-29
10	yojI, hypothetical ABC transporter	2	547	3.5	1.166e-29

Table 3.3 Relative probabilities of possible identities of the candidate fusion protein The m/z values recorded in the MALDI-TOF experiment (Figure 3.12) were compared with a database of theoretical trypsin digest fragments of GST-Cgch and all proteins and putative proteins encoded by the *E.coli* genome using PepMapper software provided privately by Dr. Simon Hubbard, UMIST. Theoretical peptides with a mass deviation of <0.05 Da from the masses inferred by the m/z values were counted as matches. Coverage indicates the percentage of that protein's length represented by the matched peptides. Scores are indicators of relative probability.

3.7 : Production of antibodies against GST-Cgch

Following positive identification of the approx. 47 kDa GST-Cgch fusion protein (3.6) it was decided to use this species in an attempt to raise antibodies active against the native *Pf* GTCase. Such antibodies, given a sufficiently strong antibody/antigen reaction, could subsequently prove a useful tool in a variety of analyses, for example as an affinity reagent in the purification of the native enzyme

from the parasite, or in cellular localisation studies. The injection of rabbits with the *Pf* valosin protein extracted from nitro-cellulose fragments excised from a western blot (2.12.1) had yielded antiserum with high reactivity against the source antigen. The serum had exhibited extensive, seemingly non-specific, cross reactivity against malaria parasite extract, however (Read M., personal communication). Similar levels of cross-reactivity were subsequently observed by the same researcher in sera taken from non-immunised experimental rabbits provided by Eurogentec (Herstal, Belgium), a company offering custom antibody production, indicating that cross-reactivity of rabbit antiserum with malaria parasite extract was perhaps a generalised problem, possibly due to exposure of the experimental animals to related coccidian parasites. Sera taken from non-immunised mice provided by the same company showed little undesirable cross-reactivity however. The initial immunisation strategy of choice was therefore the inoculation of mice with protein extracted from a western blot (2.12.1).

Four 50ml log-phase cultures of BL21 cells transformed with pGEX-Cgch were induced to produce GST-Cgch through the addition of 1mM IPTG. The cells were harvested 4 hours post-induction, resuspended in a total of 4ml 1 x PBS (2.1.2) and lysed by sonication (2.11.1). GST-Cgch was then purified from the sonicate by glutathione affinity chromatography (2.11.2), with the purified fusion protein eluted in a total of 3ml glutathione elution buffer (2.1.26). This was then dialysed against sdH_2O to remove salts (2.11.4) and a small aliquot subjected to SDS-PAGE (2.6.3) in order to visually gauge the approximate concentration of GST-Cgch (results not shown). An aliquot equivalent to 400-500µg dialysed protein was then lyophilised, resuspended in 225µl of '1 x' SDS-PAGE reducing/loading buffer (2.1.11), loaded in 9 separate 25µl aliquots onto an SDS-PAGE gel (2.6.3) and subjected to electrophoresis and western blot (2.12.1, 2.13.1). The membrane was stained for total protein with Ponceau S (2.12.3, results not shown) and stained bands containing the approx. 47 kDa fusion protein were excised from the membrane. 500µl of DMSO was added to the membrane fragments and the whole was vortexed to dissolve the nitro-cellulose matrix.

The DMSO/antigen/nitro-cellulose mixture was subsequently delivered to the laboratories of Eurogentec. Here it was inoculated into 3 experimental mice, SK8, SK9 and SK10, according to the regime outlined in 2.13.1. 350µl of serum was bled from each of the surviving mice (SK8 having died during the course of the experiment) 10 days after the third injection. These serum samples, along with samples taken from each

surviving mouse prior to the first injection, were then sent to UMIST. Each of the sera was depleted for antibodies against nitro-cellulose and *E.coli* (2.13.2).

Aliquots of purified GST-Cgch and GST alone were then western blotted onto quadruplicate nitro-cellulose membranes (2.12.1). Following initial washes, one membrane each was incubated with a $1/1000^{th}$ dilution of depleted serum from the following four sources; SK9 pre-immunisation, SK9 post-immunisation, SK10 pre-immunisation and SK10 post-immunisation. Each membrane was then incubated with AP conjugated goat anti-mouse IgG antibody, and developed with the colorimetric agents NBT and BCIP (2.12.2). No bands appeared in any of the four cases (results not shown). Further attempts were made to demonstrate antiserum/antigen reactions using enhanced quantities of serum (up to $1/100^{th}$ dilution) and serum which had been depleted against nitro-cellulose only (100µl of serum was made up to 1ml with sdH_2O, then shaken in the presence of nitro-cellulose as above). None of these modifications proved successful however, and it was therefore resolved to attempt a different immunisation strategy.

Large quantities of IgY can be purified from the yolks of chicken eggs, without any need for specialised animal training, making this species an attractive alternative to mammals for immunisation. IgY purified from eggs laid by several non-immunised birds provided by Eurogentec was tested for cross-reactivity with malaria parasite extract (Read, M., personal communication). When no deleterious cross-reactivity was observed it was decided that immunisation of experimental chickens with GST-Cgch should be attempted. It was decided that, as the GST-Cgch fusion represented the majority of the protein in the relevant purified fraction (see e.g. Figure 3.11), it would be acceptable to use this fraction for immunisation without further separating GST-Cgch from the other species present.

GST-Cgch was purified from a total volume of 200ml induced culture as above then dialysed (2.11.4) against 1x PBS (2.1.2). A small aliquot was subjected to SDS-PAGE (2.6.3) in order to visually gauge the approximate concentration of GST-Cgch (results not shown). An aliquot equivalent to 400-500µg dialysed protein was then lyophilised and delivered to the laboratories of Eurogentec. Here it was inoculated into 2 experimental chickens, SK1 and SK2, according to the regime outlined in 2.13.3. Eggs were collected from SK2 (SK1 having died during the course of the experiment) every day from 5 to 18 days after the third injection. The yolks from these eggs, along

with the yolk from an egg laid by SK2 prior to the first injection, were then sent to UMIST. IgY was partially purified from approximately 5ml each of pre-immunisation and post-immunisation yolk (2.13.4, Figure 3.13).

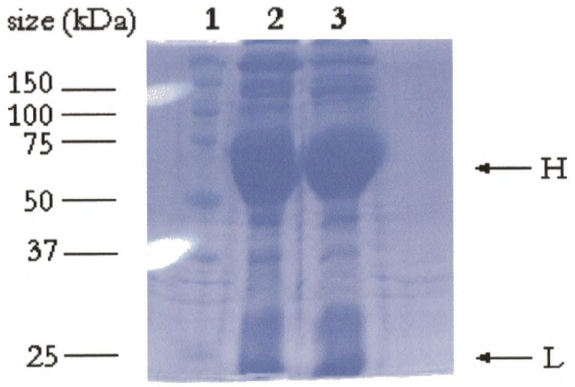


Figure 3.13 Partial purification of IgY from chicken egg yolks
Lane numbers: 1, Precision molecular weight marker; 2, IgY preparation from pre-immunisation yolk; 3, IgY preparation from post-immunisation yolk. IgY heavy (H, approximately 65 kDa) and light (L, approximately 25 kDa) chains are labelled.

Aliquots of purified GST-Cgch, GST-dhfs (another GST fusion protein, provided by Martin Read, UMIST) and MBP-Cgch (3.10) were then western blotted onto a nitro-cellulose membrane (2.12.1). Following initial washes, the membrane was incubated with a 1/200th dilution of post-immunisation IgY preparation. The membrane was then incubated with AP conjugated rabbit anti-chicken IgY antibody, and developed with the colorimetric agents NBT and BCIP (2.12.2, Figure 3.14).

The post-immunisation IgY reacted not only with the target antigen GST-Cgch, but also with the GST-dhfs and full length MBP-Cgch fusion proteins, which share with the target antigen the GST and Cgch domains respectively, indicating that the IgY contained antibodies specific to epitopes in both domains. Promisingly, the IgY did not react with the complex of bands of around 40-45 kDa present alongside the full length MBP-Cgch fusion. It is considered likely that these bands consist of predominantly MBP derived truncations or breakdown products of the full-length fusion (see 3.11); the lack of interaction between the IgY and these species suggests that the antibody/antigen reactions demonstrated are specific.

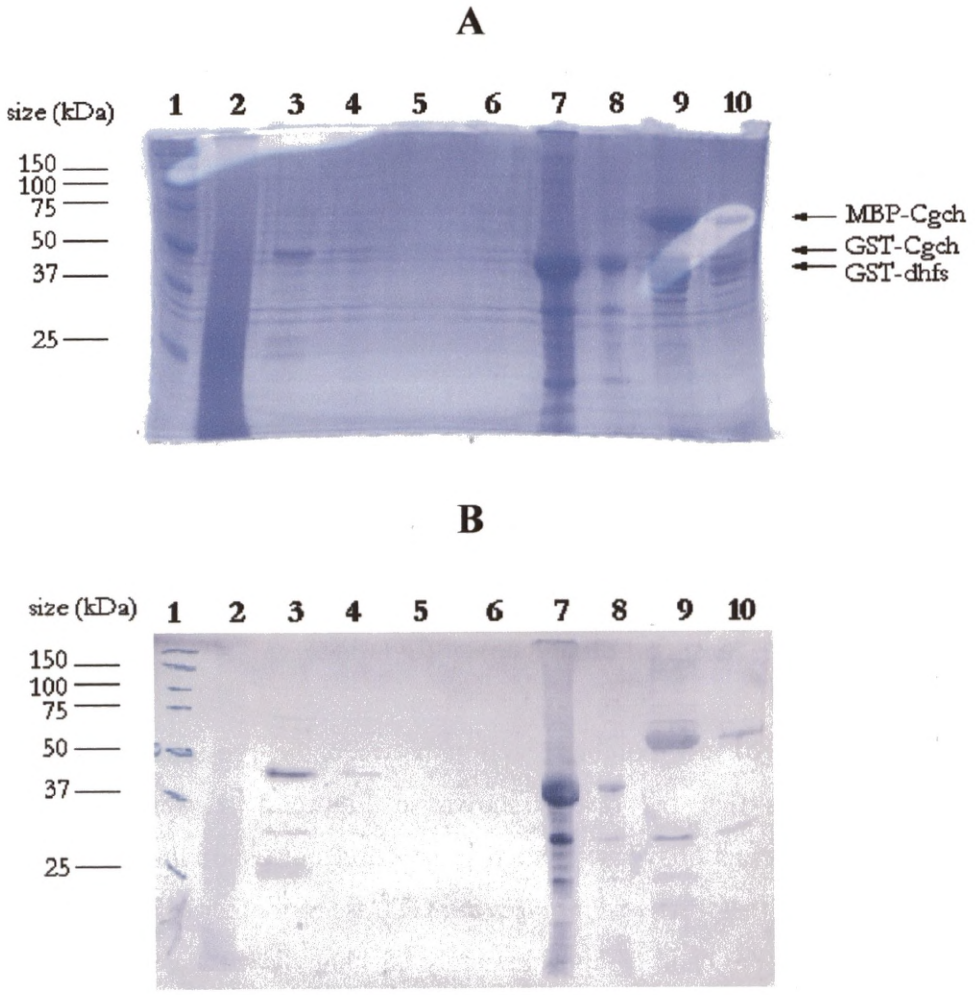


Figure 3.14 Test blot of anti-GST-Cgch chicken IgY vs. fusion proteins
A) Equivalent of original polyacrylamide gel; Lane numbers: 1, Precision molecular weight marker; 2, BL21 pGEX-Cgch cells, crude lysate (sample is degraded); 3, GST-Cgch; 4, GST-Cgch 1/10th dilution; 5, GST-Cgch 1/100th dilution; 6, GST-Cgch 1/1000th dilution; 7, GST-dhfs; 8, GST-dhfs 1/10th dilution; 9, MBP-Cgch; 10, MBP-Cgch 1/10th dilution.
B) Developed membrane; Lane numbers: as A). n.b. positions of markers were added immediately post-transfer.

In order to further test the IgY, aliquots of purified GST-Cgch and malaria parasite extract were western blotted onto duplicate nitro-cellulose membranes. Following initial washes, one membrane each was incubated with a 1/200th dilution of the pre-immunisation and post-immunisation IgY preparations respectively. The membranes were incubated with AP conjugated rabbit anti-chicken IgY antibody, and developed with the colorimetric agents NBT and BCIP (Figure 3.15).

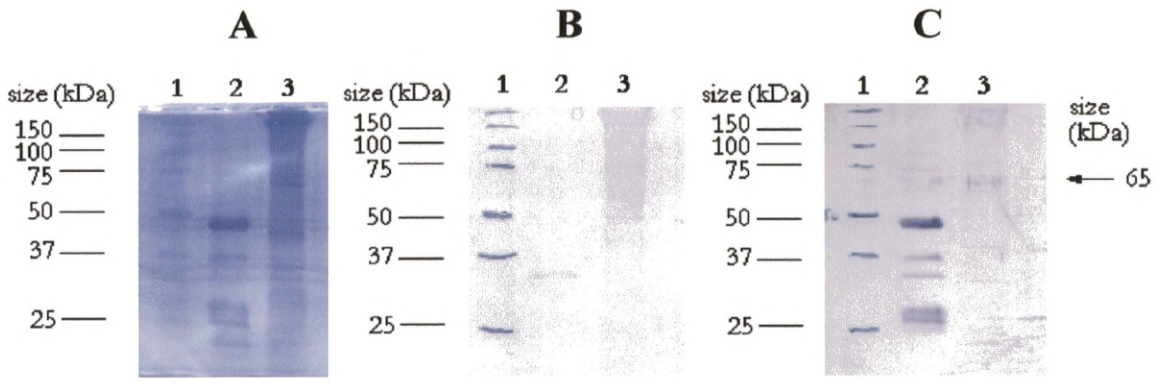


Figure 3.15 Test blot of anti-GST-Cgch chicken IgY vs. parasite extract
A) Equivalent of original polyacrylamide gel; Lane numbers: 1, Precision molecular weight marker; 2, GST-Cgch; 3, *P.falciparum* cell extract, K1 lineage, from approx. $5x10^8$ pRBCs.
B) Developed membrane, probed with IgY purified from pre-immunisation yolk;
C) Developed membrane, probed with IgY purified from post-immunisation yolk;
For B)/C), lane numbers as A). n.b. positions of markers were added immediately post-transfer.

The absence of any notable reaction between the pre-immunisation IgY and GST-Cgch confirms that the results shown in Figure 3.14 were due to specific antibody/antigen interactions. Unexpectedly, no band corresponding to the calculated molecular weight of the putative *P.falciparum* GTCase monomer (45.9 kDa) lit up from the parasite extract (the band of approximately 65 kDa shown was not apparent when this experiment was repeated (result not shown)). In a previous similar experiment a 'masking effect' was observed, whereby purified parasite DHFR, when mixed with parasite extract, was unable to react with anti-DHFR specific antiserum which was otherwise able to recognise it (Nirmalan N., personal communication). In order to test whether a similar effect was responsible for the non-appearance of a *Pf* GTCase-specific band when parasite extract was probed with the post-immunisation IgY, we probed parasite extract spiked with the antigen, GST-Cgch, as well as extract spiked with MBP-Cgch, another *Pf* GTCase derived fusion protein initially described in 3.8 (Figure 3.16).

In both cases the IgY preparation was able to recognise the purified protein when mixed with the parasite extract, although some reduction in intensity was noted in the reaction with GST-Cgch, suggesting that the 'masking effect' was not a major factor in this case. Surprisingly, a band of around 65 kDa again lit up from the unspiked parasite extract, casting doubt on the earlier negative result, however, when a fresh aliquot of parasite extract was subsequently probed alongside a GST-Cgch positive control, no 65 kDa band was seen (result not shown).

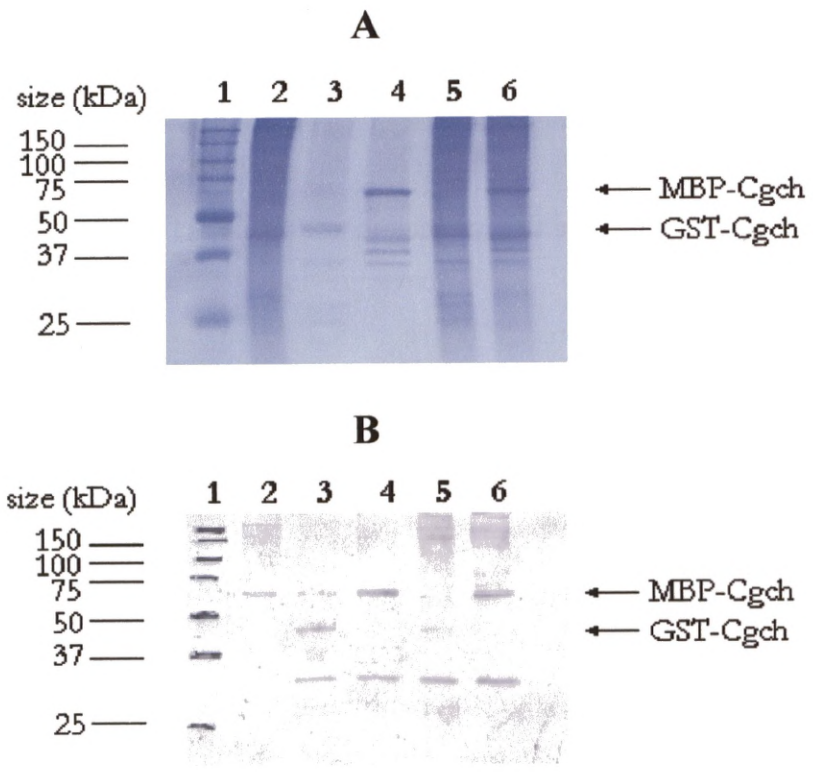


Figure 3.16 Blot of spiked parasite extracts with anti-GST-Cgch chicken IgY
A) Equivalent of original polyacrylamide gel; Lane numbers: 1, Precision molecular weight marker; 2, *P.falciparum* cell extract, K1 lineage, from approx. $5x10^8$ pRBCs; 3, GST-Cgch; 4, MBP-Cgch; 5, *P.falciparum* cell extract plus GST-Cgch; 6, *P.falciparum* cell extract plus MBP-Cgch.
B) Developed membrane; Lane numbers: as A). n.b. positions of markers were added immediately post-transfer.

3.8: Production of clones for expression of MBP fusion proteins

The pMAL series of vectors (Appendix VI) express gene product as a fusion with a cleavable N-terminal maltose-binding protein (MBP) tag. The vectors have been demonstrated to dramatically enhance solubility, and also to increase biological activity, in a wide variety of otherwise recalcitrant heterologous gene products in *E.coli* (Kapust & Waugh, 1999). It was decided to use pMAL-c2G, one of the vectors in the series, in parallel with pGEX-6P-2 to attempt to express quantities of soluble, biologically active protein derived from *Pf gch-1*. *Pf* GTCase, the product of *Pf gch-1*, is considerably larger then other GTCases (Appendix III). The polypeptide shows a high degree of homology to other GTCases in its C-terminal region, which contains residues homologous to all those implicated in subunit interactions and substrate binding in the *E.coli* GTCase. However, *Pf* GTCase has an unusually long N-terminal portion, which

possibly plays a part in physiological regulation of the enzyme activity. With reference to this fact, and in view of the apparent difficulties in expressing N-terminal regions of *Pf gch-1* (3.3, 3.4), it was decided to attempt to clone *Pf gch-1* fragments corresponding to the C-terminal half of the gene product into pMAL-c2G, rather than the full length *Pf gch-1* gene, in the hope that this truncation would not affect multimer assembly and basic activity. Auerbach *et al.* (2000) have previously demonstrated close to wild-type catalytic activity from a human GTCase with a 40 amino acid N-terminal truncation, whereby the polypeptide starts at an amino acid approximately equivalent to Lys177 in the *Pf* GTCase (see Appendices II & III).

Three oligonucleotide primers, gtpmal1, gtpmal2 and gtpmal3 (Appendix V) were designed to produce *Pf gch-1* derived DNA fragments with terminal *Bam* HI and *Hin*d III sites suitable for unidirectional cloning into pMAL-c2G.

The 2 inserts produced (Figure 3.17) were as follows:-

an approximately 600bp product amplified by gtpmal1 and gtpmal3 (Cgch), corresponding to a C-terminal fragment of the gene product starting a short distance upstream of both Leu225, the most extreme N-terminal residue which is completely conserved between different species, and those amino acids roughly aligned with the start codons of the shortest bacterial GTCases, for example the enzymes from *Campylobacter jejuni* and *Streptococcus pneumoniae* (Appendix III), and an approximately 400bp product amplified by gtpmal2 and gtpmal3 (COREgch), corresponding to a C-terminal fragment of the gene product starting just upstream of Ile266, which has been structurally aligned (Lee, 1999) to Val99 in the *E.coli* GTCase, the first residue in s_1, the first secondary structure element in the compact C-terminal domain of that species (as described by Nar *et al.*, 1995a).

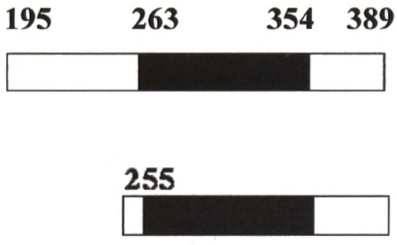


Figure 3.17 ***Pf gch-1*** **derived inserts for use with pMAL-c2G**
Upper, Cgch; lower, COREgch. Black box indicates the highly conserved 'core' region of the gene. Numbers indicate corresponding amino acids in polypeptide sequence.

PCRs (2.7.1) were set up using the clone of *Pf gch-1* cDNA in pBS as template, and the two primer combinations. The amplification products were visualised on an agarose gel (2.6.1) and purified from the reaction mixes (2.3.1), then digested with the cloning enzymes, as was the vector (2.4.1). Both constructs, pMAL-Cgch and pMAL-COREgch, were subsequently produced by cloning (2.5.1, results not shown).

3.9: Expression of the pMAL-c2G derived constructs

Each of the constructs was transformed into cells of *E.coli* strain MSD 2658 (Appendix I, hereby described as MSD, this strain was used instead of BL21 in order to comply with revised departmental safety directives) using $CaCl_2$ methodology (2.5.2, 2.5.3). 50ml log-phase cultures of each transformant were induced at 30°C to produce fusion protein through addition of 1mM IPTG. The cells were harvested by centrifugation approximately 4 hours after induction (see 2.11.3). Small quantities of cells (harvested from 1.5ml of induced culture) were resuspended in 250µl '1 x' SDS-PAGE reducing/loading buffer (2.1.11), boiled to lyse and subjected to SDS-PAGE (2.6.3). The gel was subsequently stained for total protein with Coomassie Blue (2.1.16, Figure 3.18).

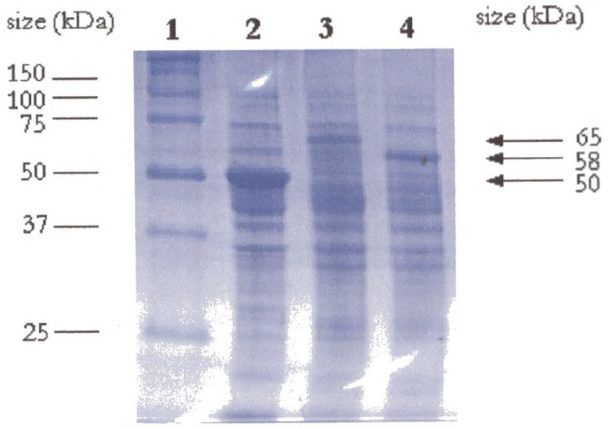


Figure 3.18 Crude lysates of induced MSD pMAL-c2G clones
Lane numbers: 1, Precision Protein Standards; 2, MSD pMAL-c2G vector only cells; 3, MSD pMAL-Cgch cells; 4, MSD pMAL-COREgch cells.

The lysate of the cells transformed with pMAL-c2G showed a band of around 50 kDa, corresponding closely to the expected size for MBP of 51 kDa, while the lysates of the cells transformed with pMAL-Cgch and pMAL-COREgch showed distinct unique

bands of around 65 kDa and 58 kDa respectively. These sizes are relatively close to those expected of the respective full length fusion proteins, MBP-Cgch (66 kDa) and MBP-COREgch (59 kDa.)

3.10 : Purification of MBP & MBP-gch fusion proteins

The remainder of the MSD transformant cells that had been harvested following induction (3.9) were resuspended in 1ml PBS (2.1.2) and subsequently lysed by sonication/freeze thaw cycles (2.11.1). These lysates was separated into soluble and insoluble fractions by centrifugation. Small quantities of each fraction were subjected to SDS-PAGE (2.6.3)(Figure 3.19). Significant quantities of both MBP and the putative MBP-Cgch and MBP-COREgch fusion proteins were found to be soluble.

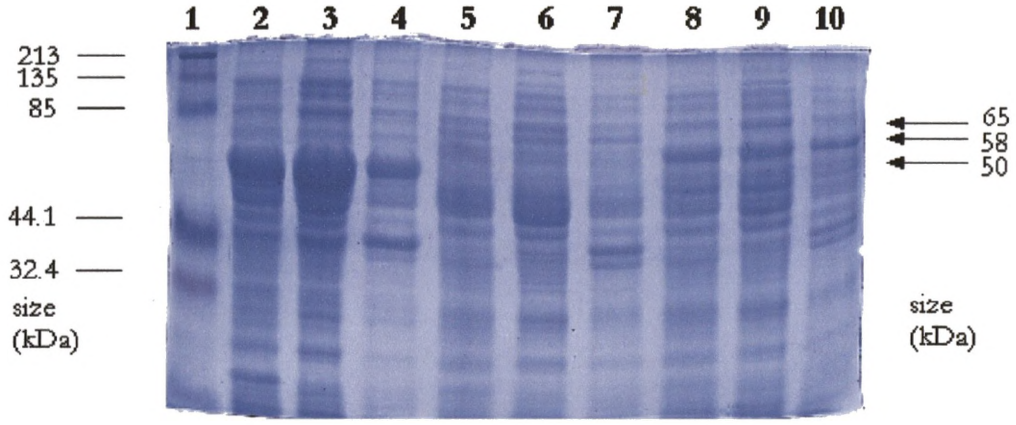


Figure 3.19 Solubility of proteins in cells transformed with pMAL-c2G derived clones grown at 30°C Lane numbers: 1, Kaleidoscope molecular weight markers; 2, MSD pMAL-c2G total protein; 3, MSD pMAL-c2G soluble fraction; 4, MSD pMAL-c2G insoluble fraction; 5, MSD pMAL-Cgch total protein; 6, MSD pMAL-Cgch soluble fraction; 7, MSD pMAL-Cgch insoluble fraction; 8, MSD pMAL-COREgch total protein; 9, MSD pMAL-COREgch soluble fraction; 10, MSD pMAL-COREgch insoluble fraction.

500µl of each of the 3 soluble fractions was subjected to maltose affinity column purification (2.11.3)(Figure 3.20). This resulted in the purification of a significant proportion of MBP, and the putative MBP-Cgch and MBP-COREgch fusions, with a low level of undesirable copurification of other species.

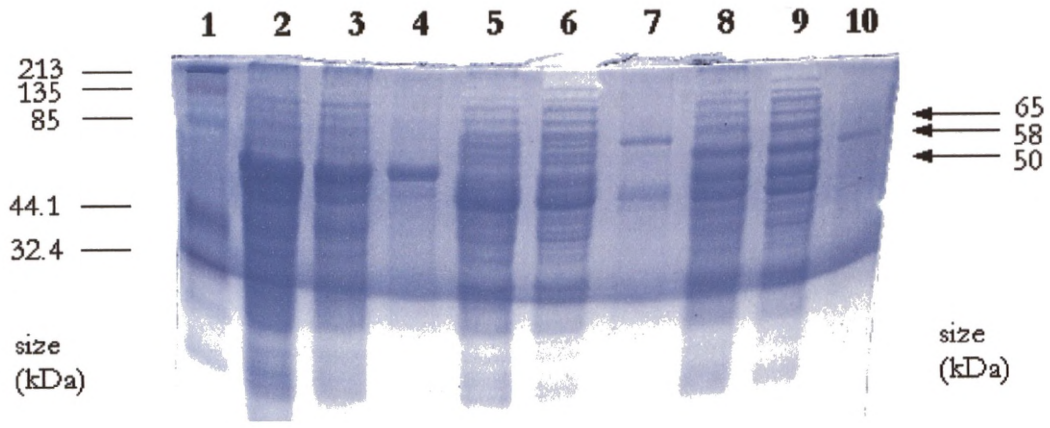


Figure 3.20 Affinity purification of MBP and MBP fusion proteins
Lane numbers: 1, Kaleidoscope molecular weight markers; 2, MSD pMAL-c2G soluble fraction; 3, MSD pMAL-c2G pre-wash eluate; 4, MSD pMAL-c2G post-column eluate; 5, MSD pMAL-Cgch soluble fraction; 6, MSD pMAL-Cgch pre-wash eluate; 7, MSD pMAL-Cgch post-column eluate; 8, MSD pMAL-COREgch soluble fraction; 9, MSD pMAL-COREgch pre-wash eluate;10, MSD pMAL-COREgch post-column eluate.

3.11 : Western blot analysis of the putative MBP-gch fusion proteins

In order to confirm the origin of the polypeptides purified in 3.10, it was resolved to subject them to Western blot analysis (2.12) using an antibody specific to the MBP domain produced by the pMAL vector. Proteins were transferred to nitro-cellulose membrane (2.12.1) from a polyacrylamide gel equivalent to Figure 3.21A.

Following initial washes, the membrane was incubated first with rabbit anti-MBP antibody, then with AP conjugated goat anti-rabbit IgG antibody, and developed with the colorimetric agents NBT and BCIP (2.12.2). The pattern of bands observed after development is shown in Figure 3.21B.

All three target proteins were seen to 'light up' from both their respective purified and non-purified fractions. The relatively high background on the blot was thought to be the result of truncation/proteolysis of the target proteins, as there was very little background produced by the soluble fraction of the untransformed cells.

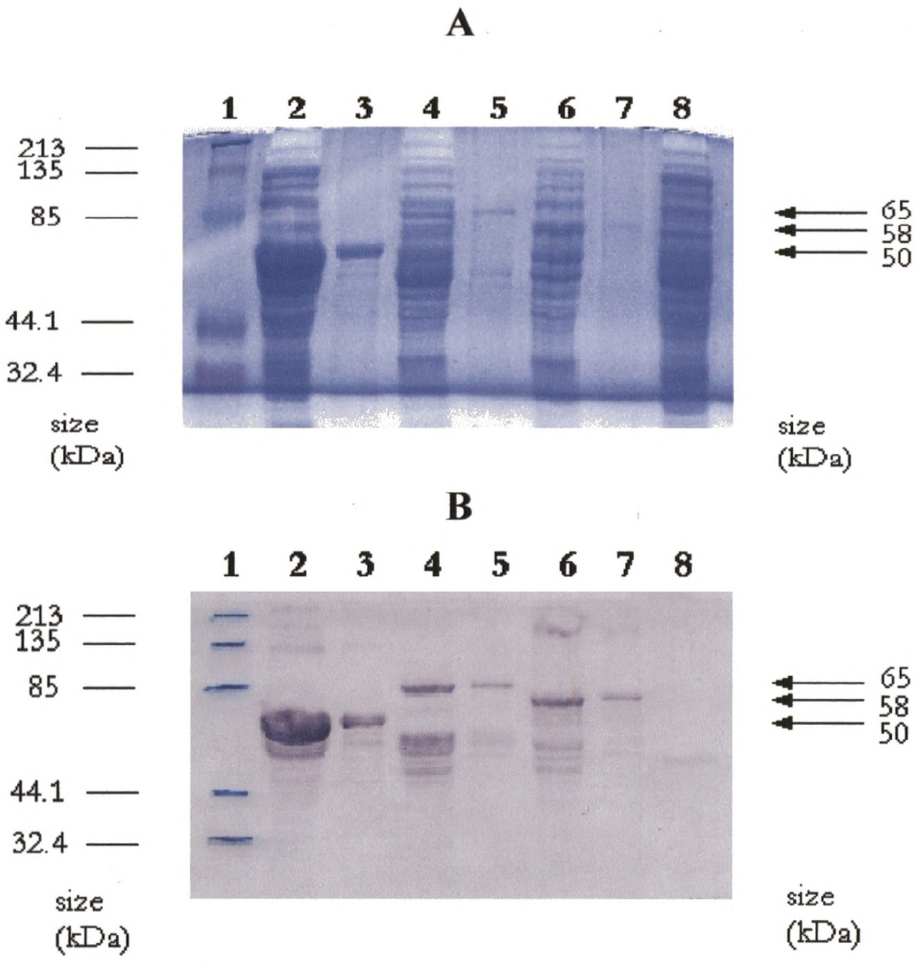


Figure 3.21 Western blot of putative MBP-gch fusion proteins
A) Equivalent of original polyacrylamide gel; Lane numbers: 1, Kaleidoscope molecular weight marker; 2, MSD pMAL-c2G soluble fraction; 3, MSD pMAL-c2G post-column eluate; 4, MSD pMAL-Cgch soluble fraction; 5, MSD pMAL-Cgch post-column eluate; 6, MSD pMAL-COREgch soluble fraction; 7, MSD pMAL-COREgch post-column eluate; 8, MSD untransformed cells, soluble fraction.
B) Developed membrane; Lane numbers: as A). n.b. positions of markers were added immediately post-transfer.

3.12 : Fusion protein cleavage

It was decided that, in order to produce material suitable for future GTCase activity assays, it was advisable to cleave the GTCase-specific domains of the GST and MBP fusion proteins described earlier from their fusion polypeptides. It was felt that the multimeric assembly almost certainly required for enzyme activity (see 1.9) would be most likely to occur without the impediment of the bulky fusion domains.

Purified samples of GST-Cgch and each of the MBP-gch fusion proteins were dialysed (2.11.4) against 1 x PBS (2.1.2) and genenase reaction buffer (2.1.30) respectively. 750µl of dialysed GST-Cgch and 1ml of each dialysed MBP-gch fusion protein were then subjected to proteolysis with PreScission protease (2.11.5), in the case of GST-Cgch, and genenase I (2.11.6), in the case of the MBP fusions. In order to monitor the proteolysis and subsequent separations, samples were taken from each reaction and quenched in SDS-PAGE reducing/loading buffer (2.1.11 or 2.1.12/2.1.13 as appropriate). These samples were then subjected in duplicate to SDS-PAGE (2.6.3). One gel of each duplicate pair was then stained for total protein with Brilliant Blue G – colloidal solution (2.6.3), while the other was subjected to western transfer (2.12.1) and the resulting blot immunoprobed (2.12.2) with anti-GST-Cgch chicken IgY (3.7). The results of these experiments are shown in Figures 3.22 to 3.24.

Each of the three proteolyses was successful to a different extent. In the case of GST-Cgch (Figure 3.22), a large percentage of the fusion protein appeared to be cleaved within 30 min of the addition of the protease (lane 4), as indicated by the significant reduction in the level of the 49 kDa fusion protein and appearance of a polypeptide of approximately 26 kDa, most likely the GST domain. Little additional visible reduction in GST-Cgch occurred throughout the remainder of the reaction time (lanes 5-7). Surprisingly, given that the Cgch domain should be generated in equimolar quantities to GST, only a faint band appeared at around the expected size (23 kDa) on both the gel and the blot; this possibly indicates that any Cgch liberated by fusion protein cleavage rapidly degenerates. As expected, the evidence of the stained gel (lane 9) suggests that GST is released by the column matrix following incubation with glutathione elution buffer (2.1.26) – the fate of the putative Cgch domain is less clear however; it did not appear to be present in the post-proteolysis eluate, and possibly bound irreversibly to the column matrix.

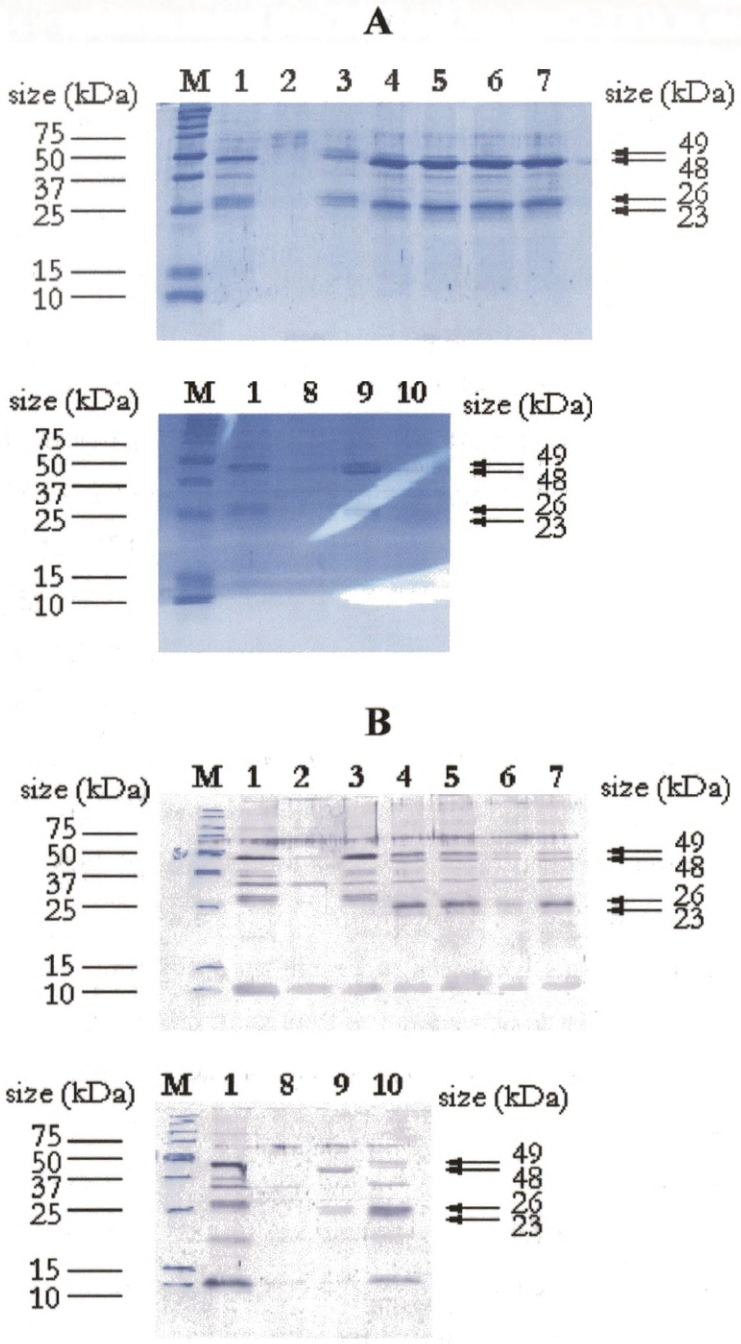


Figure 3.22 Proteolysis of GST-Cgch

Refer to 2.11.5 for experimental outline. **A)** Stained polyacrylamide gels; Lane numbers: M, Precision molecular weight marker; 1, GST-Cgch, post-dialysis; 2, eluate from column after addition of GST-Cgch; 3, column, pre-addition of protease; 4, column, 30 min post-addition of protease; 5, column, 1hr post-addition; 6, column, 2hr post-addition; 7, column, 4hr post-addition; 8, eluate from column, post-proteolysis; 9, eluate from column after addition of glutathione elution buffer; 10, column, after elution with glutathione elution buffer. **B)** Developed membranes; Lane numbers: as A).

The approximate sizes of the major protein species are marked as follows; GST-Cgch, 49 kDa; PreScission protease (this species is a GST fusion), 48 kDa; GST, 26 kDa; Cgch, 23 kDa.

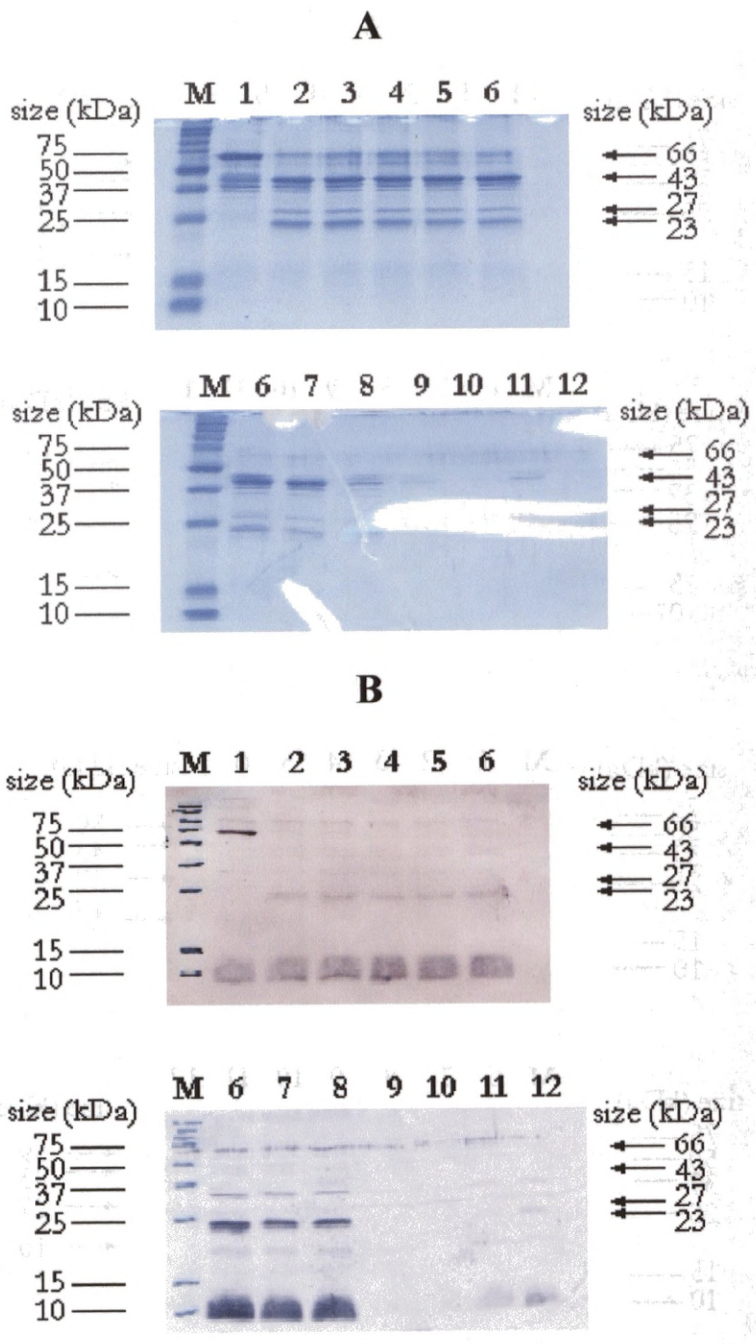


Figure 3.23 Proteolysis of MBP-Cgch

Refer to 2.11.6 for experimental outline. **A)** Stained polyacrylamide gels; Lane numbers: M, Precision molecular weight marker; 1, MBP-Cgch, post-dialysis; 2, reaction mix, 30 min post-addition of protease; 3, reaction mix, 1 hr post-addition; 4, reaction mix, 2 hr post-addition; 5, reaction mix, 4 hr post-addition; 6 reaction mix, 8 hr post-addition; 7, eluate from column following addition of reaction mix; 8-10, successive washes of column; 11, eluate from column after addition of maltose elution buffer; 12, column, after elution with maltose elution buffer. **B)** Developed membranes; Lane numbers: as A). The approximate sizes of the major protein species are marked as follows; MBP-Cgch, 66 kDa; MBP fusion domain, 43 kDa; genenase I, 27 kDa; Cgch, 23 kDa.

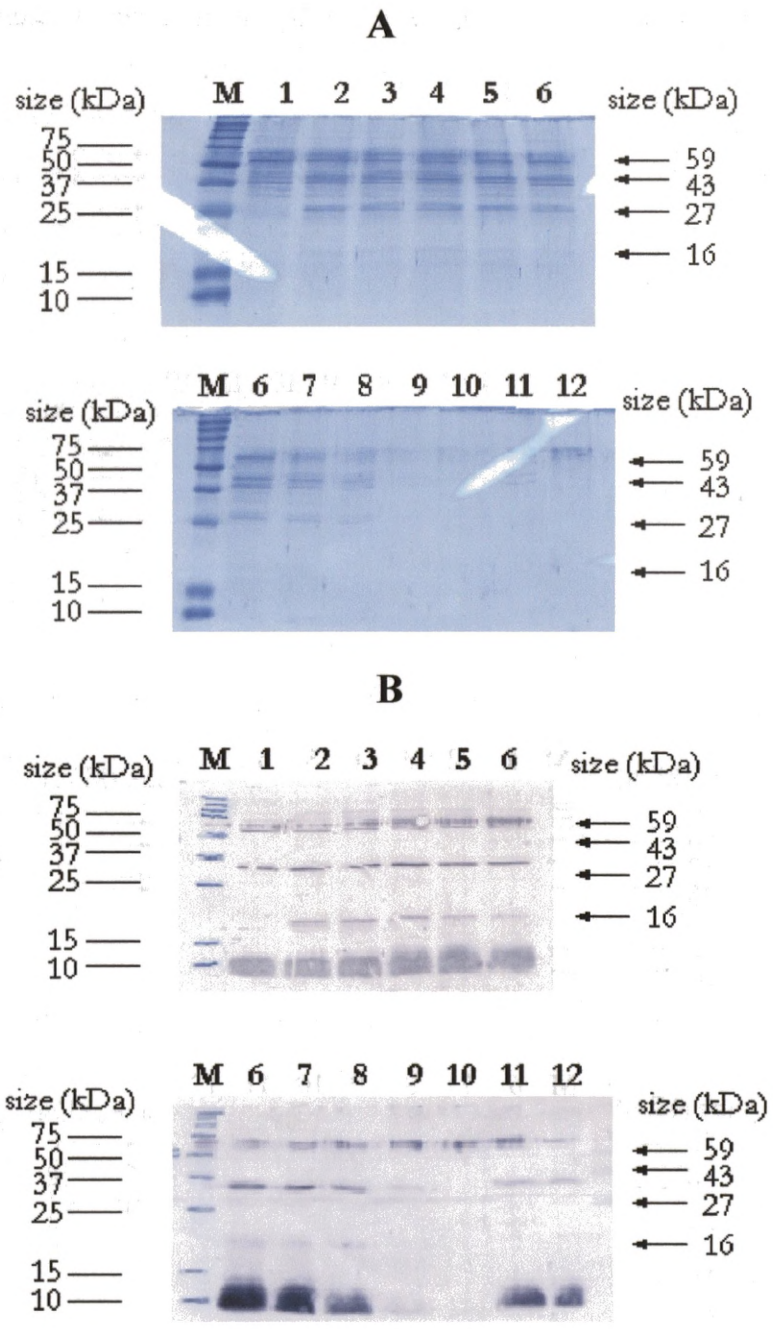


Figure 3.24 Proteolysis of MBP-COREgch

Refer to 2.11.6 for experimental outline. **A)** Stained polyacrylamide gels; Lane numbers: M, Precision molecular weight marker; 1, MBP-COREgch, post-dialysis; 2, reaction mix, 30 min post-addition of protease; 3, reaction mix, 1 hr post-addition; 4, reaction mix, 2 hr post-addition; 5, reaction mix, 4 hr post-addition; 6 reaction mix, 8 hr post-addition; 7, eluate from column following addition of reaction mix; 8-10, successive washes of column; 11, eluate from column after addition of maltose elution buffer; 12, column, after elution with maltose elution buffer. **B)** Developed membranes; Lane numbers: as A). The approximate sizes of the major protein species are marked as follows; MBP-COREgch, 59 kDa; MBP fusion domain, 43 kDa; genenase I, 27 kDa; COREgch, 16 kDa.

In the case of MBP-Cgch (Figure 3.23), again the reaction seems to proceed towards completion fairly rapidly, with much of the fusion protein apparently cleaved within 30 min of addition of the protease (lane 2), and little subsequent cleavage (lanes 3-6). In this case, however, the accumulation of Cgch occurred much less ambiguously, with a strong band of around 23 kDa appearing in concert with the 43 kDa MBP fusion domain (this 23 kDa band also appears on the IgY probed western blots); some slight reduction in intensity of this band over the course of the reaction was noted, however. Surprisingly, separation of this species from MBP by amylose affinity chromatography (see 2.11.3) was not possible, as the MBP domain appeared unable to rebind to the column matrix (see lanes 7&11).

The MBP-COREgch fusion protein appeared more resistant to proteolysis than the other two fusions studied (Figure 3.24) – the majority of this species appeared to remain intact 8hr after the addition of the protease (lane 6). Some cleavage did occur, however, as indicated by the appearance within 30 min of a polypeptide of around 43 kDa, almost certainly the MBP-fusion domain (lane 2). Surprisingly, given the relative abundance of uncleaved fusion protein, little further proteolysis was apparent after 30 min. In addition to MBP, a very small amount of a polypeptide of around 16 kDa, corresponding to the size of the COREgch domain, appeared following the addition of protease. This polypeptide was also able to react with the IgY, strongly indicating that it was indeed COREgch. Some reduction in this species over the course of the reaction was apparent. As in the previous experiment, separation of this species from MBP was not achieved, as the MBP was unable to rebind to the matrix of the amylose affinity column (see lanes 7&11).

3.13 : Discussion

3.13.1 : Expression of fragments of *Pf* GTCase incorporating the N-terminus

Expression studies of the *Pf gch-1* gene in *E.coli* have produced mixed results; while expression of fragments corresponding to the C-terminal portion of the gene product has produced significant quantities of proteins of approximately the size expected (3.4, 3.10), there appear to be problems associated with expression of fragments containing sequence corresponding to the N-terminal portion of the gene product (3.4). There are a number of potential causes of this apparent effect, for example, the upstream portion of *Pf gch-1* contains a number of long polyA tracts

(Appendix II), which may act to disrupt efficient transcription of the gene. In addition, although no full Shine-Dalgarno consensus sequences (Shine & Dalgarno, 1974) are present in *Pf gch-1*, it is possible that the presence of sub-optimal sequences causes translational interference sufficient to result in the predominance of truncated gene product. It may, nevertheless, be possible to achieve efficient expression of proteins incorporating the N-terminal portion of *Pf* GTCase in *E.coli*, for example by varying the expression vector or host strain.

3.13.2 : Failure of anti-GST antiserum to recognise GST-Cgch

Although the 47 kDa polypeptide purified from the lysates of cells transformed with pGEX-Cgch (Figures 3.7, 3.8) has subsequently been proved beyond reasonable doubt to be the anticipated GST-Cgch fusion protein (3.6), commercial anti-GST antiserum was consistently unable to recognise this species when denatured (efficient transfer of GST-Cgch to the membrane from an SDS (denaturing) gel has been established to occur (Figure 3.11)), although it was apparently able to recognise this protein when transferred from a non-denaturing gel (3.5). It is possible that this differential effect is due to the greater sensitivity of antibodies against conformational isotopes of GST within the serum; this serum is explicitly recommended for detection of GST fusion protein which has previously been denatured by boiling with SDS, however. The reasons for this protein not lighting up when subjected to western analysis following SDS PAGE remain obscure therefore; it is possible but seems unlikely that a particular aggregation of GST-Cgch occurs during western transfer following denaturing PAGE, which prevents normal antibody-antigen interactions. Despite the failure of immunoprobing with anti-GST antiserum to confirm the identity of GST-Cgch, confirmation has been achieved, to a level of certainty impossible through western analysis, through the utilisation of peptide mass fingerprinting (3.6). With the availability of the appropriate technology, this technique represents an attractive and relatively straightforward alternative route to fusion protein identification.

3.13.3 : Generation of anti-*Pf* GTCase antibodies

The initial experimental strategy chosen for the generation of antibodies specific to the GST-Cgch fusion protein was the inoculation of mice with antigen that had been blotted onto nitrocellulose then dissolved with DMSO (3.7). This strategy proved unsuccessful and possibly even precipitated the death of one of the three experimental

animals. The mice used in this experiment were highly inbred BALB/c strain animals; it is possible that the combination of the natural toxicity of DMSO with the comparatively low body weight of the inbred mice resulted in the poor immune response/ fatal effects observed. The alternative strategy, inoculation of chickens with purified antigen dissolved in saline proved far more successful, and carried the additional benefit of yielding vast quantities of 'immune' material in the form of several weeks' worth of egg yolks, rather than the meagre 2 or 3 ml of blood recoverable from a mouse. However, despite the satisfactory level of sensitivity and specificity demonstrated by the immune yolk extract when used to probe GST-Cgch and other purified fusion proteins, the post-immunisation IgY did not react with a protein of the anticipated 46 kDa (corresponding to *Pf* GTCase) when it was used to probe malaria parasite extract. On certain occasions a band of around 65 kDa was lit up from parasite extract (Figures 3.15, 3.16); it has long been established that the progress of highly basic proteins, for example the histones, through SDS-PAGE gels can exhibit retardation (Panyim & Chalkley, 1971), and given that the theoretical pI of *Pf* GTCase (as determined using the Compute pI tool at http://www.expasy.ch) is around pH 9.4, it is possible that such retardation is responsible for this anomaly. Identification of the 65 kDa species did not occur consistently, however, and, given the apparent preponderance of an artefact band of around this size on western blots (see for example Figure 3.22B), it seems likely that the appearance of the 65 kDa band in parasite extract was, in one way or another, simply a coincidence. The failure of the post-immunisation IgY to light up a band of the expected size from parasite extract casts doubts on the usefulness of this material for probing *Pf* GTCase at normal cellular concentrations against a background of parasite total protein, for example in *in situ* immunofluorescence tests. However, the availability of antibodies against *Pf* GTCase has already proved useful in other analyses (3.12).

3.13.4 : Fusion protein cleavage

The proteolytic treatments of the three *Pf* GTCase derived fusion proteins were variably successful (3.12). In each case the majority of cleavage noted occurred in the first 30 min of the reaction; this was most likely due to an excess of protease in the reaction mix. In the cases of GST-Cgch and MBP-Cgch protease digestion proceeded almost to completion, however, in the case of MBP-COREgch, much of the fusion protein remained intact, even 8 hr after the addition of protease. It is possible that the inefficiency in this case is due to the three-dimensional conformation of this species or

the particular nature of the polypeptide sequence adjacent to the protease cleavage site. In each case, regardless of the extent of the protease digestion, some degeneration of the parasite specific domain was noted; this appeared most severe in the case of GST-Cgch, where only a faint band was noted around the expected 23 kDa, contrasting with the intense band at 26 kDa corresponding to GST, produced in supposedly equimolar quantities with Cgch (Figure 3.22A). There are no PreScission protease cleavage sites in the GST-Cgch fusion polypeptide other than that at the junction of the two domains; it is possible therefore that some non-specific proteolytic degradation is responsible for the apparent instability of the Cgch domain in this system.

Of the three fusion proteins investigated, it seems therefore that MBP-Cgch is by some distance the most promising in terms of generating material suitable for GTCase activity assays. As MBP-Cgch shares its parasite-specific domain with GST-Cgch, there seems little merit in investigating this more troublesome species further, while production of free COREgch, although of academic interest in terms of its potential activity in comparison to Cgch, may be considered as of less importance than the longer polypeptide, as it is less representative of the native *Pf* GTCase. Unfortunately, separation of the Cgch domain from MBP by maltose affinity chromatography post-proteolysis proved unsuccessful (this effect was also noted in the case of MBP-COREgch). This may have been caused by incompatibility of the cleavage buffer with column binding; alternatively, the maltose added to the column used to purify the fusion protein from the cell lysate may not have disassociated from the MBP domain during dialysis, and thus inhibited rebinding. In order to separate Cgch from the other species in the protease digestion mix an alternative strategy, for example size-exclusion chromatography, may prove useful.

CHAPTER 4

EXPRESSION OF THE GTP CYCLOHYDROLASE I GENE FROM *PLASMODIUM FALCIPARUM* IN EUKARYOTIC CELLS

4.1 : Expression in *Saccharomyces cerevisiae*

The yeast *S.cerevisiae* is a popular system for the heterologous expression of cloned eukaryote genes. As a eukaryote, it is to an extent better able than a prokaryotic host to perform any post-translational modifications necessary to produce active gene product. Compared with other eukaryotic systems however, *S.cerevisiae* cells are relatively easy to grow and maintain in culture. In addition, decades of research into the yeast have led to the development of an extensive range of biochemical and genetic techniques for use with this organism.

As a result of the yeast genome project (Goffeau *et al.*, 1996) the entire sequence of the *S.cerevisiae* genome, along with the locations of around 6000 open reading frames (ORFs), is now known. It is thus comparatively straightforward to delete or disrupt any desired ORF in the genome. Complementation of the mutant through expression of a cloned gene product can then be attempted.

In this way, a number of malarial genes, including those encoding the folate synthesis/utilisation enzymes (1.8) *Pf dhfs/fpgs* (Salcedo *et al.*, 2000) and *Pf dhfr1* (Wooden *et al.*, 1997) have been shown to complement the yeast carrying the corresponding mutant gene. In addition, a strain of *S.cerevisiae* where *FOL2*, the gene encoding GTCase, has been knocked out, was rescued through expression of both the human and *E.coli* genes (Mancini *et al.*, 1999).

The UMIST Parasitology research group has made similar attempts to complement a *FOL2* knockout mutant with malarial GTCase (Lee, 1999). pMET-gch, a clone of *Pf gch-1* in the yeast shuttle expression vector pMETplus (a single copy plasmid with a repressible promoter, Appendix VI) was transformed into the haploid FY1679 Δ*FOL2*::*kanMX4* HP1 (Appendix I), a strain of *S.cerevisiae* where *FOL2* had been replaced by a cassette conferring resistance to geneticin. When the transformants

showed no ability to grow in the absence of folate, the mRNA transcripts produced by the expression vector were analysed. Northern hybridisation indicated that no full length transcripts were produced, while 3' RNA ligase-mediated rapid amplification of cDNA ends (3' RLM-RACE) and sequencing analysis identified that truncation of the mRNA was occurring at a minimum of three distinct points: at around 450, 660 and 675 nucleotides (nt) downstream of the *Pf gch-1* start codon (Figure 4.1).

```
ATGTATAAATATACGTCAATAAACAAATCTGATAAAATATATGAAACACATAATATGGAA      60
GAAAAAAAAAAAAAAGGTAATAATAACAATTTCTCAGGATTATTAAATAATGAAATCGAT     120
GATAATAATAAAAAGGAAAAGTTAAAAAATAGTATATCTAAAATGTATAGTAACCATAAG     180
AATAGAGAAAATTTTAACGAATGTGAGAAGGAAGACCTTGTTGTGATAGATGAAAAGGAT     240
AATAATAAAAAGAAAAAAAAAAATATGACAAATACATTTGAACAAGATAATAATTATAAT     300
ATGAATGATAATAAAAGGTTAGGTAGCTTTTTTAAAATAAATGATAAATGTGAATCTATT     360
AATGAGAATGTTAATAATATTAATAAACAGTCTCTTAAGGATTCTATACTATTTGACAAT     420
ATAAATGAGAAAGAATATTTTAACGAAACCAAAGAAGAAAATAAGGAAGGAAATAAAAGT     480
AATGATATAGAAAAGATAAATTGTATGAAAGTGAAGAAGAAAACTGTTAAAAAGAATAAA     540
AAGAAGATTAATAAAATTATAAACAATAAGAATAAAATATCTAAATCAAATGACATAGAA     600
GAACAAATAATTAATATTAGTAAACATATATATAAAATATTAAATATATCCAAATTACCA     660
AAATGTGATATATTAAAAAGAACAAATAGAAGATATGCTGAAACATTTTTATATTTAACT     720
AATGGTTATAATCTGGATATAGAACAAATAATAAAAAGATCTTTATATAAAAGGATGTAT     780
AAAAATAATTCAATAATCAAAGTTACAGGTATACATATATATTCATTATGTAAACACCAT     840
CTTTTACCTTTTGAAGGTACATGTGATATTGAGTATATACCCAATAAATATATTATCGGG     900
TTATCTAAATTTTCAAGAATAGTTGATGTCTTTTCTAGAAGATTACAATTACAAGAAGAT     960
TTAACTAACGATATTTGTAACGCTTTAAAAAAATACTTAAAACCATTATATATTAAAGTA    1020
TCTATTGTAGCTAAACATTTATGTATAAATATGAGGGGAGTTAAAGAGCACGATGCTAAA    1080
ACTATAACGTATGCATCTTATAAAGCAGAAAAAGAGAATCCTACAGTTCATTCTTTAAAT    1140
ATTGACTCTTCTGTGGAAAATTTAAATTAG                                   1170
```

Figure 4.1 The location of pre-mRNA 3' cleavage sites in pMET-gch
Cleavage sites are in bold, italicised and underlined. Start and stop codons are in bold & underlined.

In *S.cerevisiae*, efficient processing of the 3' ends of mRNAs is believed to require the presence of some or all of a number of elements around the cleavage and

polyadenylation site, or 3' terminus; an efficiency element, UAUAUA, AUAUAU or similar, generally positioned between 30 and 80 nt upstream of the 3' terminus, a positioning element, AAAAUA, AAGAAA or similar, positioned between 13 and 35 nt upstream and two U-rich elements, a pre-cleavage element, up to 15 nt upstream, and a downstream element, up to 30 nt downstream of the 3' terminus (Graber *et al.*, 1999). The site of pre-mRNA cleavage and polyadenylation is generally described as a pyrimidine followed by one or more A nt (Guo & Sherman, 1996)(Figure 4.2).

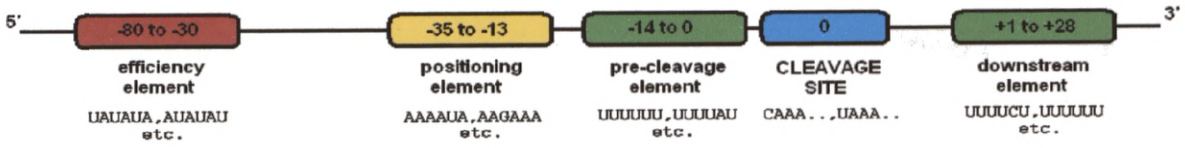


Figure 4.2 Locations and typical sequences of *S.cerevisiae* pre-mRNA 3' end processing motifs

The areas of the *Pf gch-1* gene around the mRNA cleavage sites contain a number of potential 3' end processing signals (Figure 4.3). This is unsurprising considering the A+T rich nature of the malarial genome (1.6). Initially, two experimental strategies were devised in order to attempt to overcome the problem of truncation of the mRNA transcripts of the gene in yeast.

Firstly, expression of *Pf gch-1* has been attempted in *S.cerevisiae* strains with deficiencies in RNA14 and RNA15, two proteins implicated in yeast mRNA 3' end-processing (Minvielle-Sebastia, 1994). Growth of these cells was, however, sickly, and no full length *Pf gch-1* transcript was observed (Verrall, B., personal communication).

Secondly, we resolved to remove the most powerful end processing motifs from the affected areas of the gene by changing the codon usage where necessary and possible, through site-directed mutagenesis. This approach has proved partially successful in the expression of the cDNA coding for *P.falciparum* topoisomerase II in yeast. Efficiency and positioning elements believed to be responsible for truncation of the topoisomerase II transcript at 376 nt were disrupted by mutagenesis. As a result a longer, although still truncated transcript of 617 nt was produced, which again terminated in a region characterised by putative 3' end processing motifs (Sibley *et al.*, 1997).

A

```
AATGAGAATGTTAATAATATTAATAAACAGTCTCTTAAGGATTCTATACTATTTGACAAT   420
ATAAATGAGAAAGAATATTTTAACGAAACCAAAGAAGAAAATAAGGAAGGAAATAAAAGT   480
AATGATATAGAAAAGATAAATTGTATGAAAGTGAAGAAGAAAACTGTTAAAAAGAATAAA   540
```

B

```
AAGAAGATTAATAAAATTATAAACAATAAGAATAAAATATCTAAATCAAATGACATAGAA   600
GAACAAATAATTAATATTAGTAAACATATATATAAAATATTAAATATATCCAAATTACCA   660
AAATGTGATATATTAAAAAGAACAAATAGAAGATATGCTGAAACATTTTTATATTTAACT   720
```

C

```
AAGAAGATTAATAAAATTATAAACAATAAGAATAAAATATCTAAATCAAATGACATAGAA   600
GAACAAATAATTAATATTAGTAAACATATATATAAAATATTAAATATATCCAAATTACCA   660
AAATGTGATATATTAAAAAGAACAAATAGAAGATATGCTGAAACATTTTTATATTTAACT   720
```

Figure 4.3 Potential mRNA 3'end processing signals in *Pf gch-1*
Sequences possibly acting as end processing signals (based on interpretation of the results presented by Graber *et al,* (1999)) at the cleavage sites at around **A,** 450 nt; **B,** 660 nt; **C,** 675 nt.

Key:

CLEAVAGE SITE

EFFICIENCY ELEMENT
loosely defined as ATATAT or TATATA, maximum 1 substitution, 30-90 nt upstream of cleavage site

POSITIONING ELEMENT
loosely defined as minimum 4 A residues in 6, 10-40 nt upstream of cleavage site.

PRE-CLEAVAGE/DOWNSTREAM ELEMENT
loosely defined as minimum 4 T residues in 6, from 15 nt upstream to 40 nt downstream of the cleavage site.

4.2 : Orientation of *Pf gch-1* in pMETplus

It is vital that the selection primer, conferring a change to a unique restriction site, and the mutagenesis primer(s), producing the desired changes in the target sequence, anneal to the same strand of the target plasmid in the site-directed mutagenesis protocol used (2.9). It was therefore necessary to confirm the orientation of the *Pf gch-1* gene in the expression vector pMETplus.

PCRs were set up with pMET-gch as template using the kpnsrf selection primer (designed to convert a unique *Kpn*I restriction site, upstream of the transcription start point in pMETplus, into an *Srf*I site; Appendix V) and either one of the cyclo12 and cyclo13 primers (gene specific primers with opposite orientations; Appendix V).

The presence of extension product of around the expected size (approx. 700bp) only in the mix using the kpnsrf/cyclo13 primer combination (Figure 4.4) confirmed that the mRNA synonymous strand of the gene was on the same strand as the transcription start point in pMETplus, enabling design of the gene-specific mutagenic primers.

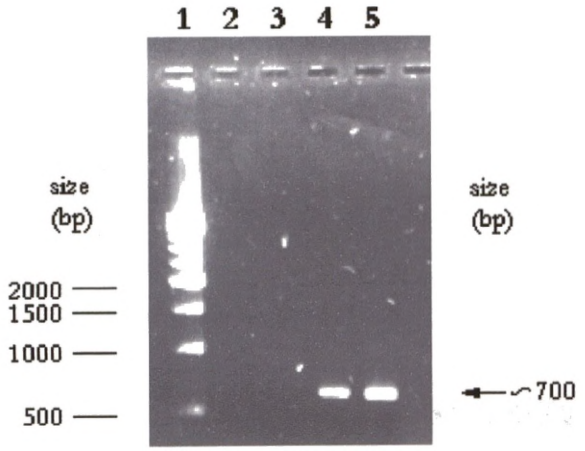


Figure 4.4 Orientation of *Pf gch-1* in pMETplus
Lane numbers: 1, molecular weight markers; 2/3, kpnsrf/cyclo12 primer combination; 4/5, kpnsrf/cyclo13 primer combination.

4.3 : Mutagenesis primers

3 mutagenic primers (Figure 4.5; Appendix V) were designed for use in site-directed mutagenesis:

gtpmut1; designed to disrupt the cleavage site at around 450 nt from the start codon and the potential positioning element around 20 nt upstream of this (see Figure 4.3).

gtpmut2; designed to disrupt the strong potential efficiency element around 30-50 nt upstream of the cleavage sites of the longer transcripts and the A+T rich tract of sequence just downstream of this efficiency element with the potential to act as either an efficiency or positioning element.

gtpmut3; designed to disrupt the cleavage site at around 675 nt and an A+T rich sequence immediately upstream with the potential to act as a downstream or pre-cleavage element at the cleavage sites of the longer transcripts or as an efficiency element at any as yet unidentified cleavage sites further downstream.

All primers used in site-directed mutagenesis were 5' phosphorylated (2.4.4).

Figure 4.5 Mutagenesis primers
Oligonucleotide sequence above, corresponding target DNA sequence below, corresponding polypeptide sequence between the two. Desired mutations are highlighted in black, target sequences are labelled as per Figure 4.3.

4.4 : Incorporation of the mutagenic primer sequences

A modified version of the site-directed mutagenesis protocol (2.9) was carried out whereby pMET-gch was incubated with all three mutagenic primers simultaneously, in addition to the kpnsrf selection primer (the manufacturer's protocol describes the annealing of just one mutagenesis primer).

Transformation yielded several hundred potentially mutant colonies as expected. 10 colonies were screened for incorporation of the mutagenic primers by colony PCR (2.5.8) and DNA sequencing (2.8). One transformant (mut8) had incorporated the gtpmut3 primer only (data not shown). The other transformants showed no mutations.

pMET-gch mut8 plasmid DNA was isolated (2.3.3) and resequenced to confirm incorporation of gtpmut3 and check for extraneous mutations. *Kpn*I and *Srf*I digests of the DNA (2.4.1) were also carried out to confirm the change in unique restriction site had occurred (data not shown).

Subsequently, additional site-directed mutageneses were carried out using pMET-gch mut8, incubated with gtpmut1 and gtpmut2 simultaneously, and also with each primer separately, in addition to the srfkpn selection primer (designed to convert the incorporated *Srf*I restriction site back to a *Kpn*I site; Appendix V).

6 potentially mutant colonies from each experiment were screened for incorporation of the mutagenic primers as described above. Two transformants (mut81 and mut85) from the experiment using both primers showed incorporation of the gtpmut1 primer only. The PCR product amplified from the mut81 transformant cells also showed two extraneous silent point mutations close to the gtpmut1 mediated mutations (data not shown), and this transformant was therefore discarded in a spirit of conservatism. The other transformants showed no mutations.

pMET-gch mut85 plasmid DNA was isolated and resequenced to confirm incorporation of gtpmut1 and check for additional extraneous mutations. *Kpn*I and *Srf*I digests of the DNA were also carried out to confirm the change in unique restriction site had occurred (data not shown).

A final round of site-directed mutagenesis was then carried out using pMET-gchmut85, incubated with the gtpmut2 and kpnsrf primers.

10 potentially mutant colonies were screened for incorporation of the mutagenic primer. One transformant (mut856) showed incorporation of the gtpmut2 primer (Figure 4.6). The other transformants showed no mutations.

pMET-gch mut856 plasmid DNA, subsequently described as pMET-gchmut^3, was isolated and resequenced to confirm incorporation of gtpmut3 and check for extraneous mutations (data not shown).

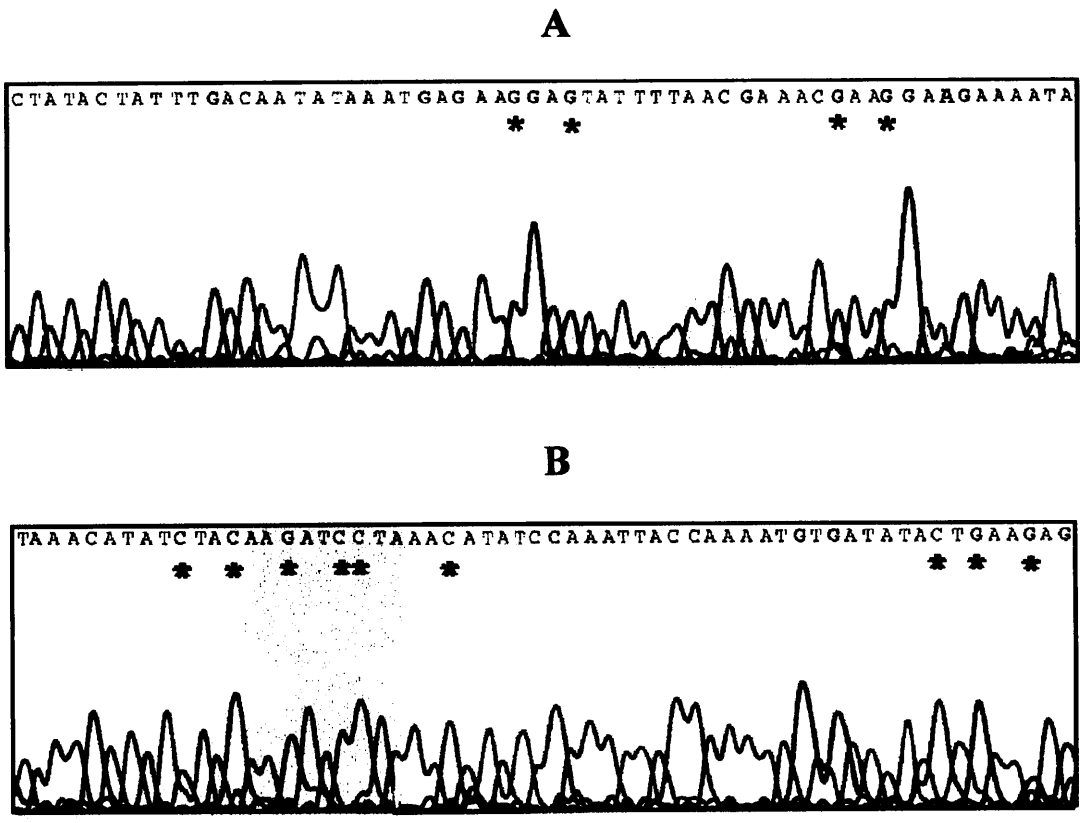


Figure 4.6 Sequence of pMET-gchmut^3
A; sequence tract around cleavage site at 450nt from start codon, B; sequence tract around cleavage sites at 660nt and 675nt from start codon. Incorporated mutations are marked by black asterisks.

4.5 : Complementation test of pMET-gchmut^3 in the *S.cerevisiae* knockout mutant

pMET-gchmut^3 was transformed (2.5.7) into the haploid yeast *FOL2* knockout mutant FY1679 Δ*FOL2*::*kanMX4* HP1, along with the original pMET-gch clone and the vector pMETplus alone. The transformants were selected on the basis of their ability to grow on uracil depleted synthetic media (2.2.4), due to the *URA3* gene encoded on the plasmid (Appendix VI), and further confirmed using PCR analysis (results not shown.)

Transformants containing each of the three plasmids were grown to mid-log phase in synthetic liquid media, and 10μl of each culture was subsequently inoculated in

a series of dilutions (10^0, 10^{-1} and 10^{-2}) onto two separate synthetic media plates; with and without a 50 μg/ml folinic acid supplement respectively. Each of the transformants grew well on the media with the folinic acid supplement, as expected (Figure 4.7A). In addition, the cells containing pMETplus and pMET-gch were unable to grow on the media lacking the supplement. Disappointingly however, the cells transformed with pMET-gchmut^3 were also unable to grow on this media (Figure 4.7B.)

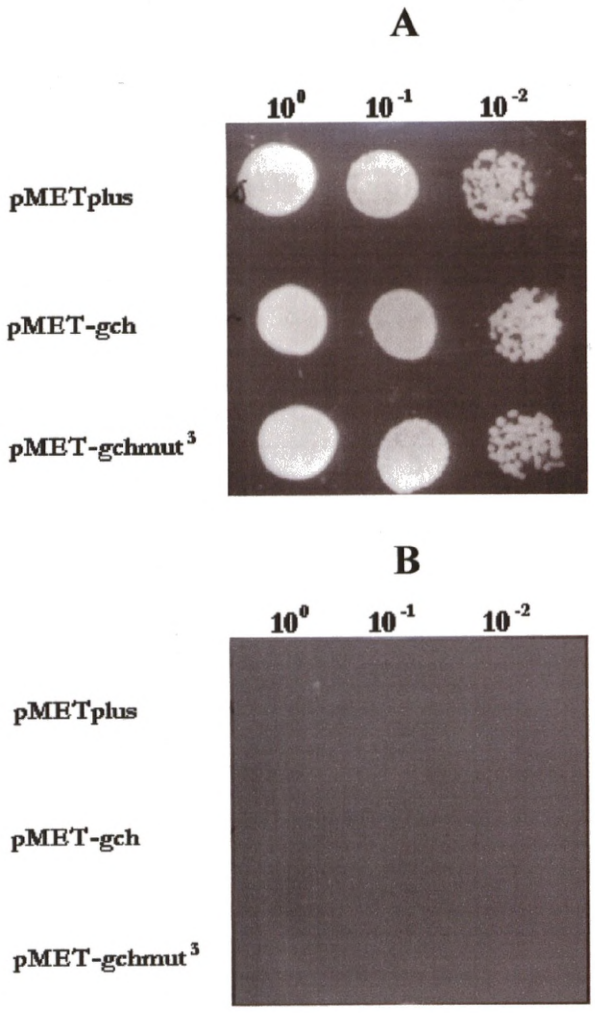


Figure 4.7 Complementation test of pMET- gchmut^3 in FY1679 Δ*FOL2*::*kanMX4* HP1 The transformants carrying either pMETplus alone, pMET-gch, or pMET-gchmut^3 were inoculated in a series of dilutions (10^0, 10^{-1} and 10^{-2}) onto two separate synthetic media plates; A) with a 50 μg/ml folinic acid supplement and B) without this supplement.

4.6 : Production of additional clones to test for complementation in the *S.cerevisiae* knockout mutant

Following the failure of pMET-gchmut^3 to rescue the yeast GTCase knockout, it was decided to explore alternative approaches to expression of *Pf gch-1* in *S.cerevisiae.* Given the general problems with expressing malaria parasite genes in yeast due to the A+T rich nature of the parasite genome (1.6), and the associated preponderance of yeast pre-mRNA processing motifs (4.1), it was decided to dispense with the portion of *Pf gch-1* corresponding to the N-terminal portion of the enzyme, as in 3.8, in some expression constructs, in the hope that this would facilitate transcription up to and beyond the stop codon, without affecting basic enzyme activity. It was also resolved to attempt *Pf gch-1* expression using the p424 GPD-U4 expression vector (pGPD for brevity, Appendix VI), a multicopy plasmid which expresses recombinant genes using an extremely powerful *S.cerevisiae* glyceraldehyde phosphate dehydrogenase (GPD) promoter, and has shown success in expression of the malaria parasite gene encoding the folate biosynthetic enzyme *Pf dhfs/fpgs* in yeast (Salcedo *et al.,* 2000). It was decided to take advantage of the modifications present in the version of *Pf gch-1* present in pMETgchmut^3, and to use the modified gene in the new expression constructs.

Five oligonucleotide primers, gtpmet1, gtpmet2, gtpgpd1, gtpgpd2 and gtpgpd3 (Appendix V) were designed to produce a *Pf gch-1* derived DNA fragment with terminal *Spe*I and *Sac*I sites suitable for unidirectional cloning into pMETplus, and two *Pf gch-1* derived DNA fragments with terminal *Bam*HI and *Xho*I sites suitable for unidirectional cloning into pGPD.

The inserts produced were as follows:-

an approximately 1200bp product amplified by gtpgpd1 and gtpgpd3, corresponding to the full length modified *Pf gch-1* gene (gchmut^3), for use with pGPD, and two similar products of approximately 600bp amplified by gtpmet1 and gtpmet2, and gtpgpd2 and gtpgpd3, corresponding to the C-terminal half of the gene product (Cgchmut^3,) for use with pMETplus and pGPD respectively.

PCRs (2.7.1) were set up using pMET-gchmut^3 as template, and the three different primer combinations. The amplification products were visualised on an agarose gel (2.6.1) and purified from the reaction mixes (2.3.1), then digested with the cloning enzymes (2.4.1), as were the vectors.

The pMET-Cgchmut^3 and pGPD-Cgchmut^3 constructs were subsequently produced by cloning (2.5.1)(data not shown). However, early attempts to produce the pGPD-gchmut^3 construct were unsuccessful.

It was therefore decided to attempt cloning of this species using a different pair of restriction enzymes. The terminal *Spe*I site serendipitously produced by amplification using the gtpgpd1 primer (Appendix V) facilitated this experimental strategy, but it was necessary to design an additional oligonucleotide primer, gtpgpd4 (Appendix V), in order to incorporate a terminal *Eco*RI site at the 3' end of the gene.

A PCR was then set up using pMET-gchmut^3 as template and the gtpgpd1/gtpgpd4 primer combination. The amplification product was visualised on an agarose gel and purified from the reaction mix, then digested with the cloning enzymes, as was pGPD. Cloning using this vector/insert combination yielded the pGPD-gchmut^3 construct at the first attempt (data not shown).

It was also decided that, in order to provide a positive control for the transformation process (2.5.7) and the expression vectors used, clones of *FOL2*, the gene encoding GTCase in yeast (Nardese *et al,* 1996), in each of the expression vectors used should be produced. 3 oligonucleotide primers, fol2A, fol2B and fol2C (Appendix V) were designed to produce a *FOL2* derived DNA fragment with terminal *Spe*I and *Sac*I sites suitable for unidirectional cloning into pMETplus, and a *FOL2* derived DNA fragment with terminal *Spe*I and *Cla*I sites suitable for unidirectional cloning into pGPD. The inserts produced were as follows:-

two similar products of approximately 750bp, corresponding to the full length *FOL2* gene, amplified by fol2A and fol2B, and fol2A and fol2C, for use with pMETplus and pGPD respectively.

Yeast colony PCRs (described in 2.5.9) were set up using cells of FY1679wt HP1 (Appendix I), a haploid line wild type with respect to *FOL2,* and the two primer combinations. The amplification products were visualised on an agarose gel and purified from the reaction mixes, then digested with the cloning enzymes, as were the vectors.

Both the pMET-fol2 and pGPD-fol2 constructs were subsequently produced by cloning (data not shown).

4.7 : Complementation test of additional *Pf gch-1* derived clones in the *S.cerevisiae* knockout mutant

The five new constructs, three *Pf gch-1* derived, and two *FOL2* derived (4.6) were transformed (4.5.7) into the haploid yeast *FOL2* knockout mutant, along with pGPD alone.

Transformant colonies, containing each of these plasmids along with knockout cells transformed with pMETplus, pMET-gch and pMET-gchmut^3, and wild-type haploid cells, were streaked onto separate segments of synthetic media plates (2.2.4) some with and some without a 50μg/ml folinic acid supplement (Figure 4.8).

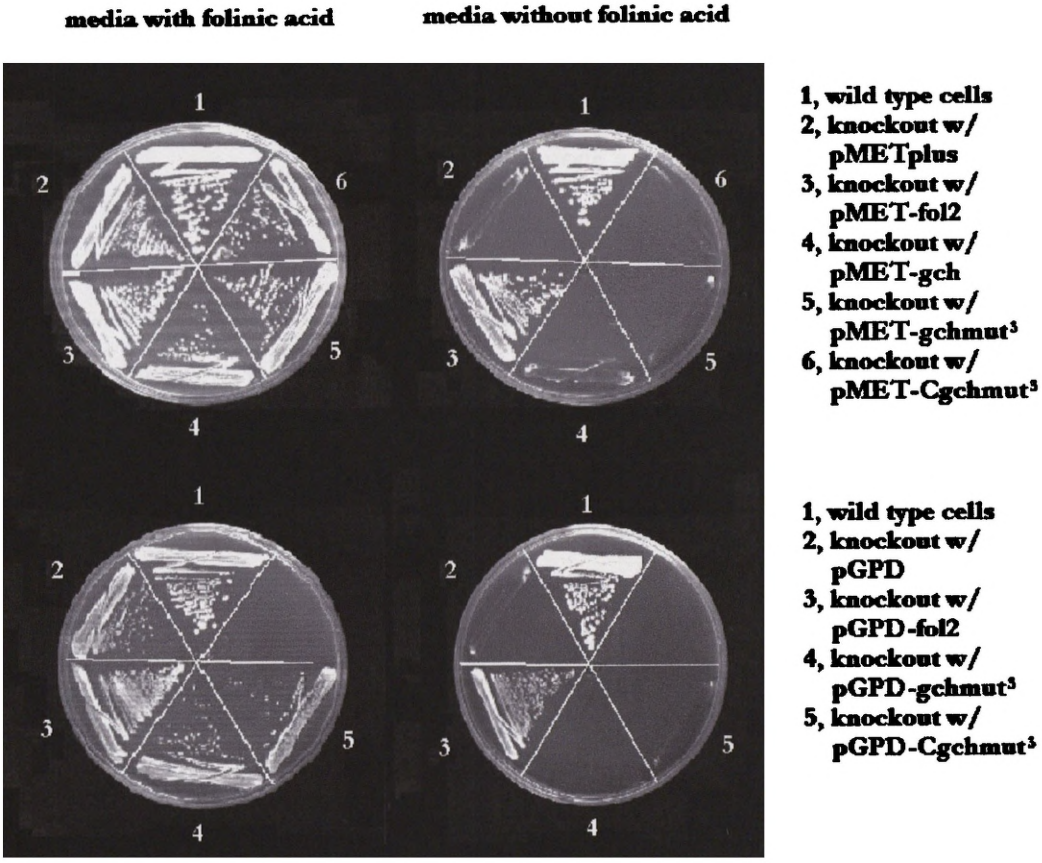


Figure 4.8 Complementation test of additional *Pf gch-1* derived clones in FY1679 Δ*FOL2*::*kanMX4* HP1

Although the knockout cells transformed with the pMET-fol2 and pGPD-fol2 constructs were able to propagate on the media without the folinic acid supplement, demonstrating the suitability of both vector systems for testing for complementation of the *FOL2* knockout, the new *Pf gch-1* derived constructs, pMET-Cgchmut^3, pGPD-gchmut^3 and pGPD-Cgchmut^3 were unable to grow on this media, and hence failed to

complement the knockout, where pMET-gch and pMET-gchmut[3] had failed before (4.5).

4.8 : Partial complementation test of *Pf gch-1* derived clones in the *S.cerevisiae* knockout mutant

Although none of the *Pf gch-1* derived constructs were able to effect rescue of the yeast knockout in the absence of folate (4.5, 4.7), it was thought possible that these constructs might enable growth at folinic acid concentrations too low to support growth by untransformed mutant cells.

In order to establish parameters for a partial complementation test, growth of untransformed knockout mutant yeast on solid media at a variety of folinic acid concentrations was observed and qualitative assessments made (results not shown).

It was determined that apparently normal, if comparatively slow, colony growth was possible at folinic acid concentrations down to approximately 5μg/ml (compared to the standard concentration of 50μg/ml), while at approximately 1μg/ml, no growth was observed. At approximately 2μg/ml 'sickly' growth, typified by the emergence of some small, incipient colonies, which were seemingly unable to grow beyond this initial phase, was observed.

It was therefore decided to carry out partial complementation tests on the *Pf gch-1* derived constructs at approximate folinic acid concentrations of 1μg/ml and 2μg/ml (parallel plates, spread with folinic acid to a final concentration of approximately 5μg/ml and without a folinic acid supplement were also used, as positive and negative controls).

Knockout yeast cells containing the *Pf gch-1* derived clones, along with wild type cells, untransformed knockout cells and cells transformed with the pMETplus and pGPD plasmids were streaked onto separate segments of synthetic media plates at each of the four folinic acid concentrations (Figure 4.9).

As was the case with the untransformed knockout yeast, the yeast transformed with each of the *Pf gch-1* derived constructs were unable to grow at a folinic acid concentration of approximately 1μg/ml. At 2μg/ml these cells exhibited either 'sickly' growth or no growth at all, but not normal growth.

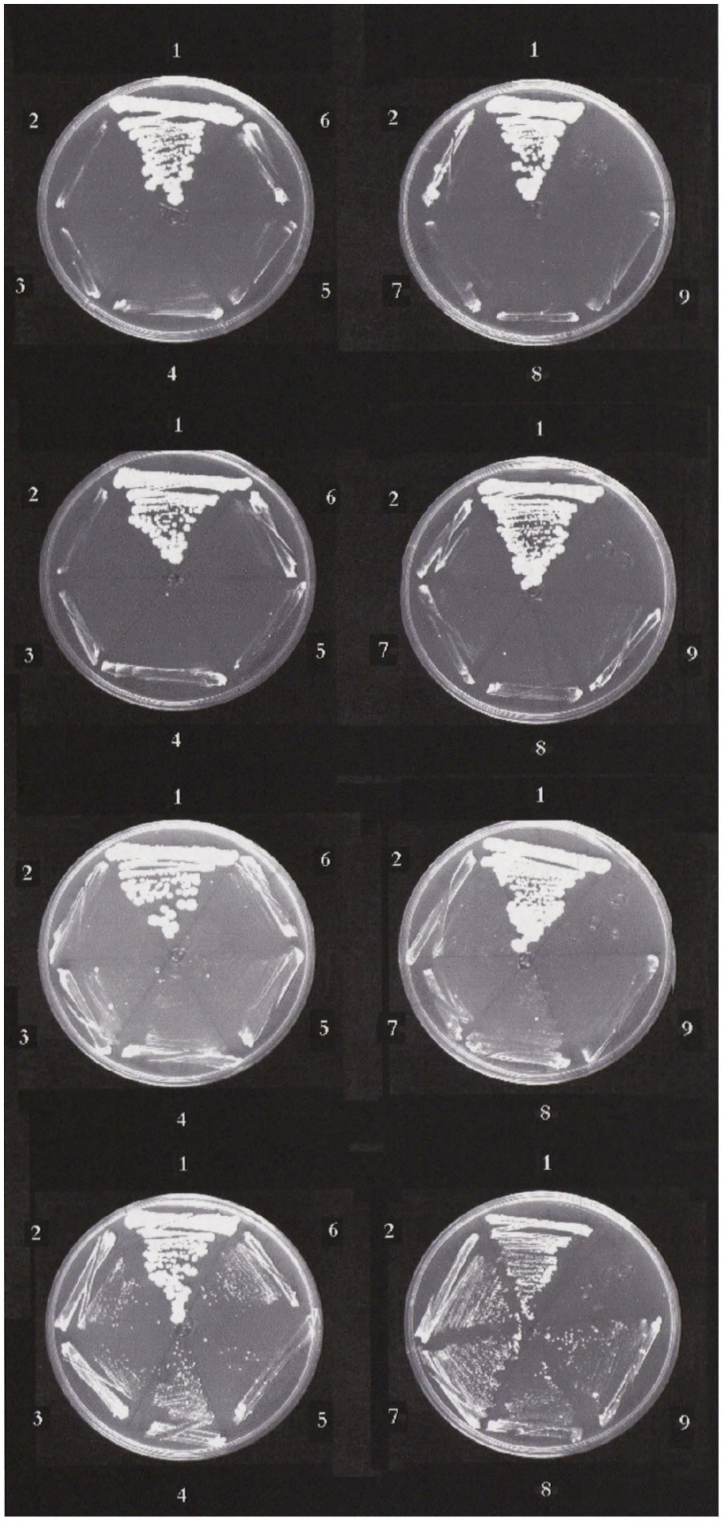


Figure 4.9 Partial complementation test of *Pf gch-1* derived clones in FY1679 Δ*FOL2*::*kanMX4* HP1 1, wild-type cells; 2, knockout cells; 3, knockout w/ pMETplus; 4, knockout w/ pMET-gch; 5, knockout w/ pMET-gchmut[3]; 6, knockout w/ pMET-Cgchmut[3]; 7, knockout w/ pGPD; 8, knockout w/ pGPD-gchmut[3]; 9, knockout with pGPD-Cgchmut[3]. Plates containing folinic acid at; A, 0μg/ml; B 1μg/ml; C, 2μg/ml; D, 5μg/ml.

It is therefore impossible to claim on this evidence that any of the *Pf gch-1* derived constructs produced are able to effect either full or partial complementation of the *FOL2* knockout.

4.9 : Investigation of the transcripts from the yeast transformants

Although the modified version of *Pf gch-1* carried on the pMET-gchmut^3 clone was unable to complement the GTCase deficiency in the yeast knockout mutant (4.5), it was resolved to analyse the mRNA transcripts produced by the expression vector, in order to discover what change in transcription pattern had been effected by the removal of the end-processing motifs from the gene. Each of the three initial yeast transformants, (haploid knockout cells transformed with pMETplus, pMET-gch and pMET-gchmut^3), was grown in liquid synthetic medium supplemented with folinic acid. Total RNA was subsequently extracted from each culture (2.3.4), and an aliquot of each RNA preparation was subjected to formaldehyde denaturing agarose gel electrophoresis (2.6.2) along with aliquots of *E.coli* and *P.falciparum* RNA used as size markers (Figure 4.10). The RNA was then transferred from the gel to a nylon membrane by northern blot (2.10.1), which was subsequently probed with a DIG-labelled RNA fragment (2.10.2, 2.10.4) complementary to *Pf gch-1,* synthesised from a copy of *Pf gch-1* in the plasmid pBluescript (Stratagene). Following development (2.10.5), no signal was observed.

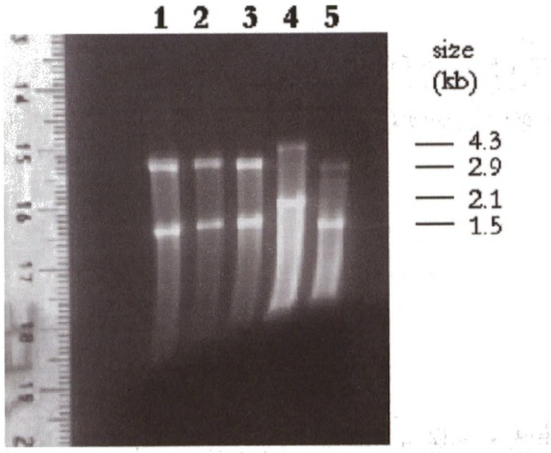


Figure 4.10 RNA gel for Northern blot analysis using DIG-labelled probe
Samples are total RNA preparations. Lane numbers; 1, knockout w/ pMETplus; 2, knockout w/ pMET-gch; 3, knockout w/ pMET-gchmut^3; 4, *P.falciparum*, K1 lineage; 5, *E.coli.*

When a number of subsequent attempts to detect transcript using chemiluminescently labelled RNA fragments and the associated hybridisation and wash regime were similarly unsuccessful, it was decided to switch to hybridising the membrane with radiolabelled DNA probe, synthesised from a PCR product corresponding to the full length *Pf gch-1* (2.10.3, 2.10.6). The advantages conferred by this protocol were deemed to include greater sensitivity, a more robust probe and the reduction in wash and development steps, thus reducing the likelihood of RNA degradation.

When this change of approach proved unable to effect detection (results not shown) it was postulated that a low occurrence of *Pf gch-1* derived transcripts in the yeast cells, as a result of transcriptional difficulties including those previously detailed (4.1), might be partially responsible for the observed failure. It was therefore decided to probe RNA samples with rabiolabelled DNA sourced from another gene, in order to act as a positive control for the protocol.

The initial gene of choice was the actively transcribed *S.cerevisiae act-1* (actin) gene. Radiolabelled probe was synthesised (2.10.3) from a DNA fragment excised from a clone of *act-1* in the vector pGEM-T (Promega), and this was used to hybridise a membrane onto which the same yeast aliquots had been blotted as before (Figure 4.10). Again, no signal was observed (results not shown). This reinforced minor doubts about the provenance of the template fragment used to synthesise the probe. Fortunately, with the production of the pMET-fol2 and pGPD-fol2 constructs, with which complementation of the *FOL2* knockout, and hence production of *FOL2* derived transcript, had been proved (4.6, 4.7), an attractive alternative positive control gene was readily available.

Radiolabelled probe was synthesised from a PCR product corresponding to the full length *FOL2* gene (2.10.3) and this was used to hybridise a membrane onto which aliquots of total RNA from yeast cells transformed with pMET-plus, pMET-fol2, pGPD and pGPD-fol2, along with *E.coli* and *P.falciparum* RNA as size markers, had been blotted. Following washes, the membrane was exposed to X-ray film, and following development this was checked for signal (Figure 4.11)

A strong band of approximately 1.35kb was observed in the lane corresponding to the RNA sample from the knockout cells transformed with pGPD-fol2. Although the size value must be treated with some caution due to the relative paucity of comparators,

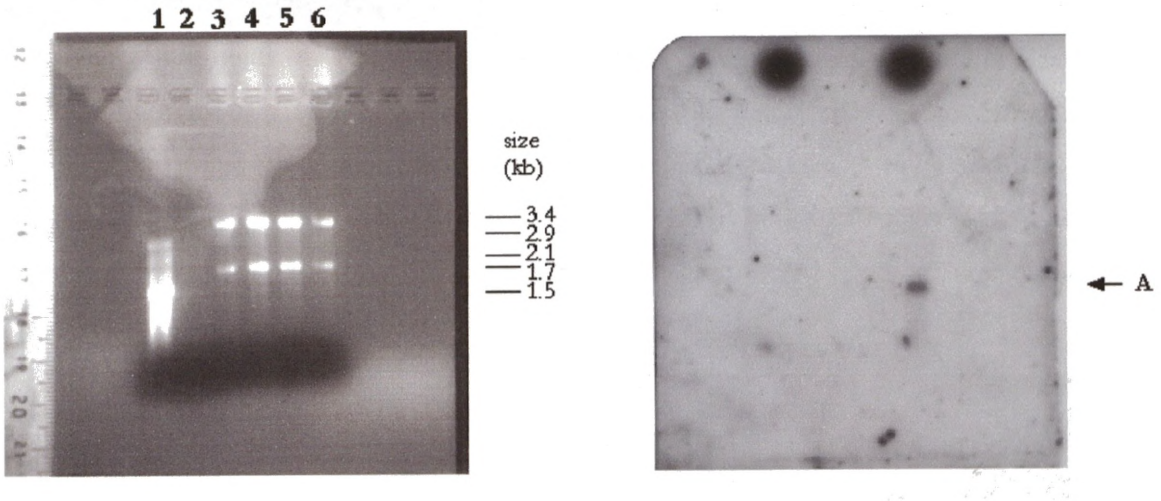


Figure 4.11 Positive control northern blot using *FOL2* derived probe
Gel on left, autoradiograph on right. Lane numbers for gel; 1, *E.coli*; 2, *P.falciparum*, K1 lineage (underloaded); 3, knockout w/ pMETplus; 4, knockout w/ pMET-fol2; 5, knockout w/ pGPD; 6, knockout w/ pGPD-fol2. Signal **A**, marked on the autoradiograph, corresponds to a band of approx. 1.35kb in lane 6. The large dark spots at the top of the autoradiograph correspond to spots of denatured template DNA, dried on to the membrane post-transfer as a control for the hybridisation procedure.

and the distortion in the gel, this fits in reasonably with the size of the *FOL2* gene (750bp), allowing for the 5' and 3' untranslated regions of the transcript. The lack of signal in the lane corresponding to the pMET-fol2 transformed yeast RNA is almost certainly due to a combination of the multicopy nature of the pGPD plasmid, as opposed to the single copy pMETplus plasmid (Appendix VI), and the relative weakness of the *MET3* promoter on pMETplus compared to the potent *GPD* promoter on pGPD. These results clearly demonstrated that the RNA preparation/ northern blot/ probe synthesis/ hybridisation regime used was capable of detecting transcript given sufficient quantities. However, the failure to detect transcript from pMET-fol2 indicated that sensitivity was a problem in this system.

It was therefore decided to prepare RNA from the yeast cells using the RNeasy Midi Kit from Qiagen, rather than the similar Mini Kit, which had been used previously (2.3.4), in order to boost levels of RNA. When these preparations were run on a gel prior to transfer, the midiprepped RNA, although seemingly much more highly concentrated than the miniprepped material, appeared as a bright smear rather than the usual bright bands of large and small subunit rRNA, contrasted with a faint smear of mRNA (Figure 4.12). This suggested that degradation of the RNA preparations had occurred. Nevertheless, the RNA was blotted onto a membrane, and this was hybridised

with a *FOL2* derived probe. Following development of the resulting autoradiograph, a dark, but highly smeared or elongated band, extending to a maximum 1.45kb, was observed in the lane corresponding to the RNA sample from the knockout cells transformed with pGPD-fol2 (Figure 4.12). This reinforced the previous observation of *FOL2* transcript of around 1.35kb in these cells (Figure 4.11). Again, no signal was observed from the pMET-fol2 transformed yeast RNA.

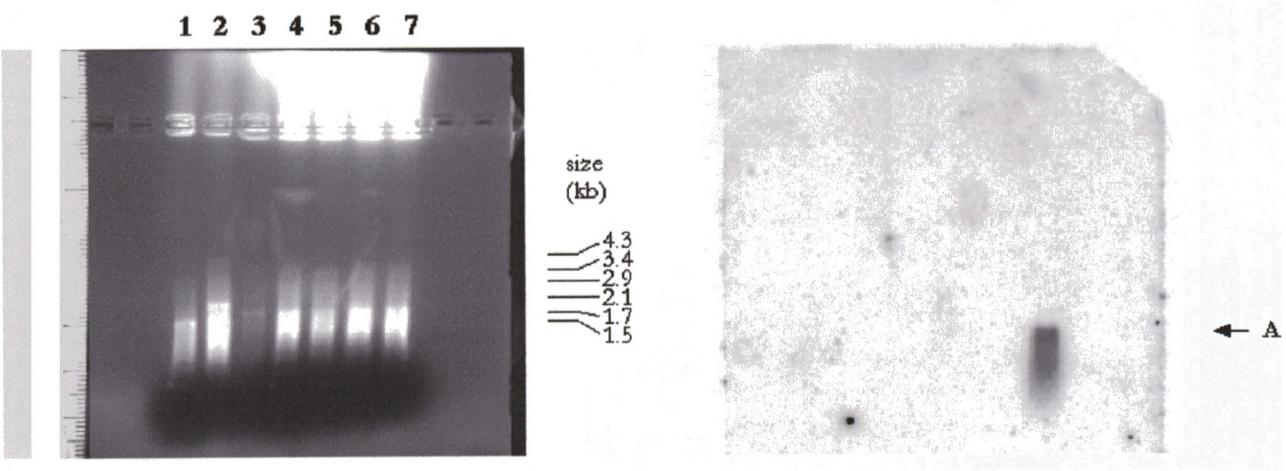


Figure 4.12 Northern blot using *FOL2* derived probe and midiprepped RNA
Gel on left, autoradiograph on right. Lane numbers for gel; 1, *E.coli*; 2, *P.falciparum*, K1 lineage (underloaded); 3, knockout w/ pMETplus, miniprep for comparison; 4, knockout w/ pMETplus; 5, knockout w/ pMET-fol2; 6, knockout w/ pGPD; 7, knockout w/ pGPD-fol2. Signal **A**, marked on the autoradiograph, corresponds to a band of approx. maximum 1.45kb in lane 7.

As a result of the failure of the midiprep protocol adopted to enable detection of transcript from pMET-fol2, coupled with the apparent problems with RNA degradation that this protocol effected, it was decided to use the original miniprep protocol to prepare RNA for subsequent northern analysis. However, given the problems with sensitivity experienced, and in light of the fact that the pGPD expression system produces greater quantities of transcript than the pMETplus vector, it was decided to produce a clone of the unmodified *Pf gch-1* gene in pGPD. In the event that *Pf gch-1* derived transcripts could be detected from pGPD clones only, this new construct would provide, in comparison with pGPD-gchmut[3], information with respect to the effect of mutagenesis on the pattern of transcription of *Pf gch-1* in yeast. A PCR (2.7.1) was therefore set up using pMET-gch as template and the gtpgpd1/gtpgpd4 primer combination (4.6, Appendix VI). The amplification product was visualised on an

agarose gel (2.6.1) and purified from the reaction mix (2.3.1), then digested with the cloning enzymes (2.4.1), as was pGPD.

Following cloning (2.5.1) just a single recombinant transformant *E.coli* colony was recovered. When the insert of the plasmid carried by this strain was sequenced, an extraneous point mutation, changing a T to a C at nucleotide position 349, corresponding to an amino acid change of cysteine to arginine at position 117 (Appendix II) was observed (results not shown). As a result of time constraints, it was decided to retain this clone (pGPD-gch') for use in subsequent experiments, rather than attempting to produce a clone without extraneous mutations. The mutation occurred at a position fairly distantly removed from the truncation points identified in transcripts from pMET-gch (Figure 4.1), and was therefore deemed unlikely to affect the transcription pattern. The new construct was transformed (2.5.7) into the haploid knockout mutant. Unsurprisingly, when streaked onto solid synthetic media (2.2.4) without folinic acid, the transformant cells were unable to grow (results not shown). Total RNA was subsequently prepared from these cells, along with knockout cells transformed with pMETplus, pMET-gch, pMET-gchmut[3], pMET-Cgchmut[3], pGPD, pGPD-gchmut[3] and pGPD-Cgchmut[3] (4.6). Aliquots of each RNA preparation were run on a gel along with aliquots of *E.coli* and *P.falciparum* RNA (Figure 4.13). The RNA was then blotted to a membrane, and this was hybridised with a radiolabelled *Pf gch-1* derived probe. Disappointingly, following exposure of the membrane to X-ray film, and subsequent development of the autoradiograph, no signal was observed.

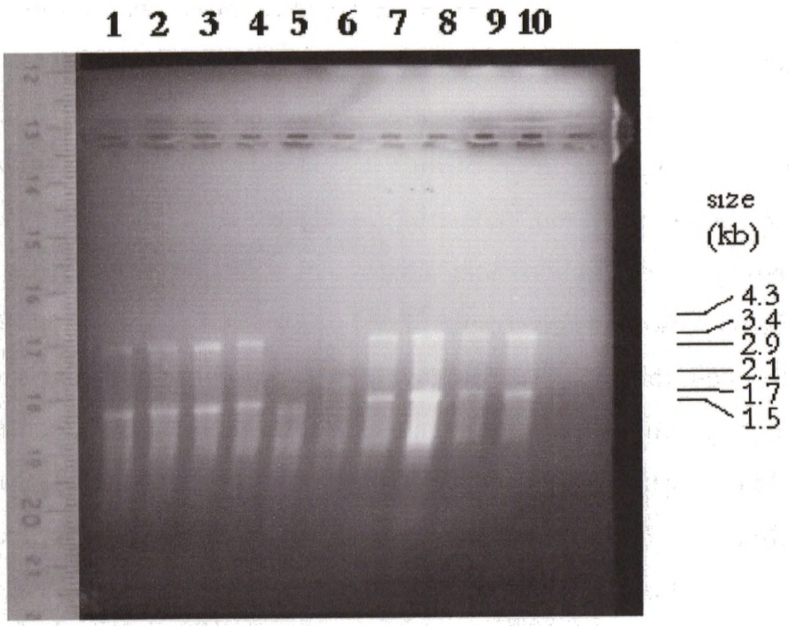


Figure 4.13 RNA gel for Northern blot analysis using radiolabelled, *Pf gch-1* derived probe Lane numbers; 1, knockout w/ pMETplus; 2, knockout w/ pMET-gch; 3, knockout w/ pMET-gchmut[3]; 4, knockout w/ pMET-Cgchmut[3]; 5, *P.falciparum*, K1 lineage; 6, *E.coli*; 7, knockout w/ pGPD; 8, knockout w/ pGPD-gch'; 9, knockout w/ pGPD-gchmut[3]; 3, knockout w/ pGPD-Cgchmut[3].

4.10 : Expression in a human cell line

The various difficulties encountered in heterologous expression of the *Pf gch-1* gene in *S.cerevisiae* and *E.coli* (this chapter & Chapter 3), particularly in terms of expression of full length, biologically active gene product, can be either tentatively or explicitly ascribed to incompatibilities between the molecular biology of *P.falciparum* and the respective host organisms. We therefore resolved to investigate alternative host cell types for the expression of *Pf* GTCase, in the first instance vertebrate cell lines, in the hope that this strategy might yield favourable results. Initially, some consideration was given to the expression of the gene in chicken cells in which the gene encoding GTCase had been knocked out. However, Kaspers *et al.* (1997) have showed that inhibition of GTCase in chicken macrophage cells is non-lethal and results only in subtle changes in phenotype. These experimenters measured GTCase activity indirectly through detection of nitric oxide synthase activity; this enzyme requires the presence of the cofactor BH_4, the end product of the biosynthetic pathway in animals in which GTCase catalyses the first step (1.9). Disruption of a specific gene in chicken cells is a complex and lengthy procedure and, given the anticipated difficulties in monitoring

GTCase mutant rescue, it was decided that the vastly simpler procedure of overexpression of *Pf gch-1* following transient transfection of this gene into a human cell line was a preferable experimental strategy.

The expression vector chosen for this transfection study was pcDNA4/His MaxA (Invitrogen; Appendix VI). This vector contains the strong CMV (human cytomegalovirus) promoter and expresses gene product as a fusion with a cleavable N-terminal tag, including a polyhistidine tract allowing affinity purification and the Express epitope for easy detection of expressed product by immunoprobing (2.12.2). PCRs (2.7.1) were set up using a clone of *Pf gch-1* cDNA in pBS (Stratagene) as template, and the gtpgex1/gtpgex4 and gtpgex2/gtpgex4 primer combinations (3.2; Appendix V), in order to generate *Pf gch-1* derived DNA fragments (gch and Cgch respectively; see 3.2) with terminal *Eco* RI and *Not* I sites suitable for unidirectional cloning into pcDNA4/HisMaxA. The amplification products were visualised on an agarose gel (2.6.1) and purified from the reaction mixes (2.3.1), then digested with the cloning enzymes (2.4.1), as was the vector. Both constructs were subsequently produced by cloning (2.5.1)(results not shown).

Human embryonic kidney 293T cells were plated into 60mm culture dishes, and grown for a day to reach 80% confluence. 5μg of each construct, pcDNA-gch and pcDNA-Cgch, plus a control expression construct encoding β-galactosidase was then used to transfect the cells in one plate each using the calcium phosphate transfection protocol described in Doyle (1996). Depending on the cell type, calcium phosphate mediated transfection can result in up to 20% of the cells expressing the foreign DNA, it is therefore the method of choice for this type of transient transfection experiment (Sambrook *et al.*, 1989). The cells were then harvested 36 hr post-transfection and lysed in 400μl NP40 lysis buffer (Sambrook *et al.*, 1989) each.

In order to screen the cells for expression, the cells were then subjected to western analysis (2.12). Lysates from the cells transfected with each of the three constructs, plus the lysate from untransfected cells, were mixed with 10 x SDS-PAGE reducing buffer (2.1.12) and 5 x PAGE loading buffer (2.1.13) to final concentrations of 1 x each. Duplicate 25μl aliquots of each sample were then subjected to SDS-PAGE (2.6.3) and western transfer (2.12.1). The two blots were then subjected to immunoprobing (2.12.2) with mouse anti-Express epitope (Invitrogen)/mouse anti-

polyhistidine tag (Novagen) monoclonal antibody mix and chicken anti GST-Cgch yolk extract (3.7) as primary antibodies respectively (Figure 4.14).

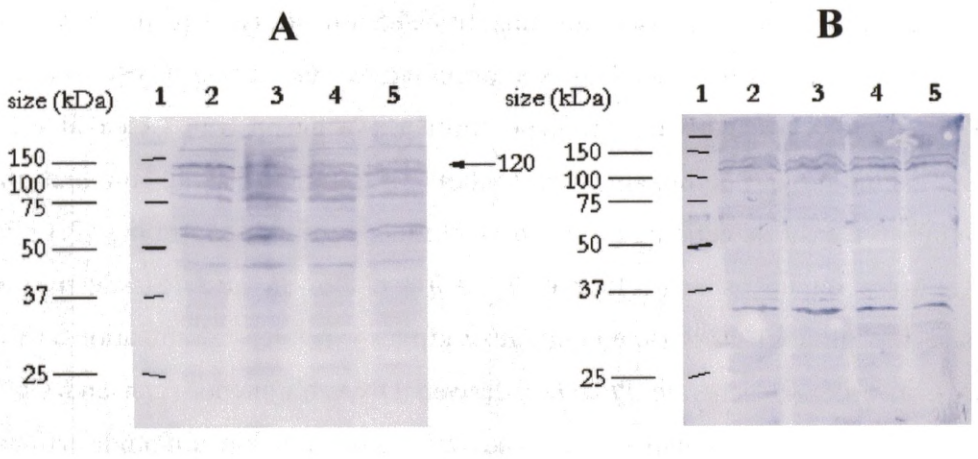


Figure 4.14 Screening of transfected human cells for expression of *Pf gch-1* derived gene products A) Blot probed with mouse anti-Express epitope/mouse anti-polyhistidine tag monoclonal antibody mix; Lane numbers: 1, Precision molecular weight marker; 2, lysate of positive control (pcDNA-lacZ) transfected cells; 3, lysate of untransfected cells; 4, lysate of pcDNA-gch transfected cells; 5, lysate of pcDNA-Cgch transfected cells.
B) Blot probed with chicken anti-GST-Cgch yolk extract; Lane numbers as A). n.b. positions of markers were added immediately post-transfer.

Unfortunately, the results yielded by the western analysis of the transfected cells were not promising. A band slightly above background of around the anticipated size (120 kDa) was noted when the cells transfected with the positive control (β-galactosidase encoding) plasmid were probed with the anti-Express epitope/anti-polyhistidine tag mix. However, no bands of the anticipated sizes (50 kDa and 27 kDa for pcDNA-gch and pcDNA-Cgch respectively), nor any additional bands not present in the lysate of untransfected cells, were noted in the lysates of the cells transfected with the *Pf gch-1* derived constructs when probed with either the anti-Express epitope/anti-polyhistidine tag mix, or the anti-GST-Cgch yolk extract, suggesting that expression of these recombinant genes was not occurring at a significant level in the transfected cells.

4.11 : Discussion

4.11.1 : Efficiency of mutagenic primer incorporation

Although successive rounds of mutagenesis (2.9) resulted in a clone of pMET-gch with all the desired mutations (4.4), the efficiency of primer incorporation was surprisingly low. In all, 64 segments of DNA sequence from 38 transformants were screened for mutations. The mutations were only present in four cases. This contrasts markedly with the normal efficiency quoted in the manufacturer's protocol of 70-95% of transformants showing incorporation of mutagenic primer sequence. A control experiment using mutagenesis primer, selection primer and target plasmid supplied by Stratagene produced a rate of incorporation of the mutagenesis primer of around 80% (results not shown).

It was originally suggested that the low efficiency observed was the result of the modified protocol whereby several mutagenic primers were annealed to the target DNA simultaneously. However, as the efficiency appeared no higher in experiments using the traditional protocol, this explanation is unlikely.

A low rate of mutagenic primer incorporation could be caused by an imbalance in the proportions of primers and target DNA; in this case either a low level of mutagenic primer compared with the other components or an excess of plasmid could result in the observed efficiency. However, given that the mutagenic and selection primers were supplied by the same source (MWG), while plasmid DNA was quantified prior to use (2.3.6), it seems unlikely that imbalances sufficient to lower the efficiency to such an extent occurred (the protocol recommends a primer:target sequence molar ratio of 100:1).

Although the primer-annealing step of site-directed mutagenesis is carried out at room temperature (23-25°C), low primer/target sequence melting temperatures (T_m) might affect efficiency. Although the T_m values given for the three mutagenic primers used are not particularly low (64-68°C) this would be affected by the mismatches between primer and target sequence inherent in site-directed mutagenesis. It is possibly instructive that the primer conferring the fewest mutations, gtpmut3 (3 point mutations), proved the easiest to incorporate, while that conferring the most, gtpmut2 (6 point mutations), proved most difficult.

It is also possible that the relatively large size of the plasmid (6.7 kb) or incomplete digestion by the selection enzyme played a part in lowering efficiencies, although *Kpn*I was used as the selection enzyme in both experimental and control reactions.

4.11.2 : Failure of northern hybridisation

Despite the detection of mRNA transcribed from the pGPD-fol2 plasmid (Figures 4.11, 4.12), the consistent failure to detect transcript from either pMET-fol2 or any of the *Pf gch-1* derived clones indicates that the sensitivity of the protocol for northern analysis was not sufficiently high to be of general use. While transcript was detected from a yeast gene carried in multiple copies per cell, we were unable to detect transcript either from parasite genes carried in presumably similar copy numbers, or from a yeast gene with just a single copy per cell. All probes used were assayed to determine the extent of labelling (2.10.2, 2.10.3) and found to be within reasonable limits of activity. Post-blot examination of both the gel and membrane under UV demonstrated that RNA transfer was indeed occurring, while the limited success experienced with the *FOL2* derived probe suggests that conditions conducive to probe/membrane bound RNA hybridisation were present. It is therefore most likely that the lack of success experienced was due to the poor quality of the RNA samples used.

The high susceptibility of mRNA to degradation by RNases is well documented (e.g. Garcia-Sanz, 1997), and extensive steps were undertaken in order to eliminate these enzymes from the solutions and glassware used in the preparation and manipulation of RNA, most obviously in terms of the use of the irreversible RNase inhibitor DEPC (2.3.4). Nevertheless, it is possible that RNA degradation was occurring to deleterious levels, most likely during the preparation steps (2.3.4), although this was not indicated by the appearances of the gels in most cases. It is possible that these problems could be alleviated by using an alternative RNA preparation protocol. *Pf gch-1* derived transcript has previously been detected from yeast total RNA purified from cells by organic (phenol/chloroform) extraction following disruption of the cell wall by mechanical, rather than enzymatic, means (Lee, 1999). The sensitivity of the system may also have been affected by chemical modification of the RNA. Ethidium bromide, DEPC and oxidation products of both formamide and formaldehyde are all capable of effecting modifications to RNA nucleobases that can interfere with hybridisation

(Sambrook *et al.*, 1989; Garcia-Sanz, 1997), although exposure of the RNA to these chemicals was minimised wherever possible.

4.11.3 : Expression in a human cell line

The failure to detect any recombinant *Pf* GTCase derived protein expressed by human embryonic kidney 293T cells transiently transfected with pcDNA-gch and pcDNA-Cgch (4.10) can possibly be ascribed to a number of causes. It is likely that the A+T richness of *P.falciparum* DNA (1.6) contributed to the low level of expression observed, while the quality of the DNA preparations used may have reduced the efficiency of transfection. However, transfection of pcDNA-lacZ, the commercially supplied positive control construct encoding β-galactosidase resulted in expression of this protein at a level that was only marginally detectable (Figure 4.14). It is possible, therefore, that the efficiency of transfection was unusually low in this series of transfections for reasons other than DNA quality; a number of factors, including the pH of the transfection buffer used, can affect the efficiency of calcium phosphate transfection (Doyle, 1996). For this reason, it would seem injudicious to abandon investigations into expression of *Pf gch-1* in vertebrate cell lines. It is entirely possible that alternative host cell strains or transfection protocols might result in useful expression of *Pf* GTCase.

CHAPTER 5

FURTHER MOLECULAR STUDIES OF THE GTP CYCLOHYDROLASE I GENES FROM *PLASMODIUM*

5.1 : Intraerythrocytic stage-specific transcription of *Pf gch-1*

The intraerythrocytic cycle of *Plasmodium*, whereby a single parasite nucleus provided by an invasive merozoite divides by repeated rounds of mitosis to yield a variable number (usually 8-32) of daughter parasites (1.3.1), naturally requires considerable synthesis of biomolecules, including DNA and protein. There therefore exists a parallel requirement during this time for folate cofactors, which act as donors and acceptors of one-carbon units in the biosynthesis of dTMP, a precursor of DNA, and several amino acids, most notably methionine (1.8).

It has been demonstrated (Inselburg & Banyan, 1984) that DNA synthesis in *P.falciparum* does not occur uniformly throughout the intraerythrocytic cycle; initiation of this process occurs at around 30hr following invasion (corresponding to the late trophozoite stage of the cycle) and synthesis increases logarithmically throughout schizogony (no synthesis of DNA is observed in the early trophozoite or ring stage of the cycle). Given that parasite amino acid requirements are also likely to fluctuate through the life cycle, it is reasonable to postulate that rates of folate biosynthesis will show stage-specific variation.

Activity of GTCase (1.9), the enzyme catalysing the first step in folate biosynthesis, has been shown to be higher in mature trophozoites than in either ring or schizont forms of the simian malaria parasite *P.knowlesi* (Krungkrai *et al.*, 1985). No similar study has been conducted in any of the human malaria parasites, however. We therefore resolved to examine the intraerythrocytic stage-specific expression of GTCase in *P.falciparum* through detection of mRNA transcript from *Pf gch-1* using real-time quantitative RT-PCR (2.7.2). In this technique, changes in the fluorescence intensity of the RT-PCR product under ultraviolet illumination are monitored during the exponential phase of the amplification reaction. A 'threshold' cycle (C_t), where product first becomes detectable under standardised conditions, is determined, and this is related to a standard curve of C_t values obtained through PCR of a dilution series of stock genomic

DNA of known concentration using the same pair of primers. In this way, relative (but not absolute) mRNA concentrations, in this case between different life-cycle stages, can be extrapolated.

In order to make allowances for variability in starting RNA concentrations it was necessary to include an internal control gene in this experiment. In traditional comparative analyses of transcript levels by northern hybridisation it is standard practice to normalise total RNA levels in different samples by visually comparing the fluorescence intensity under UV of ethidium bromide-stained LSU and SSU rRNAs (for example; Horrocks & Newbold, 2000; Al-Khedery *et al.*, 1999). It was therefore decided to use the gene encoding asexual stage SSU rRNA in *P.falciparum* (McCutchan *et al.,* 1988) as the internal control (subsequent work by Goidin *et al.* (2001) has shown SSU rRNA to be superior to both glyceraldehyde-3-phosphate dehydrogenase and ß-actin genes as an internal standard for RT-PCR experiments in human tumour cells).

In the standard RT-PCR protocol (2.7.2), developed by Niroshini Nirmalan (UMIST), with the detection of mRNA species in mind, 7µl of 1st strand cDNA is used as template in a 100µl orthodox PCR reaction (2.7.1). In order to allow detection of the vastly more numerous SSU rRNA molecules within a similar time-frame it was necessary to modify the protocol by using lower concentrations of 1st strand cDNA.

1st strand cDNA was synthesised from *P.falciparum* strain HB3 (Honduran isolate) RNA using primer ssu3 (Appendix V) as described in 2.7.2. A dilution series of this 1st strand cDNA was then prepared, and 7µl of each dilution used as template in a 100µl PCR reaction (2.7.1) using an ssu1/ssu2 primer combination (Appendix V). Previous RT-PCR experiments with a variety of mRNA species had shown that the C_t usually fell between 10 and 24 PCR cycles, depending on the gene and life-cycle stage in question. It was therefore determined to sample the PCR reactions using the dilution series after 10 cycles (Figure 5.1).

A band of the expected size was apparent at the 10th cycle in PCRs using cDNA at dilutions down to 1/30 (not apparent in reproduction). It was therefore decided to use SSU rRNA-specific cDNA at a dilution of 1/100 in the stage-specific RT-PCR experiments in the expectation that this would allow assignation of a C_t by sampling from the 10th cycle of PCR onwards.

Cultures of *P.falciparum* strain HB3 in human RBCs were synchronised by two treatments with 5% sorbitol at 30hr intervals as described by Read & Hyde (1993).

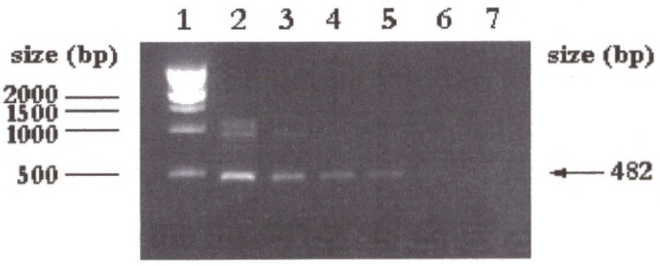


Figure 5.1 SSU rRNA specific RT-PCR using a 1st strand cDNA dilution series
All samples were taken after 10 PCR cycles; lane numbers: 1, DNA marker; 2, undiluted cDNA; 3, 1/2 dilution; 4, 1/5 dilution; 5, 1/10 dilution; 6, 1/30 dilution; 7, 1/100 dilution.

Total RNA was extracted from parasites pelleted from 10ml of culture at intervals of approximately 8hr, 28hr and 34 hr after RBC invasion, corresponding to ring, trophozoite and schizont stages respectively, and resuspended in 50µl RNase free water.

For each life-cycle stage, two separate 1st strand cDNA syntheses were carried out as described in 2.7.2; one using primer ssu3 in order to synthesise SSU rRNA-specific cDNA, and one using the oligo-p$(dT)_{15}$ primer (Roche) in order to synthesise cDNA from mRNA species in a non-specific manner. For each life-cycle stage three separate 100µl PCR reaction mixes were then prepared; one using 7µl of SSU-rRNA specific cDNA at a 1/100 dilution and the ssu1/2 primer combination, one using 7µl of mRNA-specific cDNA and the *Pf gch-1*-specific cyclo16/cyclo31 primer combination (Appendix V), and one using 7µl of mRNA-specific cDNA and the *Pf dhfs/fpgs*-specific dhfs17/dhfs20 primer combination (Appendix V). This last mix was intended to serve as a control on the quality of the RNA preparations; the dhfs17/dhfs20 priming sites are separated by three introns in the genomic DNA sequence – any PCR product amplified from contaminating genomic DNA would therefore be of an observably greater size than that produced by amplification from cDNA.

Each of the reaction mixes was then divided into 6 aliquots and subjected to PCR as described in 2.7.1. Aliquots were removed from the PCR machine at intervals of two cycles starting at either the 10th, 12th or 14th cycle depending on the nature of the cDNA used and life-cycle stage in question. Each set of aliquots was then subjected to agarose gel electrophoresis (2.6.1) in order to determine the relevant C_t (Figure 5.2, Table 5.1).

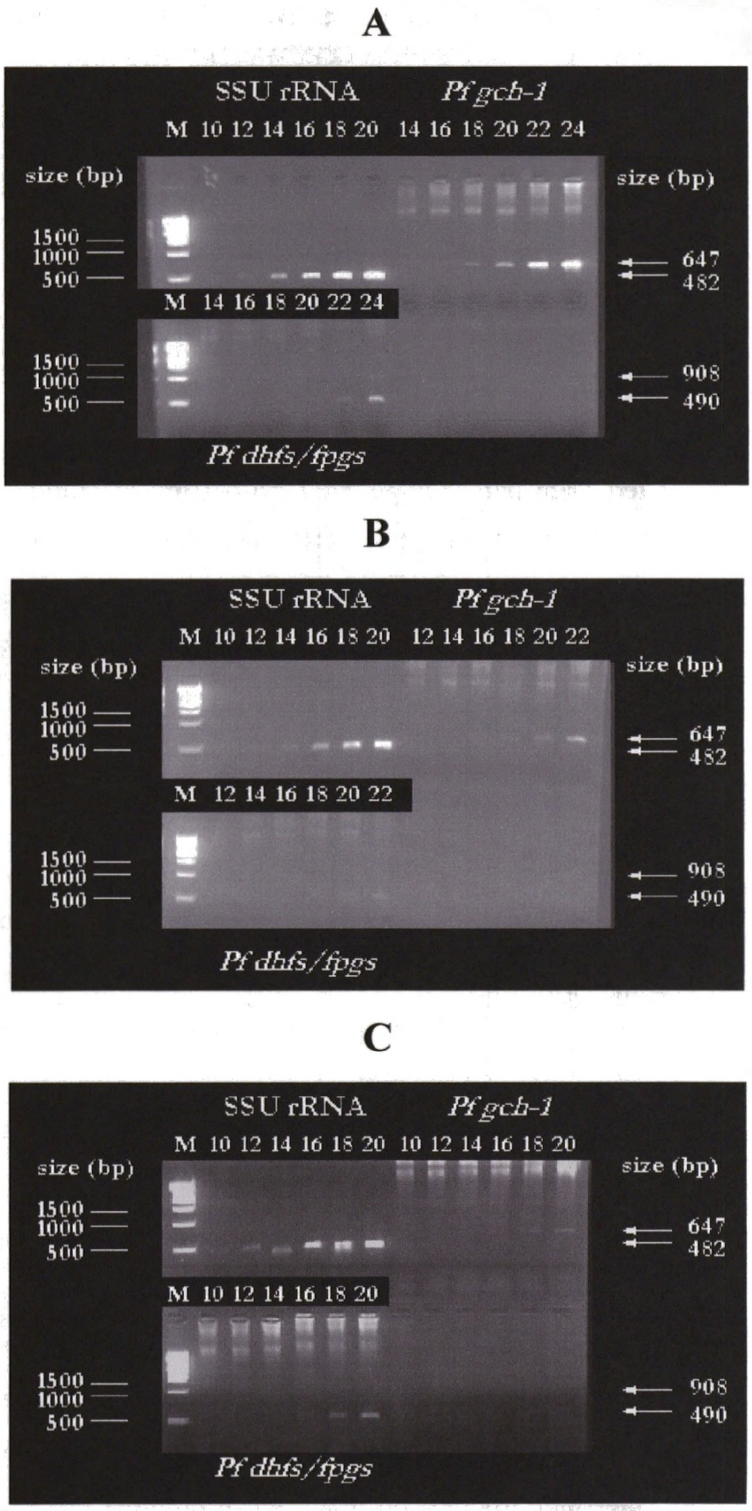


Figure 5.2 Real time RT-PCR of *Pf gch-1* from asexual cycle stage-specific parasite RNA A, ring; B, trophozoite; C, schizont. 6 separate aliquots of a PCR reaction using primers specific to asexual stage SSU rRNA, *Pf gch-1* or *Pf dhfs/fpgs* were taken between the 10th and 24th cycle as indicated. The sizes at the right hand side correspond to the expected products of the SSU rRNA specific reaction (482bp), the *Pf gch-1* specific reaction (647bp), the *Pf dhfs/fpgs* cDNA specific reaction (490bp) and the *Pf dhfs/fpgs* genomic DNA specific reaction (908bp). M=DNA markers.

Asexual cycle stage	C_t (SSU rRNA)	C_t (*Pf gch-1*)
Ring	12	18
Trophozoite	14	19
Schizont	11	19

Table 5.1 Threshold cycles for stage-specific RT-PCRs
An extremely faint band at cycle x was interpreted as a C_t value of $x+1$.

From all three RNA sources *Pf dhfs/fpgs* cDNA-specific product appeared before the larger *Pf dhfs/fpgs* genomic DNA-specific product. The RNA preparations were therefore judged to be of sufficiently high quality for the requirements of this experiment. The C_t values recorded were then compared against a set of genomic DNA standard curves prepared using data provided by Niroshini Nirmalan, UMIST (Figure 5.3).

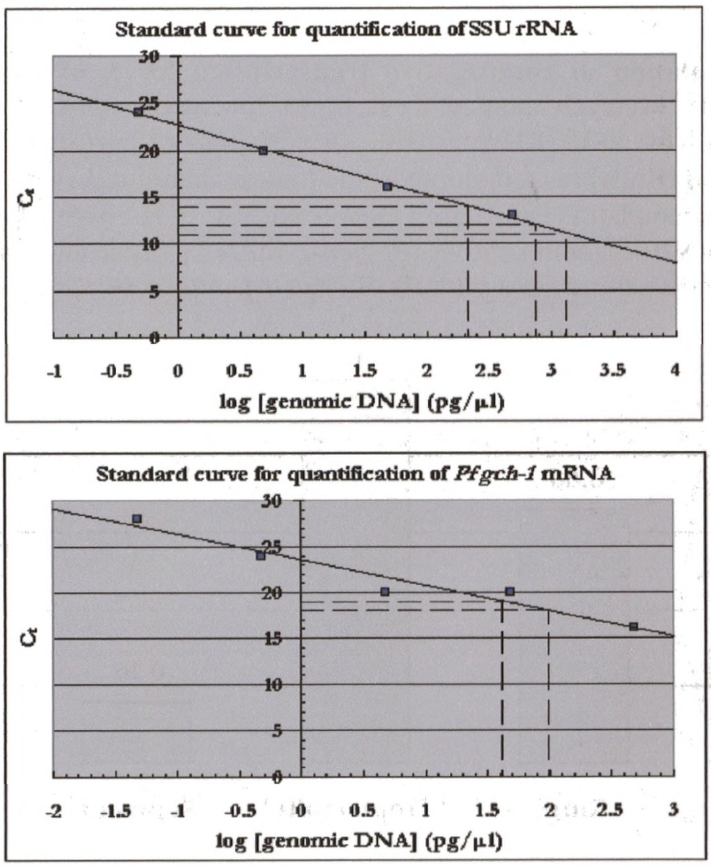


Figure 5.3 Standard curves for RNA quantification
Blue squares represent data points provided by Niroshini Nirmalan.[genomic DNA] values represent concentrations of *P.falciparum* genomic DNA in the final PCR reaction. Dashed lines indicate the extrapolations of template concentrations from the C_t values recorded in Table 5.1, although the values listed in Table 5.2 were derived mathematically.

The equivalent template concentrations for the SSU rRNA-specific PCR reactions, calculated using the equation of the SSU rRNA standard curve, were used to provide an index for the relative concentrations of total RNA in the preparations from the three life-cycle stages. The equivalent template concentrations for the *Pf gch-1*-specific PCR reactions (calculated in a similar manner) were then adjusted according to this index to give a measure of the relative transcription levels of this gene in ring, trophozoite and intraerythrocytic schizont stages of the parasite life-cycle (Table 5.2, Figure 5.4).

	a	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*
Ring	12	2.87	741	**3.46**	18	1.97	93	**2.27**	**0.66**
Trophozoite	14	2.33	214	**1.00**	19	1.61	41	**1.00**	**1.00**
Schizont	11	3.14	1380	**6.45**	19	1.61	41	**1.00**	**0.16**

Table 5.2 Calculation of comparative transcription levels of *Pf gch-1* between intraerythrocytic life-cycle stages Key: *a*, C_t for SSU rRNA; *b*, equivalent log [template] (pg/µl) for SSU rRNA; *c*, equivalent [template] (pg/µl) for SSU rRNA; *d*, comparative [total rRNA] (as a multiple of the lowest value); *e*, C_t for *Pf gch-1* mRNA; *f*, equivalent log [template] (pg/µl) for *Pf gch-1* mRNA; *g*, equivalent [template] (pg/µl) for *Pf gch-1* mRNA; *h*, comparative [*Pf gch-1* mRNA] (as a multiple of the lowest value); *i*, comparative transcription levels of *Pf gch-1* mRNA (*h/d*).

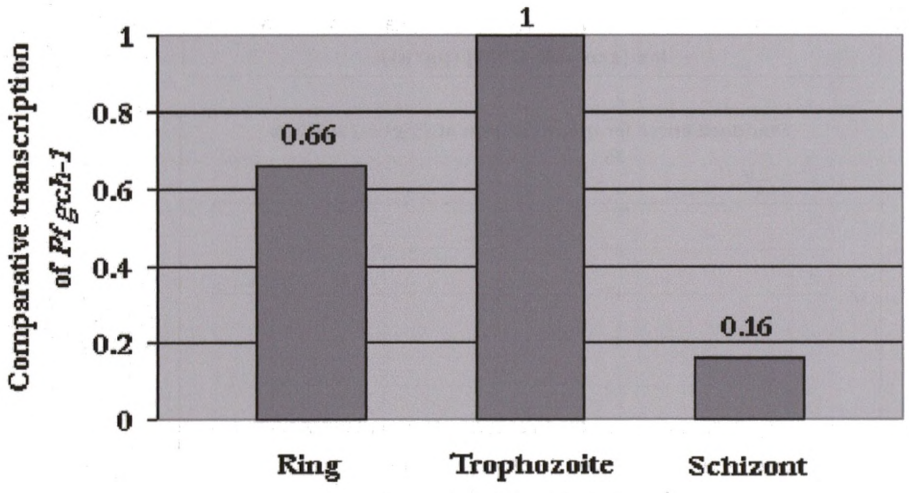


Figure 5.4 Comparative transcription levels of *Pf gch-1* between intraerythrocytic life-cycle stages

These results indicate that the transcription of *Pf gch-1* as a proportion of total RNA synthesis is highest in the mature trophozoite stage of the intraerythrocytic cycle,

and lowest in the schizont stage, although a difference of less than one order of magnitude between the two stages was noted in this case. This observation has subsequently been repeated by another experimenter (Nirmalan N., personal communication) and is not inconsistent with the observation that GTCase activity peaks in the mature trophozoite stage of the intraerythrocytic cycle of *P.knowlesi* (Krungkrai *et al.*, 1985).

5.2 : Identification and analysis of the genes encoding GTCase in other *Plasmodium* spp.

In addition to the *P.falciparum* genome sequencing project (1.6), co-ordinated efforts are also underway to sequence the genome of a number of other *Plasmodium* spp., namely *P.vivax*, the rodent parasites *P.chabaudi, P.berghei* and *P.yoelii*, and the simian parasite *P.knowlesi*. Preliminary data from these projects is available in the public domain (e.g. at http://www.plasmoDB.org). By analysing this data as it accumulated, we hoped to identify genes encoding GTCase in *Plasmodium* species other than *P.falciparum*. The *P.falciparum* GTCase shows a high level of conservation in its C-terminal portion compared with the enzymes from other organisms, both eukaryotic and prokaryotic, but possesses an unusually long N-terminal region (1.9.1). It has been postulated that this portion of the protein might play a role in the regulation of enzyme activity (Lee, 1999). If this were the case, then those motifs actively involved in such regulation might reasonably be expected to show conservation between different *Plasmodium* spp.. Analysis of the N-terminal portions of the GTCases of these other spp. would possibly be instructive therefore. In addition, although all malaria parasites possess A+T rich genomes, the degree of A+T richness varies between species (Weber, 1988); for example, *P.falciparum* possesses a particularly A+T rich genome (1.6). This variation is reflected in the codon usage of the different species. Of the species currently subject to genome sequencing projects, *P.falciparum* and *P.yoelii* show the most marked A+T bias, while the bias is least pronounced in *P.vivax* and *P.knowlesi* (Table 5.3). Given that many of the problems encountered in expression of *Pf* GTCase in heterologous systems (Chapters 3 & 4) can either be tentatively or explicitly linked to the extremely biased codon usage in this species (for example see 4.1), the identification of a less A+T rich parasite GTCase gene might prove useful in providing a model for functional analysis of this enzyme in *Plasmodium*.

Plasmodium species	A+T content in coding areas (%)
P.falciparum	74.02
P.yoelii	72.13
P.berghei	69.30
P.chabaudi	68.19
P.knowlesi	60.11
P.vivax	59.30

Table 5.3 A+T content in coding regions of the genomes of *Plasmodium* spp.
Data extracted from the Codon Usage Database (http://www.kazusa.or.jp/codon). Only those species currently the subject of genome sequencing projects are listed.

Probing of the *Plasmodium* genome databases with the C-terminal amino acid sequence of *Pf* GTCase using a TBLASTN (protein vs. DNA) search algorithm initially yielded an ORF from *P.yoelii* encoding a polypeptide with significant homology to *Pf* GTCase (Figure 5.5), close to the end of the sequence contig on which it was found.

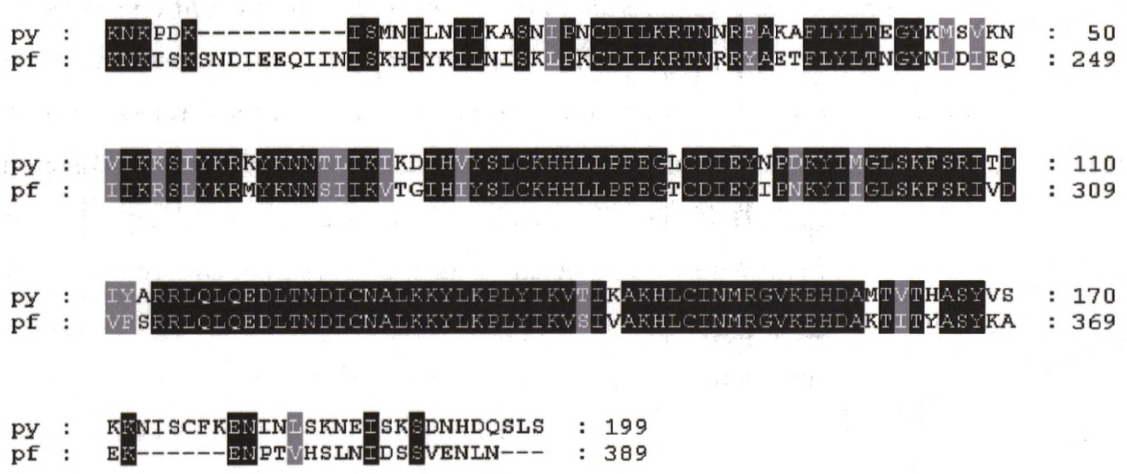


Figure 5.5 Alignment of the putative partial *P.yoelii* GTCase with *Pf* GTCase
Residues that are conserved or semi–conserved between the two species are shaded in black and dark grey respectively. py, *P.yoelii*; pf, *P. falciparum*.

Due to the comparatively short length of this polypeptide (199aa, compared with 389aa for *Pf* GTCase) it was suspected that this represented only part of the *P.yoelii* GTCase. The apparent absence of non-coding type sequence and intron splicing signals (as defined by e.g. Padgett *et al.*, 1986) from the sequence around the 5' end of the ORF, coupled with the presence of an overlapping, uninterrupted ORF in a different reading frame, extending as far as the extreme 5' terminus of the sequence contig

(Figure 5.6) suggested that a frameshift type error had been generated by the automated sequencing.

```
          .         .         .         .         .         .
GGTAAAAAAATAATAAAAGGGGAAAGAATAGACAAAAAGAATAAGGAGGATAATAATTTT      60
 G  K  K  I  I  K  G  E  R  I  D  K  K  N  K  E  D  N  N  F

          .         .         .         .         .         .
ATTATAGACGATTTAGAAAAACAAACCAGATAAAATAAGTATGAATATATTAAATATTTT      120
 I  I  D  D  L  E  K  Q  T  R  <
                <  K  N  K  P  D  K  I  S  M  N  I  L  N  I  L

          .         .         .         .         .         .
AAAAGCATCAAATATTCCAAATTGTGATATATTAAAAAGAACAAATAACAGATTTGCTAA      180
  K  A  S  N  I  P  N  C  D  I  L  K  R  T  N  N  R  F  A  K
```

Figure 5.6 Part of a *P.yoelii* sequence contig including a putative GTCase encoding ORF Nucleotide 1 represents the extreme 5' terminus of the contig as determined when this ORF was originally isolated. The polypeptides encoded by two overlapping ORFs are shown.

In order to test this hypothesis, we resolved to sequence this tract of the *P.yoelii* genome independently. A PCR (2.7.1) was carried out using *P.yoelii* genomic DNA as template and a PyGTP1/PyGTP2 primer combination (Appendix V), designed to produce an amplification product spanning the region of interest. The PCR product was then directly sequenced (2.8) using each of the primers used in the amplification (Figure 5.7).

Figure 5.7 Partial sequence chromatogram of a PyGTP1/PyGTP2 PCR product That part of the sequence affected by the frameshift error in the genome database is indicated with a red asterisk. Additional differences with respect to the database sequence are indicated by black asterisks.

The sequence revealed that, as anticipated, an apparent frameshift error was present in the database sequence (the duplication of a C residue at nt 86 as shown in Figure 5.6). Taking this error into account the two overlapping ORFs shown in Figure

5.6 are effectively stitched together, with the Thr residue in the overlap region of the upstream ORF adjoining the Asp residue in the same portion of the downstream ORF. The reading frame of interest was thus seen to be still open at the 5' extremity of the sequence contig; the polypeptide thus encoded was therefore incomplete. In addition to the frameshift error, two other differences between the database sequence and the sequence of the PCR product were recorded (these are marked by black asterisks in Figure 5.7). The first of these, a change of a G to a C at nt 39 produced an inferred change in the in frame polypeptide of a Lys to an Asn residue. The second change, of a C to a T at nt 69, was equivalent to a silent mutation. Three subspecies of *P.yoelii*, namely *P.y.yoelii*, *P.y.killicki* and *P.y.nigeriensis*, have been isolated from the blood of shiny thicket rats, *Thamnomys rutilans,* from the Central African Republic and Cameroon, Congo-Brazzaville and Nigeria respectively (Killick-Kendrick & Peters, 1978; Landau & Chabaud, 1994). The genome project standard strain is 17XNL, a strain of *P.y.yoelii*, however the genomic DNA used in the PCR had been extracted from the lethal strain of *P.y.nigeriensis*; it is likely that these additional differences between the two sequences were due to the different source subspecies.

Subsequent to this analysis, the database sequence contig of interest was revised, eliminating the frameshift error, and also extended a considerable distance in the 5' direction. The ORF encoding the putative GTCase is shown to extend backwards for approximately 100 additional amino acids before being interrupted by a cluster of stop codons. Given the absence of intron splicing signals in the sequence tract downstream of these stop codons (Figure 5.8), it seems reasonable to postulate that the complete putative GTCase is encoded as a single exon within this ORF. Two possible start (Met encoding) codons are present in the vicinity of the stop codons (Figure 5.8).

```
TATTTACCATTTTAATGAAAATAAAAT[*]GAAACCTATATCGAAAAAAAGCAACTCGAT
 Y  L  P  F  <  <  K  <  N  M  E  T  Y  I  E  K  K  Q  L  D

GAAGGTTGC[*]TTAAGCAAGAAAAAAAATGAAGCCAAAAAATGTGAAGGCCAAAAAAT
 E  G  C  M  L  S  K  K  K  N  E  A  K  K  C  E  G  Q  K  N
```

Figure 5.8 Part of a *P.yoelii* sequence contig encoding the 5' end of a putative GTCase encoding ORF Putative start codons are highlighted.

The work of Kozak (1984) suggests that for efficient translation in eukaryotes a purine, usually A, is necessary at position -3 relative to the start codon, while a G at position +4 is essential in the absence of this signal. Of the two candidate start codons in the putative *P.yoelii* GTCase gene, the upstream one possesses both these signals, while the downstream one possesses neither. It is hence likely that the upstream Met codon of the two represents the start codon of this polypeptide. The full presumed sequence of this putative *P.yoelii* GTCase and the gene that encodes it are shown in Figure 5.9.

```
                .         .         .         .         .         .
ATGGAAACCTATATCGAAAAAAAGCAACTCGATGAAGGTTGCATGTTAAGCAAGAAAAAA      60
 M  E  T  Y  I  E  K  K  Q  L  D  E  G  C  M  L  S  K  K  K       20

                .         .         .         .         .         .
AATGAAGCCAAAAAATGTGAAGGCCAAAAAAATGAAGGCAAAAAATATGAAGGCCAAAAA      120
 N  E  A  K  K  C  E  G  Q  K  N  E  G  K  K  Y  E  G  Q  K       40

                .         .         .         .         .         .
AATGAAGCCAGCAAAAATGAAGCCAGCAAAAATGACGCCAGCAAAAATGAAGCCAGCAAA      180
 N  E  A  S  K  N  E  A  S  K  N  D  A  S  K  N  E  A  S  K       60

                .         .         .         .         .         .
AATGACGCCAGCAAAAATGACATTTCAAGCGATACAACACAATTGATAAGTATAGATAAT      240
 N  D  A  S  K  N  D  I  S  S  D  T  T  Q  L  I  S  I  D  N       80

                .         .         .         .         .         .
CCTAAATCGATAAATAATAAAAATAATGGTGAAGGTAAAAAAATAATAAAAGGGGAAAGA      300
 P  K  S  I  N  N  K  N  N  G  E  G  K  K  I  I  K  G  E  R      100

                .         .         .         .         .         .
ATAGACAAAAAGAATAAGGAGGATAATAATTTTATTATAGACGATTTAGAAAAACAAACA      360
 I  D  K  K  N  K  E  D  N  N  F  I  I  D  D  L  E  K  Q  T      120

                .         .         .         .         .         .
GATAAAATAAGTATGAATATATTAAATATTTTAAAAGCATCAAATATTCCAAATTGTGAT      420
 D  K  I  S  M  N  I  L  N  I  L  K  A  S  N  I  P  N  C  D      140

                .         .         .         .         .         .
ATATTAAAAAGAACAAATAACAGATTTGCTAAAGCATTTTTATATTTAACTGAAGGATAT      480
 I  L  K  R  T  N  N  R  F  A  K  A  F  L  Y  L  T  E  G  Y      160

                .         .         .         .         .         .
AAAATGAGTGTAAAAAATGTAATAAAAAAATCTATATATAAAAGAAAATATAAAAATAAT      540
 K  M  S  V  K  N  V  I  K  K  S  I  Y  K  R  K  Y  K  N  N      180

                .         .         .         .         .         .
ACACTAATTAAAATAAAAGATATACATGTTTATTCTTTATGTAAACATCATTTATTGCCA      600
 T  L  I  K  I  K  D  I  H  V  Y  S  L  C  K  H  H  L  L  P      200

                .         .         .         .         .         .
TTTGAAGGATTGTGTGATATTGAATATAATCCAGATAAATATATAATGGGTTTATCCAAA      660
 F  E  G  L  C  D  I  E  Y  N  P  D  K  Y  I  M  G  L  S  K      220

                .         .         .         .         .         .
TTTTCAAGAATTACTGATATATATGCAAGACGTTTACAATTGCAAGAAGATCTTACTAAT      720
 F  S  R  I  T  D  I  Y  A  R  R  L  Q  L  Q  E  D  L  T  N      240
```

```
          .         .         .         .         .         .
GATATTTGCAATGCTTTAAAAAAATATTTAAAACCCTTATATATAAAAGTTACAATTAAA      780
 D  I  C  N  A  L  K  K  Y  L  K  P  L  Y  I  K  V  T  I  K       260

          .         .         .         .         .         .
GCTAAACATTTATGCATTAATATGAGAGGGGTAAAAGAACATGATGCTATGACAGTTACA      840
 A  K  H  L  C  I  N  M  R  G  V  K  E  H  D  A  M  T  V  T       280

          .         .         .         .         .         .
CATGCATCATATGTATCAAAAAAAAATATTAGTTGTTTTAAAGAAAATATTAATTTATCG      900
 H  A  S  Y  V  S  K  K  N  I  S  C  F  K  E  N  I  N  L  S       300

          .         .         .         .
AAAAATGAAATATCAAAATCAGATAACCATGATCAATCTTTATCATAA          948
 K  N  E  I  S  K  S  D  N  H  D  Q  S  L  S  <           315
```

Figure 5.9 Nucleotide and deduced amino acid sequence of the putative *P.yoelii* GTCase gene Start and stop codons are in bold and underlined. This sequence is that found in the *P.yoelii* genome sequence database.

In addition to this sequence, further probing of the *Plasmodium* genome databases with the C-terminal amino acid sequence of *Pf* GTCase using a TBLASTN search algorithm has also yielded an ORF seemingly encoding a partial GTCase from a *P.knowlesi* sequence contig. The ORF is truncated in the 5' direction by the end of the contig, but apparently includes the complete 3' terminus of the gene (Figure 5.10).

An alignment of the three putative GTCases is shown in Figure 5.11. Unsurprisingly, given the relatively small evolutionary distance between the three species and the high degree of similarity between the C-terminal 'core' regions (corresponding to aa 262-367 in *P.falciparum*) of GTCases in widely different taxa (1.9), the three polypeptides show a very high degree of identity and similarity in this region (Table 5.4, cf. Table 1.1).

	Pf	**Py**	**Pk**
Pf	n/a	*82*	*81*
Py	93	n/a	*77*
Pk	92	87	n/a

Table 5.4 Identities and similarities of the 'core' region of the three putative malaria parasite GTCases Values are percentages. Identity figures are italicised. Pf, *P.falciparum*; Py, *P.yoelii*; Pk, *P.knowlesi*.

```
                                                          .
                                                          G        1

          .         .         .         .         .         .
GGTAGAAGATTTTCGGATACCTTCCTGTACTTAACAAAAGGGTACCATATGAGCGTAGAA      61
 G  R  R  F  S  D  T  F  L  Y  L  T  K  G  Y  H  M  S  V  E       20

          .         .         .         .         .         .
AAAGTTATTAAAAAATCCTTGTACAAAAGGAATTACAAAAATAACTCTGTCATAAAAATA      121
 K  V  I  K  K  S  L  Y  K  R  N  Y  K  N  N  S  V  I  K  I       40

          .         .         .         .         .         .
AGTGGCATTCATATTTACTCTTTGTGCAAGCATCATTTGTTGCCATTCGAAGGAGAGTGC      181
 S  G  I  H  I  Y  S  L  C  K  H  H  L  L  P  F  E  G  E  C       60

          .         .         .         .         .         .
ACCATCGAGTATATCCCGAATAAATATATTATGGGTCTGTCCAAATTTTCCAGGGTTATA      241
 T  I  E  Y  I  P  N  K  Y  I  M  G  L  S  K  F  S  R  V  I       80

          .         .         .         .         .         .
GATATTTTTGCGAGAAGGTTACAACTACAGGAGGATTTAACCAATGACATTTGTAATGCT      301
 D  I  F  A  R  R  L  Q  L  Q  E  D  L  T  N  D  I  C  N  A      100

          .         .         .         .         .         .
CTGGGGAAATATTTGAAGCCGAAATATCTACATGTAAATCTTGTGGCCAGACATTTATGC      361
 L  G  K  Y  L  K  P  K  Y  L  H  V  N  L  V  A  R  H  L  C      120

          .         .         .         .         .         .
ATTAATATGCGCGGAGTGAAGGAACACGATGCGACTACTATCACGAATGCATACTATGAG      421
 I  N  M  R  G  V  K  E  H  D  A  T  T  I  T  N  A  Y  Y  E      140

          .         .         .         .         .         .
GTAAAAAATGATACTTGTGTAAATCACGCCGGGAATGGATTTACGTGTCCCTCCTTCCAT      481
 V  K  N  D  T  C  V  N  H  A  G  N  G  F  T  C  P  S  F  H      160

          .         .         .         .
AATGGGAATAATCCATCTCGAGAAGATATAGCGCCCGTGTGA                        523
 N  G  N  N  P  S  R  E  D  I  A  P  V  <                         173
```

Figure 5.10 Nucleotide and deduced amino acid sequence of the putative partial *P.knowlesi* GTCase gene The putative stop codon is in bold and underlined.

Outside this region markedly less similarity is evident, although the regions of the polypeptides directly upstream of the 'core' (for example corresponding to aa 198-261 in *P.falciparum*) show a degree of inter-species similarity similar to that noted between the corresponding regions of GTCases in vertebrate and invertebrate animal species (data not shown). Further upstream however, the sequences of the putative GTCases in *P.falciparum* and *P.yoelii* diverge more dramatically (this part of the polypeptide is not present in the incomplete *P.knowlesi* sequence). Although the putative *P.yoelii* GTCase, at 315 aa, is longer than any non-plasmodial GTCase sequenced, it is still significantly shorter than *Pf* GTCase; its N-terminal

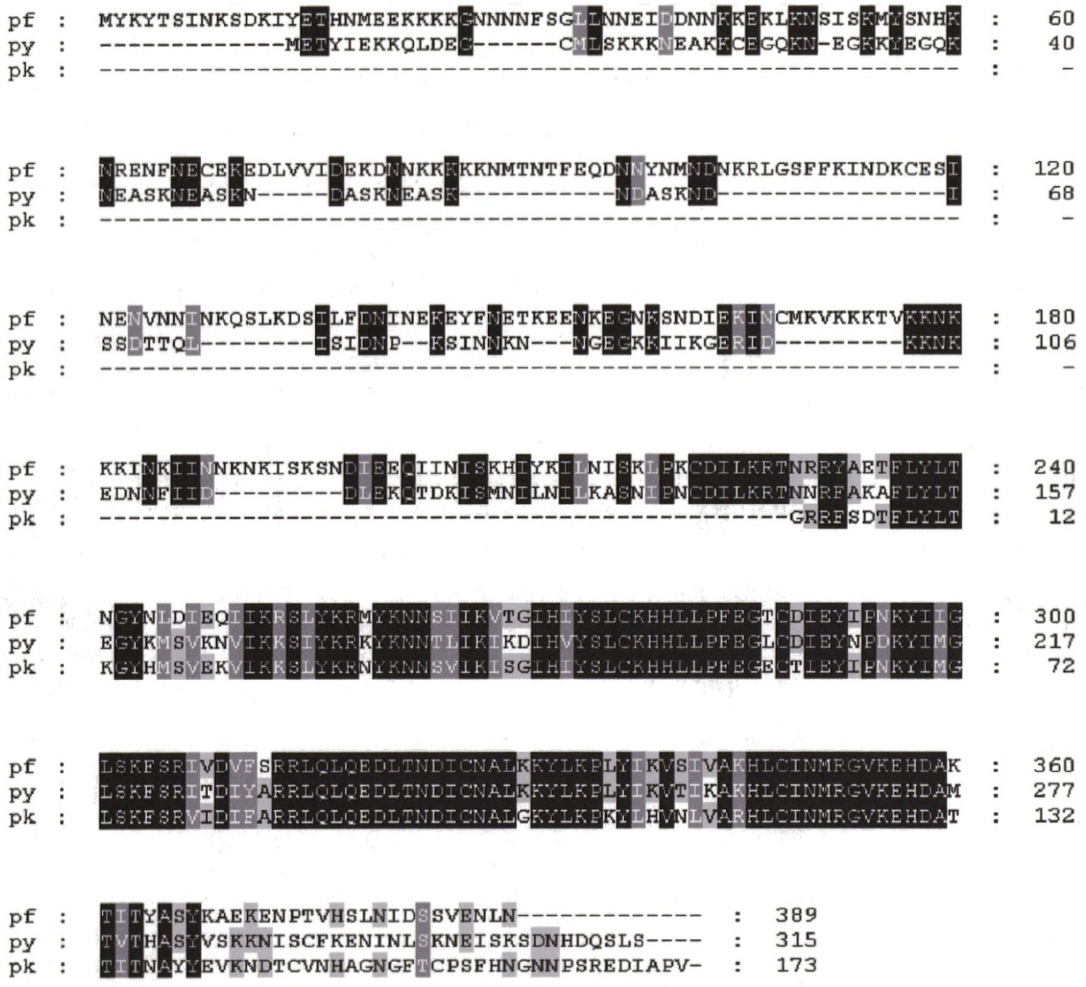


Figure 5.11 Alignment of the putative GTCases from *P.yoelii* and *P.knowlesi* with *Pf* GTCase In that portion of the alignment from which *P.knowlesi* sequence is absent, residues that are conserved or semi-conserved between the other two species are shaded in black and dark grey respectively. In the portion of the alignment in which sequence is available for all three parasite species, residues that are conserved between all species, semi-conserved between all species and conserved or semi-conserved between only two of the species are shaded in black, dark grey and light grey respectively. pf, *P. falciparum*; py, *P.yoelii*; pk, *P.knowlesi*.

extension in fact contains 83 aa fewer. Superficially, the alignment of the two sequences (Figure 5.11) appears to suggest that the extra length of the *P.falciparum* sequence is due to the presence of several oligopeptide inserts relative to the *P.yoelii* sequence, although it is possible that this effect is an artefact produced by the alignment algorithm; many of the matches in this area are between residues that are over-represented, for example lysine, asparagine and glutamate.

Downstream of the 'core' region, the divergence between the three polypeptides is even more striking. In non-plasmodial species this C-terminal extension is moderately conserved, both in terms of sequence, with, for example, these portions of the human and *E.coli* GTCases sharing 38% identity and 57% similarity, and also in length; extensions beyond the 'core' range from 19 aa in the enzyme from vertebrates and some bacterial species to 25aa in *Campylobacter jejuni*, with the majority of extensions 21 aa or fewer (see Appendix III). In *Pf* GTCase, the C-terminal extension is not of unusual length, being 22 aa; the equivalent regions of the putative *P.yoelii* and *P.knowlesi* GTCases are notably long, however, at 31 and 34 aa respectively. In addition, none of the three plasmodial sequences shows any real homology to the corresponding areas of non-plasmodial GTCases, and, even more remarkably considering the not inconsequential level of conservation of the C-terminus of GTCase in distantly related organisms, these three sequences exhibit little, if any, homology to each other (Table 5.5).

	Pf	**Py**	**Pk**	**Hs**	**Sc**	**Ec**
Pf	n/a	*9*	*8*	*9*	*0*	*4*
Py	19	n/a	*5*	*3*	*6*	*3*
Pk	32	26	n/a	*5*	*0*	*0*
Hs	22	9	8	n/a	*54*	*38*
Sc	9	12	0	59	n/a	*31*
Ec	22	16	8	57	54	n/a

Table 5.5 Identities and similarities of the C-terminal extension region of the three putative malaria parasite GTCases and several non-plasmodial GTCases All sequence downstream of the 'core' region (e.g. aa 368-389 in *P.falciparum*) was considered. Values are percentages. Identity figures are italicised. Pf, *P.falciparum*; Py, *P.yoelii*; Pk, *P.knowlesi*. Hs, *H.sapiens*; Sc, *S.cerevisiae*; Ec, *E.coli*.

Following the unexpected discovery of point mutation differences between the *P.yoelii yoelii* 17XNL genome database and *P.yoelii nigeriensis* lethal strain GTCase sequences described above, it was decided to give further consideration to intra-species comparisons of GTCase. Previous work (Lee C.S., 1999) has shown that no differences in the coding sequences of *Pf gch-1* exist between two separate Thai isolates (K1 and Tak9/96) of *P.falciparum*, and a comparison of this sequence with that in the genome database (derived from the African 3D7 isolate) has also revealed 100% sequence

identity (results not shown). This is consistent, for example, with the observation that the coding sequence of the gene encoding the bifunctional folate biosynthesis/utilisation enzyme DHFS/FPGS is identical in ten different strains of *P.falciparum*, isolated from localities in Africa, Southeast Asia and the Americas (Salcedo *et al.*, 2001).

In order to further investigate the possibility of differences between the GTCase genes of *P.y.yoelii* and *P.y.nigeriensis* a set of primers was designed with the intention of amplifying and sequencing the entire GTCase gene from *P.yoelii* genomic DNA. Five new oligonucleotides (PyGTP3-7; Appendix V) were produced, including a pair designed to bind to the 5' UTR (PyGTP5 and PyGTP6, the latter binding closer to the putative start codon), and one designed to bind to the 3' UTR (PyGTP4).

In order to circumvent the possibility of errors in the database, it was decided to sequence the GTCase gene from *P.y.yoelii* independently. The entire gene was amplified by PCR (2.7.1) from *P.y.nigeriensis* genomic DNA and DNA extracted from lysates of mouse RBCs infected with *P.y.yoelii* 17XL (a lethal variant derived from the same isolate as the genome database strain) using the PyGTP6/PyGTP4 primer combination (the PyGTP5/PyGTP4 combination produced no amplification product). The PCR products were then sequenced directly with the PyGTP primers as appropriate (2.8) to generate full length *P.y.yoelii* and *P.y.nigeriensis* GTCase gene sequences (results not shown).

The sequence derived from *P.y.yoelii* genomic DNA was found to be identical to the database sequence as anticipated; the coding sequence from *P.y.nigeriensis* was found to be considerably shorter however, 891bp as opposed to 948bp. Alignment of the two sequences suggested that this length polymorphism was best explained by the presence of two separate sequence inserts in the *P.y.yoelii* sequence with respect to the *P.y.nigeriensis* sequence (nt 42-57 and nt 72-117 as shown in Figure 5.9), and a single trinucleotide (GTT) insertion in the *P.y.nigeriensis* sequence with respect to the *P.y.yoelii* sequence (between nt 264 and 265 as shown in Figure 5.9). In addition to these length polymorphisms a number of additional point mutation polymorphisms were found (predominantly inferring changes at the amino acid level also). Details of all codons showing intra-species differences in the putative *P.yoelii* GTCase gene are given in Table 5.6. An alignment of the polypeptides encoded by the two gene sequences is shown in Figure 5.12.

Just three silent mutation polymorphisms and no missense mutations are found in that portion of the gene corresponding to the 'core' region (aa 179-284 in

Codon affected & aa change conferred	*P.y.yoelii* nt sequence	*P.y.nigeriensis* nt sequence
Thr3→Asn	ACC	A[redacted]C
Gly13→Ala	GGT	G[redacted]
Cys14→Ser	TGC	[redacted]GC
Lys24→Arg	AAA	A[redacted]A
Ser44→Arg	AGC	AG[redacted]
Ser49→Arg	AGC	AG[redacted]
Asp52→Glu	GAC	GA[redacted]
Ser59→Arg	AGC	AG[redacted]
Asp62→Lys	GAC	[redacted]A[redacted]
Asn89→Asp	AAT	[redacted]AT
Gly90→Asp	GGT	G[redacted]T
Gly98→Arg	GGG	[redacted]GG
Lys104→Asn	AAG	AA[redacted]
Asp114; SILENT	GAC	GA[redacted]
Ser219; SILENT	TCC	TC[redacted]
Pro252; SILENT	CCC	CC[redacted]
Gly270; SILENT	GGG	GG[redacted]
Ile290→Val	ATT	[redacted]TT
Asn309→Ser	AAC	A[redacted]C

Table 5.6 Point mutation polymorphisms between the *P.y.yoelii* and *P.y.nigeriensis* GTCase gene sequences Amino acid numbering corresponds to that shown in Figure 5.9. Those nucleotides affected by polymorphisms are highlighted in the third column.

P.y.yoelii), while no polymorphisms are apparent in that portion corresponding to that latter part of the N-terminal extension described above as showing significant conservation between different plasmodial species (aa 115-178 in *P.y.yoelii*). Outside these regions however, a surprising degree of divergence is noted. In the rest of the N-terminal extension (aa 1-114 in *P.y.yoelii*), in addition to the three length polymorphisms described above, a total of 18 point mutation polymorphisms in 16 codons were found. In the case of the 14 singly mutated codons, all but one mutation conferred a change at the polypeptide level. Of the two codons with two mutations each, in one case (Gly13) the change at the third position would not be sufficient to effect a change in amino acid sequence whether it preceded or followed the change at the second position, and can therefore be regarded as silent; in the other case (Asp62) either change produces a change at the polypeptide level.

In the C-terminal extension region (aa 285-315 in *P.y.yoelii*) two point mutations in two codons are found, each effecting a change in amino acid.

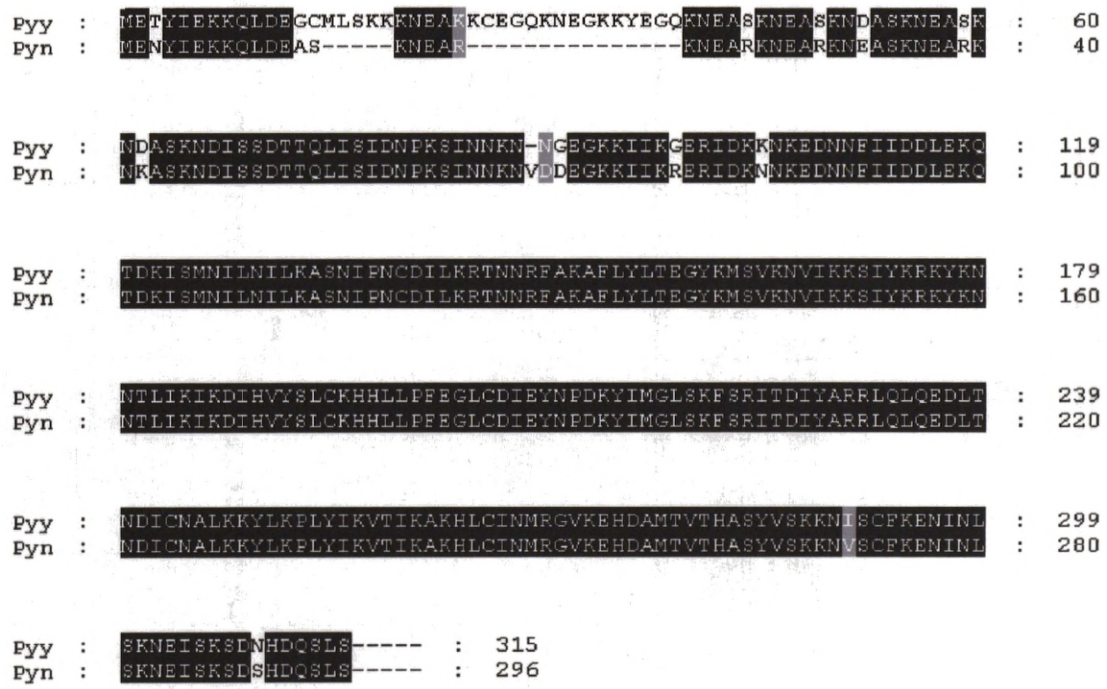


Figure 5.12 Alignment of the putative *P.y.yoelii* and *P.y.nigeriensis* GTCases
Residues that are conserved or semi–conserved between the two species are shaded in black and dark grey respectively. Pyy, *P.y.yoelii*; Pyn, *P.y.nigeriensis*.

5.3 : Discussion

5.3.1 : Intraerythrocytic stage-specific mRNA expression of *Pf gch-1*

The results yielded by this experiment suggest that, within the intraerythrocytic cycle of *P.falciparum*, expression of *Pf gch-1* is highest in the mature trophozoite stage of the cycle and lowest in the schizont stage (Figure 5.4). This is broadly consistent with previous work (Krungkrai *et al.*, 1985) showing that in the simian parasite *P.knowlesi*, GTCase activity was higher in the trophozoite stage by factors of 2.6 and 2 times compared with activity in the ring and schizont stages respectively. The apparent disparity between these two sets of data, whereby, when ring and schizonts are compared, expression is higher in the former by a factor of approximately 4 times, whereas activity is higher in the latter by a factor of 1.3 times, can reasonably be explained by the delays between transcription of the GTCase genes, translation of the

then be used to calculate the equivalent template concentration, provided efforts were made in order to ensure that such measurements fell in the linear part of the standard curve. This strategy would effectively render the troublesome concept of the C_t redundant. Finally, in the case of the problem of genomic DNA contamination, the treatment of each RNA preparation with RNase free DNase before RT-PCR has proved effective in completely eliminating amplification from genomic DNA in RT-PCR (Nirmalan N., personal communication).

5.3.2 : Comparison of the genes encoding GTCase in *Plasmodium* spp.

The identification of genes encoding apparent GTCases in *P.yoelii yoelii*, *P.y.nigeriensis* and *P.knowlesi* has revealed a surprisingly low level of conservation of areas of the protein outside the C-terminal 'core' region (1.9). Although moderate inter-species homology is apparent in the tract of the polypeptide immediately upstream of the 'core', the degree of conservation lessens as the N-terminus is approached (unfortunately no data for this part of the sequence in *P.knowlesi* is available as yet). This is perhaps to be expected as the N-terminal extensions of GTCases show little similarity between species; it does however, raise the question of exactly what the function of this part of the enzyme is in *Plasmodium* spp.. It has been speculated that this part of the *P.falciparum* GTCase contains regulatory motifs such as phosphorylation and metal binding sites (Lee, 1999). It is also possible that the enzyme undergoes some sort of subcellular localisation post translation, in which case the N-terminal extension might be expected to contain targeting information. Some degree of conservation of this area might be expected between different plasmodial species if this were the case however. The *P.yoelii* sequences contain different numbers of an apparent pentapeptide tandem repeat (10 in *P.y.yoelii*, 7 in *P.y.nigeriensis*); this repeat has a consensus sequence of KNEAS (or KNEAR in *P.y.nigeriensis*). No similar repeats are found in the *P.falciparum* sequence. Such repetitive sequences are commonly found in parasite antigens (e.g. Schofield, 1991), however their possible relevance to enzyme function in this case is unclear. A comparison of the N-terminal sequences from the two subspecies of *P.yoelii* reveals an unusual pattern of polymorphisms. In addition to a number of insertion/deletion mutations, a total of 16 point mutations in 14 codons were found (Table 5.5) of which just two were silent. In an unrestrained system, missense mutations are approximately three times more likely to occur than silent mutations; in this part of the *P.yoelii* GTCase gene however, they outnumber silent mutations by a

mRNA, and degeneration of the protein, however, experimental error is likely also to play a role; although these results are encouraging, the conduct of the experiment was beset by a number of minor problems. Efforts were made in order to standardise electrophoresis conditions to as great a degree as possible, however variation, for example in background fluorescence, was still apparent between different gels, raising questions as to the validity of drawing comparisons between them. In addition, the special measures taken to accommodate the high proportion of SSU rRNA molecules in the parasite total RNA proved barely sufficient. SSU rRNA-specific product was apparent at cycles that fell outside the data set used to construct the standard curve. It would therefore be necessary either to add additional data points to the curve or to use cDNA at greater dilutions in the SSU rRNA-specific reactions in the future.

The experimental strategy used also possesses a number of inherent limitations however. A certain amount of subjectivity is involved in the determination of C_t values. This problem is amplified by the logarithmic nature of the standard curves used to calculate template concentrations; a difference in C_t of a single cycle corresponds to a significant change in inferred template concentration. Due to the absence of introns from the SSU rRNA and *Pf gch-1* it is also difficult to discount the effects of genomic DNA contamination on the results obtained. The method used to guard against this effect (the inclusion of the *Pf dhfs/fpgs* gene as a control) is useful only as a crude indicator of genomic DNA contamination, and although cDNA specific product was in all cases observed before genomic DNA specific product, the comparative sizes of the two (the cDNA specific product is naturally smaller) means that absolute comparisons of the relative molar concentrations of cDNA and genomic DNA templates cannot be drawn. None of these problems is insoluble however. The LightCycler system (Roche), allows simultaneous amplification and highly sensitive fluorescent detection of PCR products (Wittwer *et al.*, 1997) and is used frequently as a tool for quantitative RT-PCR. In the absence of such equipment however, use of an electronic densitometer to measure fluorescence intensity would eliminate the need for subjective interpretation of the gels. The deleterious effects of the combination of a digital method of monitoring product formation with a logarithmic relationship between product formation and template concentration could be lessened if, for example, standard curves were produced by measuring the fluorescence intensities of a set of reactions using a range of different template concentrations, but, crucially, sampled at the same cycle; a measurement of the fluorescence intensity of the RT-PCR reaction under scrutiny at the same cycle could

factor of 7 times. This pattern might be anticipated if polymorphisms were actively selected for – it would perhaps be intemperate to draw such a conclusion in this case however.

The plasmodial GTCases also exhibit a marked and unusual degree of divergence in their C-terminal extensions, in terms of both length and homology (Table 5.5). In *E.coli* and human GTCases (Nar *et al.*, 1995a, Auerbach *et al.*, 2000), the largely α-helical C-terminal extension is located at the core of the GTCase homodecamer, with the absolute C-terminus found close to the interface of the two pentameric tori (see Figure 5.13 for this part of the molecule in *E.coli*).

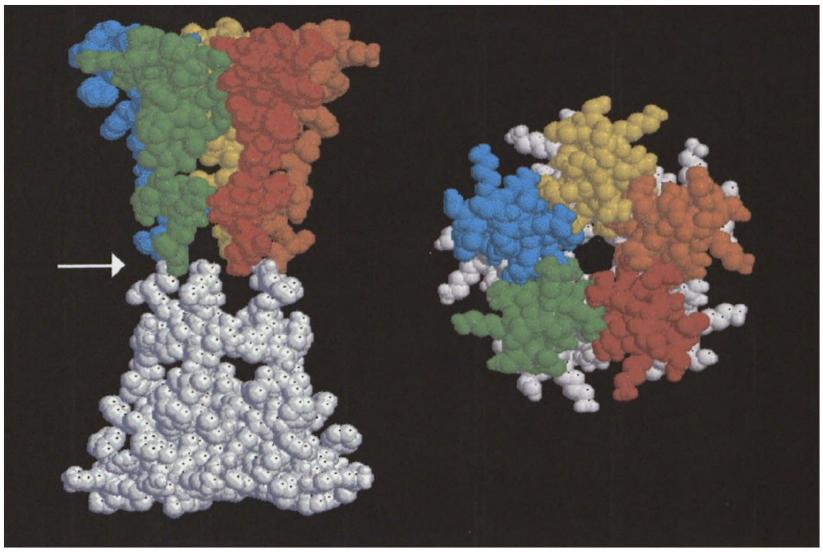


Figure 5.13 Two views of the C-terminal extensions within an *E.coli* GTP cyclohydrolase I decamer The structures shown are space-filling. The five partial subunits of the upper torus are coloured red, green, cyan, yellow, and orange respectively. The partial subunits of the lower torus are coloured grey. The C-terminus of each subunit is found at the junction of the two tori, as indicated by the arrow (images produced using RasMol software).

Although in this part of the molecule there is a degree of empty space, it is questionable as to whether this would be sufficient to accommodate an additional 10 or 13 amino acids per subunit as would be the case for the *P.yoelii* and *P.knowlesi* GTCases respectively, without a considerable rearrangement. The GOR secondary structure prediction algorithm (Garnier *et al.*, 1978), which predicts the α-helical nature of the C-terminal extensions of the *E.coli* and human GTCases with reasonable accuracy, predicts a very diverse set of structures for this part of the three plasmodial

GTCases, varying from predominantly α-helical in the case of *P.y.yoelii*, to predominantly β-sheet in *P.falciparum*. It seems likely therefore, that the structures of this part of the enzyme in the three *Plasmodium* spp. are both different from the structures already observed in non-plasmodial species, and also different from each other.

Analysis of the variation of GTCase and the genes that encode it within *P.falciparum* and *P.yoelii* has illuminated the different population structures of these two species. *Pf gch-1*, the gene encoding *Pf* GTCase, shows identical coding sequences when sequenced from geographically diverse isolates, for example K1 (Thailand) and 3D7 (Africa). It has been suggested that all extant strains of *P.falciparum* have evolved from a recent common ancestor, in existence between 5,000 and 50,000 years ago (e.g. Rich *et al.*, 1998; Volkman *et al.*, 2001). The perfect conservation of the gene sequence in this case seemingly fits with this assertion. On the other hand the GTCase gene sequences in *P.yoelii yoelii* and *P.y.nigeriensis* show a marked degree of divergence. The strains of these two subspecies used in this study are derived from field isolates from the Central African Republic and Nigeria respectively (Killick-Kendrick & Peters, 1978). Although the blood forms of these two subspecies are morphologically indistinguishable (as are the blood forms of all *P.yoelii* and *P.berghei* subspecies), they have been classified on the basis of morphological differences in their respective sporogonic and extraerythrocytic form parasites and also isoenzyme polymorphisms, in addition to their differing geographical distribution. There is however, no biological basis for classifying these organisms as separate species, as they are able to hybridise in the laboratory (although the apparent localisation of *P.y.nigeriensis* in an ecologically isolated fragment of the central African lowland forest system in south-west Nigeria, coupled with continuing destruction of the habitat of the vertebrate host, *Thamnomys rutilans*, means that *P.y.nigeriensis* is seemingly well down the road to speciation 'in the wild'). The discovery of this level of polymorphism in an apparent housekeeping gene suggests that *P.y.yoelii* and *P.y.nigeriensis* have endured many millennia of evolutionary isolation, and provides some justification for the trinomial nomenclature applied to these taxa.

One of the primary reasons cited for the search for GTCase genes in plasmodial species other than *P.falciparum* was the possibility that a less A+T rich gene might be uncovered, and subsequently used as a model for investigating the basic activity of the

gene product of *Pf gch-1*. Unsurprisingly (see Table 5.3) the putative GTCase gene in *P.yoelii* has a similar A+T content to *Pf gch-1* (75.5% in *P.y.yoelii* as opposed to 78.6% in *P.falciparum*;). The putative partial GTCase gene from *P.knowlesi* has a comparatively low A+T content of 62.3%. Given the high level of homology between the 'core' regions of the enzyme in this species and *P.falciparum*, the *P.knowlesi* GTCase gene is worthy of serious consideration as an alternative to *Pf gch-1* in heterologous expression studies.

CHAPTER 6

EFFORTS TO ISOLATE THE GENE ENCODING DIHYDRONEOPTERIN ALDOLASE IN *PLASMODIUM FALCIPARUM*

6.1 : Introduction

Most of the genes encoding enzymes involved in the dihydrofolate biosynthetic pathway in *P.falciparum* (1.8.1) have been cloned and sequenced (Brooks *et al.*, 1994; Triglia & Cowman, 1994; Lee C.S. *et al.*, 2001). The only established folate biosynthetic enzyme for which no gene or polypeptide sequence has been determined in this species is dihydroneopterin aldolase (DHNA)(1.10), which catalyses the conversion of dihydroneopterin to 6-hydroxymethyl-dihydropterin.Many of the genes of this pathway which have been identified thus far have been isolated from the genome using degenerate consensus primers based on highly conserved regions of the respective enzyme (see Hyde & Holloway (1993) for general methodology); the polypeptides encoding DHNA in different species show a comparatively low level of homology however (1.10, Appendix IV), meaning that a PCR-based approach was impractical in this case. We therefore resolved to analyse the data produced by the malaria parasite genome sequencing projects (1.6) as it accumulated in the hope that this would enable the identification of the gene encoding DHNA in *P.falciparum*.

6.2 : Exploitation of an apparent *P.yoelii* DHNA sequence for PCR-based probing of the *P.falciparum* genome

Initial efforts towards electronic isolation of the gene encoding DHNA in *P.falciparum* involved probing of the *Plasmodium* genome databases with amino acid sequences of DHNAs from a number of species using a TBLASTN (protein vs. DNA) search algorithm. Parasite sequences that scored highly against more than one exogenous DHNA sequence were investigated more carefully. This strategy yielded an apparent DHNA sequence from the genome of *P.yoelii* (Figure 6.1). Despite the comparatively low level of homology in the DHNAs of different species, it is a reasonable assumption that this protein would be quite highly conserved between

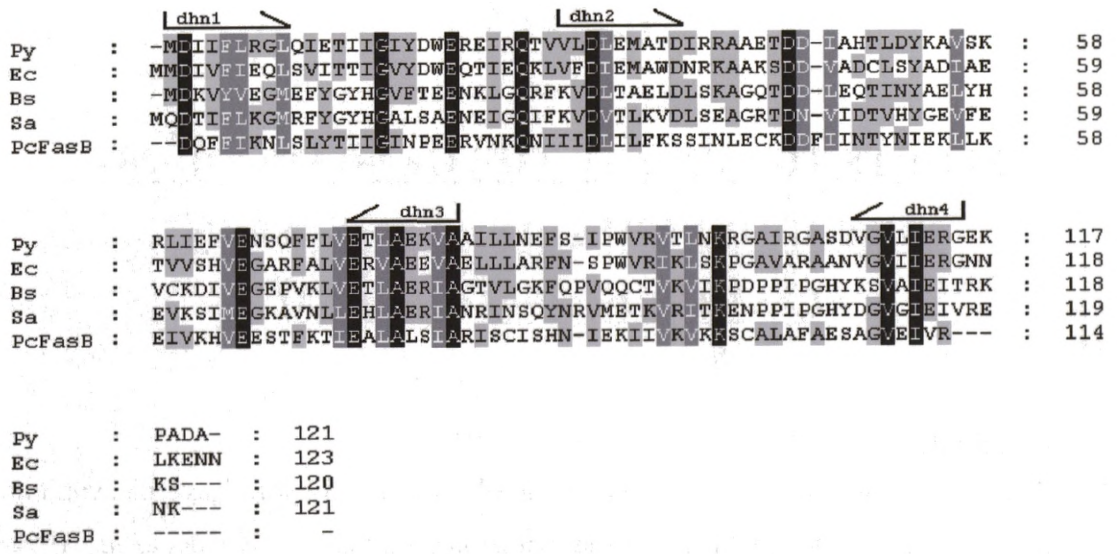


Figure 6.1 Alignment of the putative DHNA from *P.yoelii* with DHNAs from a number of different species Residues that are conserved between all species, semi-conserved between all species and conserved or semi-conserved between 60% of the species are shaded in black, dark grey and light grey respectively. Py, *P.yoelii*; Ec, *E.coli*; Bs, *Bacillus subtilis*; Sa, *Staphylococcus aureus*; PcFasB, *Pneumocystis carinii* folic acid synthesis protein, FasB domain. The position and orientation of the degenerate primers used to probe the *P.falciparum* genome for DHNA are also shown.

different *Plasmodium* spp.. We therefore resolved to use the *P.yoelii* DHNA sequence as a basis for PCR probing of the *P.falciparum* genome. We designed four degenerate primers, dhn1, dhn2, dhn3 and dhn4 (Figure 6.1; Appendix V), based on tracts of this sequence corresponding to those parts of DHNA which show the greatest degree of inter-species homology, but incorporating *P.falciparum* codon bias. Four PCR reaction mixes using *P.falciparum* K1 isolate genomic DNA as template and each different upstream/downstream primer combination were prepared. Each mix was then divided into 6 separate aliquots, and each aliquot subjected to PCR (2.7.1) at a different annealing temperature using an Eppendorf Mastercycler gradient PCR machine (Figure 6.2).

No amplification products were noted when the dhn1/dhn3, dhn1/dhn4 and dhn2/dhn4 primer combinations were used. A series of bands were noted when *P.falciparum* genomic DNA was probed with the dhn2/dhn3 primer combination at low annealing temperatures (<43.5°C)(Figure 6.2A); this series resolved to a single product of around 280bp when annealing temperatures of 44.3°C-47.6°C were used (Figure 6.2B). Assuming 100% sequence length conservation between the putative *P.yoelii* and

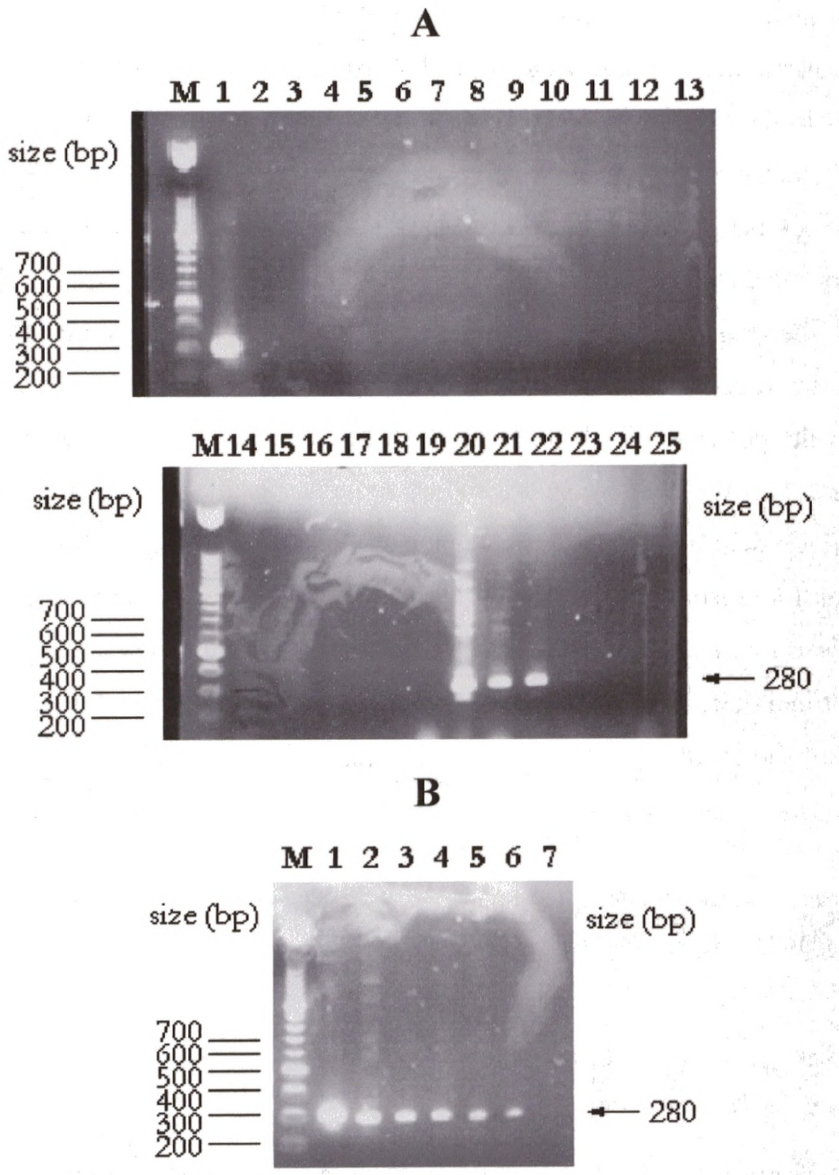


Figure 6.2 Probing of the *P.falciparum* genome with degenerate primers based on the putative *P.yoelii* DHNA A) Broad range of annealing temperatures; Lane numbers: M, 100bp DNA ladder; 1, positive control for template (amplified with *Pf gch-1* specific primers); 2-7, **dhn1/dhn4** primer combination, annealing temperature at: 2, **37.7°C**; 3, **39.3°C**; 4, **43.5°C**; 5, **48.7°C**; 6, **53.1°C**; 7, **55.2°C**; 8-13, **dhn1/dhn3** primer combination, annealing temperatures as per lanes 2-7; 14-19, **dhn2/dhn4** primer combination, annealing temperatures as per lanes 2-7; 20-25, **dhn2/dhn3** primer combination, annealing temperatures as per lanes 2-7.
B) Narrow range of annealing temperatures; Lane numbers: M, 100bp DNA ladder; 1, positive control for template; 2-7, **dhn2/dhn3** primer combination, annealing temperature at: 2, **43.3°C**; 3, **44.3°C**; 4, **45.4°C**; 5, **46.5°C**; 6, **47.6°C**; 7, **48.5°C**.

P.falciparum DHNAs, and no introns in the relevant area of the *P.falciparum* genome, this primer combination should yield a product of 155bp. In order to test whether this disparity was in fact due to the presence of one or more introns, the dhn2/dhn3 primer combination was used to probe a representative aliquot of a cDNA library of *P.falciparum* K1 isolate prepared in λ NM1149. This PCR again yielded a product of around 280bp, indicating that no introns were present in the area of the genome amplified by the original PCR (results not shown). The 280bp genomic DNA specific product was then purified (2.3.1) and sequenced (2.8) using each of the primers used in the PCR (results not shown). Using the full sequence as a probe in a BLASTN (DNA vs. DNA) search of the *P.falciparum* genome sequence database, the PCR product was found to derive not from a DHNA-like sequence, but from a large (>9000bp) ORF on chromosome 14 encoding a polypeptide with considerable homology to acetyl coenzymeA carboxylase, the rate-limiting enzyme in fatty acid synthesis. The sequences of the two primers aligned with the genomic sequence tracts to which they are binding in this case are shown in Figure 6.3.

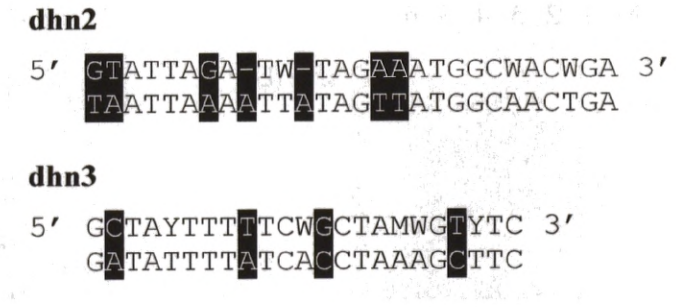


Figure 6.3 Alignment of the sequences of dhn2 & dhn3 with sequence tracts from an ORF encoding a putative acetyl coA carboxylase Primer sequences above, genomic sequences below. Mismatches are highlighted. W=A+T, Y=C+T, M=A+C.

Both oligonucleotides show a large degree of similarity to their priming sites within the acetyl coA carboxylase ORF, with just 7 and 4 mismatches for dhn2 and dhn3 respectively.

Given the paucity of conserved motifs within DHNA it was decided that, having seemingly exhausted the best candidate priming sites, the strategy of PCR-probing with degenerate primers based on the putative *P.yoelii* DHNA was not worthy of further investigation.

Subsequently, closer analysis of the sequence contig including the putative *P.yoelii* DHNA gene revealed that it had an A+T content of just 37.7%, including an

A+T content of 41.3% in that portion encoding the apparent DHNA. This is not consistent with usual nucleotide and codon usage in this species, whose genome is almost as A+T rich as that of *P.falciparum* (1.6); the A+T content of coding areas of the *P.yoelii* genome averages around 72.13% (see Table 5.3). At the time of writing this contig has been excised from the *P.yoelii* genome sequence database at http://www.tigr.org; it therefore seems likely that this sequence derives from contamination of the *P.yoelii* genomic DNA libraries used to generate sequence data, possibly with bacterial DNA.

6.3 : Analysis of a putative *P.falciparum* DHNA sequence identified through protein motif conservation

Following the failure to isolate a *P.falciparum* DHNA sequence using PCR-probing, and the continuing lack of success towards the same end with conventional BLAST searches of the *Plasmodium* genome sequence databases (identification of the spurious *P.yoelii* DHNA described above notwithstanding), an alternative bioinformatic strategy was utilised. The Amino Acid Motif Search facility on the PlasmoDb *Plasmodium* genome resource website (http://www.plasmodb.org) enables the user to identify from the *P.falciparum* genome database all ORFs of greater than 50 aa, and all polypeptides predicted using PHAT, GlimmerM or Genefinder algorithms, which contain a particular amino acid motif. When the amino acid sequence GxxxxExxxxQxxxx[D/S], based on a moderately conserved region of DHNA near the N-terminus, and including several important active site residues (1.10; Appendix IV) was used as a probe, a manageably small number of matched ORFs and predicted polypeptides were found. These sequences were then examined visually, and a promising candidate DHNA sequence, carried on an ORF from chromosome 4, was identified on the basis of its length and, more particularly, additional areas of homology with other DHNAs (Figure 6.4).

This sequence, with no potential in-frame start codon upstream of the conserved motif, appeared to encompass an exon encoding part of a DHNA molecule, with the start codon encoded by an additional upstream exon. It also appeared likely, considering the differences between the polypeptide encoded by the ORF and other DHNA sequences, that the C-terminus of the putative DHNA was also included in another exon. Further indication that this was indeed the case was provided by the discovery of a short sequence, from less than 200bp downstream of the end of the DHNA-like ORF,

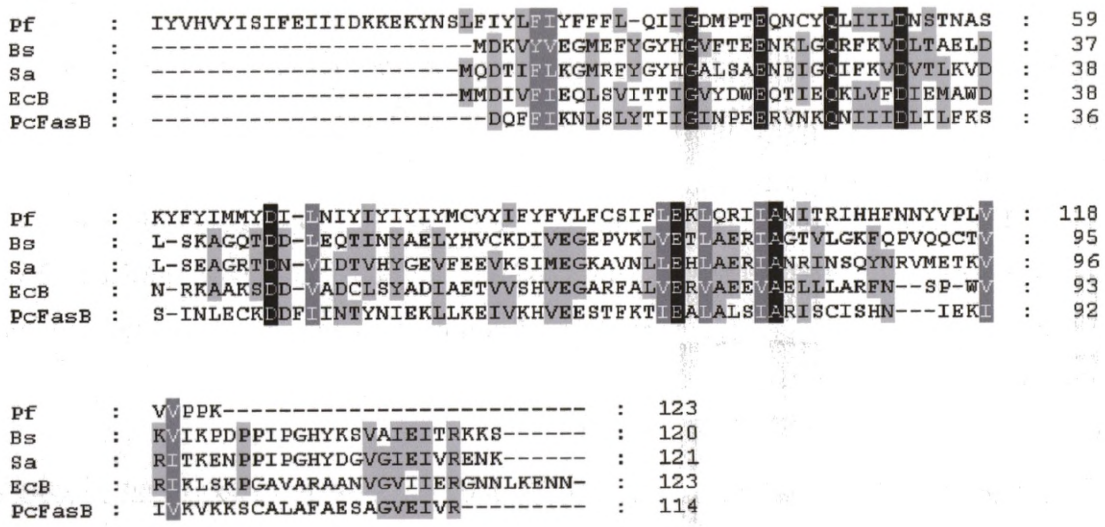


Figure 6.4 Alignment of the DHNA-like polypeptide from *P.falciparum* with DHNAs from a number of different species Residues that are conserved between all species, semi-conserved between all species and conserved or semi-conserved between 60% of the species are shaded in black, dark grey and light grey respectively. Pf, *P.falciparum*; Ec, *E.coli*; Bs, *Bacillus subtilis*; Sa, *Staphylococcus aureus*; PcFasB, *Pneumocystis carinii* folic acid synthesis protein, FasB domain.

in a *P.falciparum* EST (expressed sequence tag) library generated at the University of Tokyo (Figure 6.5). In one particular frame this sequence included a short ORF followed by a tract of tandem AT repeats, and the EST terminated with the sequence CAAA, a tetranucleotide with an established capacity to act as a pre-mRNA 3' cleavage site in eukaryote systems (Wahle & Rüegsegger, 1999). It was therefore deemed likely that this EST represented the 3'end of the mRNA encoding the putative DHNA, and that it included the in frame stop codon. In order to test this hypothesis a number of primers were designed to bind to complementary sequences in this region of chromosome 4. Primers Pfdhn1 and Pfdhn2 (Appendix V) pointed downstream and were complementary to overlapping tracts around the GxxxxExxxxQxxxxD motif, while Pfdhn3 and Pfdhn4 pointed upstream and were complementary to sequence tracts within the EST corresponding to regions of the putative C-terminus encoding ORF and 3' UTR respectively.

1st strand cDNA was synthesised from *P.falciparum* strain HB3 (Honduran isolate) RNA using oligo-p$(dT)_{15}$ primer (Roche) as described in 2.7.2. This was then used as template in a PCR reaction (2.7.1) using the outermost Pfdhn1/Pfdhn4 primer combination. An aliquot of the reaction mix was then subjected to agarose gel electrophoresis (2.6.1) and a single band of around 350bp was noted (result not shown).

```
                                                                    .
aaaaatatatatatatatatagatagatagATATATGTGCATGTTTATATTTCCATTTTT       60
                                < I  Y  V  H  V  Y  I  S  I  F      10

                    .         .         .         .         .         .
GAGATTATAATTGACAAAAAAGAAAAATATAATTCTTTATTTATTTATTTATTTATTTAT       120
 E  I  I  I  D  K  K  E  K  Y  N  S  L  F  I  Y  L  F  I  Y       30

                    .         .         .         .         .         .
TTTTTTTTTTTGCAGATTATTGGCGATATGCCAACTGAGCAAAATTGTTACCAGTTAATA       180
 F  F  F  L  Q  I  I  G  D  M  P  T  E  Q  N  C  Y  Q  L  I       50

                    .         .         .         .         .         .
ATATTAGATAATAGTACAAATGCAAGTAAGTATTTTTACATTATGATGTATGATATATTA       240
 I  L  D  N  S  T  N  A  S  K  Y  F  Y  I  M  M  Y  D  I  L       70

                    .         .         .         .         .         .
AATATATATATATATATATATATATATATGTGTGTGTATATTTTTTATTTTGTTTTATTT       300
 N  I  Y  I  Y  I  Y  I  Y  M  C  V  Y  I  F  Y  F  V  L  F       90

                    .         .         .         .         .         .
TGTTCTATTTTTTTAGAAAAATTACAACGTATAATTGCTAATATTACAAGAATTCATCAT       360
 C  S  I  F  L  E  K  L  Q  R  I  I  A  N  I  T  R  I  H  H      110

                    .         .         .         .         .         .
TTTAATAATTACGTGCCCTTAGTTGTGGTCCCACCAAAGTAAATATATTTATAAAGAGAA       420
 F  N  N  Y  V  P  L  V  V  V  P  P  K  <                         123
                      <  L  W  S  H  Q  S  K  Y  I  Y  K  E  K      13

                    .         .         .         .         .         .
AATAAATAATAAATATATATATATATATATATATATATATATATATATATGTATATATATTT       480
  I  N  N  K  Y  I  Y  I  Y  I  Y  I  Y  I  Y  M  Y  I  Y  L      33

                    .         .         .         .         .         .
ATGTGTGCATATTTTTTTGCCCTTTTGTAGATCGTCATTGGAAGAACACTTTATTTCATT       540
  C  V  H  I  F  L  P  F  C  R  S  S  L  E  E  H  F  I  S  L      53

                    .         .         .         .         .         .
ATGTGTTGAGGATAAAGTTGAGAA[N]GAATATAGTTATGTAAACTTTTTATGTTTTATTCA       600
  C  V  E  D  K  V  E  K  E  Y  S  Y  V  N  F  L  C  F  I  H      73

                    .         .         .         .         .         .
TAAGCTGGTTCATAAAAAAATCGAAGAATCATAGgatgtatatatatatatatatatata       660
  K  L  V  H  K  K  I  E  E  S  <                                  83

                    .         .         .         .         .         .
tatatatatatatatatatatatgaaatatccaagaacttaaagaaaataataaaaagaa       720

                    .         .         .         .         .         .
caa[N]gaaaagagagaaaagaatatttttaaatattatatatatataatttttatatatt       780
```

Figure 6.5 Part of *P.falciparum* chromosome 4 including an ORF encoding a putative partial DHNA Translations of the ORF encoding the DHNA-like polypeptide and the (overlapping) ORF encoding the putative C-terminus of the DHNA are shown. Those residues identified by the amino acid motif search are in bold and underlined. The limits of the sequence isolated from the EST library are highlighted.

The theoretical size of the amplification product produced when these two primers are used with genomic DNA is 563bp. This result therefore suggested that one or more introns, totalling around 200bp, was present in this part of the gene. In order to confirm this suspicion, the PCR product was purified from the reaction mix (2.3.1) and sequenced (2.8) using each of the PCR primers. Two apparent exon junctions were found (Figure 6.6) enabling the identification of two introns in the putative *dhna* gene, hereby described as *x-1* (upstream, 111bp) and *x* (downstream, 112bp) (Figure 6.7). In addition, fewer AT repeats were noted after the in frame stop codon, just 12 as opposed to 22 in the genome database sequence. This was attributed to the different strains used to prepare the two sequences (3D7 (African isolate) for the database sequence, HB3 for the cDNA sequence).

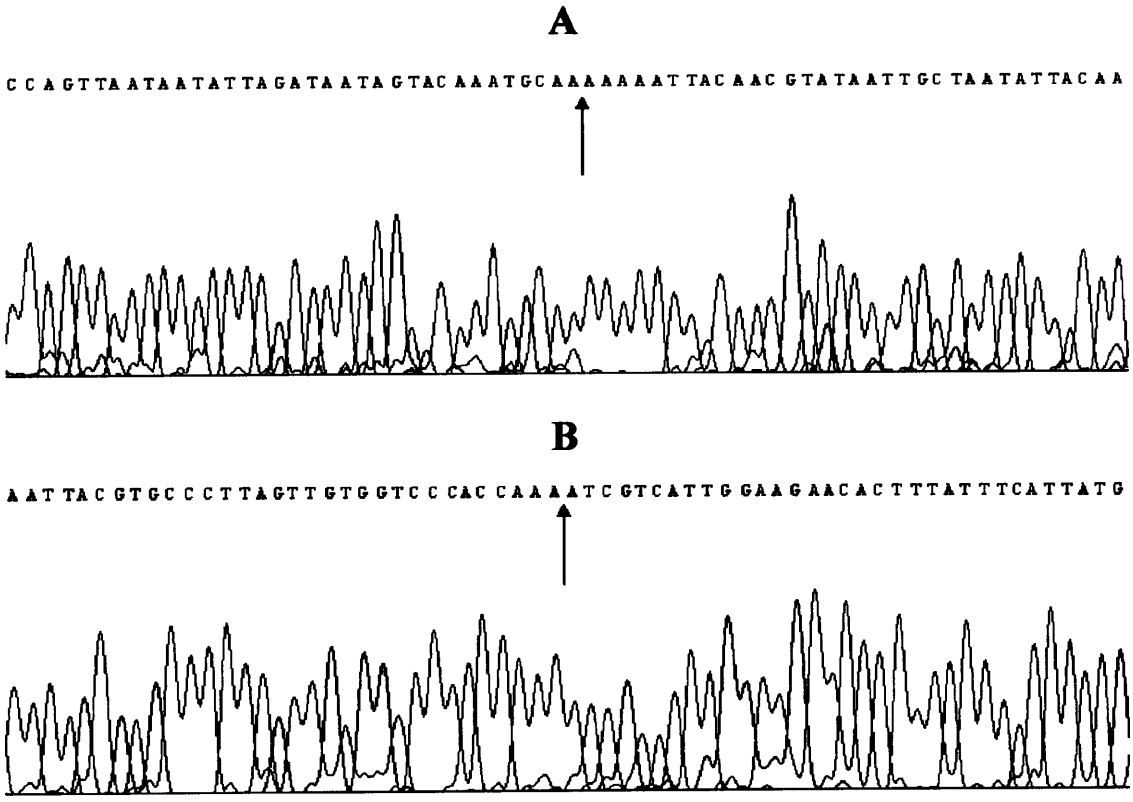


Figure 6.6 Sequence chromatograms indicating the presence of introns *x-1* and *x* in the putative *P.falciparum dhna* gene Exon junctions are indicated with arrows. A) Upstream sequence tract surrounding intron *x-1*. B) Downstream sequence tract surrounding intron *x*. The position of these introns with respect to the genomic DNA sequence is shown in Figure 6.7.

```
        .         .         .         .         .         .
aaaaatatatatatatatatagatagatagATATATGTGCATGTTTATATTTCCATTTTT      60
                              <  I   Y   V   H   V   Y   I   S   I   F      10

        .         .         .         .         .         .
GAGATTATAATTGACAAAAAAGAAAAATATAATTCTTTATTTATTTATTTATTTATTTAT      120
 E   I   I   I   D   K   K   E   K   Y   N   S   L   F   I   Y   L   F   I   Y      30

        .         .         .         .         .         .
TTTTTTTTTTTGCAGATTATTGGCGATATGCCAACTGAGCAAAATTG TACCAGTTAATA      180
 F   F   F   L   Q   I   I   G   D   M   P   T   E   Q   N   C   Y   Q   L   I      50

        .         .         .         .         .         .
ATATTAGATAATAGTACAAATGCAAgtaagtatttttacattatgatgtatgatatatta      240
 I   L   D   N   S   T   N   A                                                      58

        .         .         .         .         .         .
aatatatatatatatatatatatatatatgtgtgtgtatatttttttatttgttttattt      300

        .         .         .         .         .         .
tgttctatttttttagAAAAATTACAACGTATAATTGCTAATATTACAAGAATTCATCAT      360
                  K   K   L   Q   R   I   I   A   N   I   T   R   I   H   H      73

        .         .         .         .         .         .
TTTAATAATTACGTGCCCTTAGTTGTGGTCCCACCAAAgtaaatatatttataaagagaa      420
 F   N   N   Y   V   P   L   V   V   V   P   P   K                              86

        .         .         .         .         .         .
aataaataataaatatatatatatatatatatatatatatatatatatgtatatatattt      480

        .         .         .         .         .         .
atgtgtgcatatttttttgccctttgtagATCGTCATTGGAAGAACACTTTATTTCATT      540
                                  S   S   L   E   E   H   F   I   S   L      96

        .         .         .         .         .         .
ATGTGTTGAGGATAAAGTTGAGAAGGAATATAGTTATGTAAACTTTTTATGTTTTATTCA      600
  C   V   E   D   K   V   E   K   E   Y   S   Y   V   N   F   L   C   F   I   H      116

        .         .         .         .         .         .
TAAGCTGGTTCATAAAAAAATCGAAGAATCATAGgatgtatatatatatatatatatata      660
  K   L   V   H   K   K   I   E   E   S   <                                      126
```

Figure 6.7 Location of introns *x-1* and *x* within the putative *P.falciparum* DHNA gene Introns *x-1* and *x* are shown between nucleotide positions 206-316 and 399-510 respectively. The extreme 5' limit of the cDNA as sequenced at this juncture is highlighted.

The polypeptide encoded by the new, modified version of the putative *dhna* gene (the ORF including the conserved GxxxxExxxxQxxxxD motif, plus the two complete exons identified from the cDNA sequence) was aligned with the DHNAs of several other species (Figure 6.8).

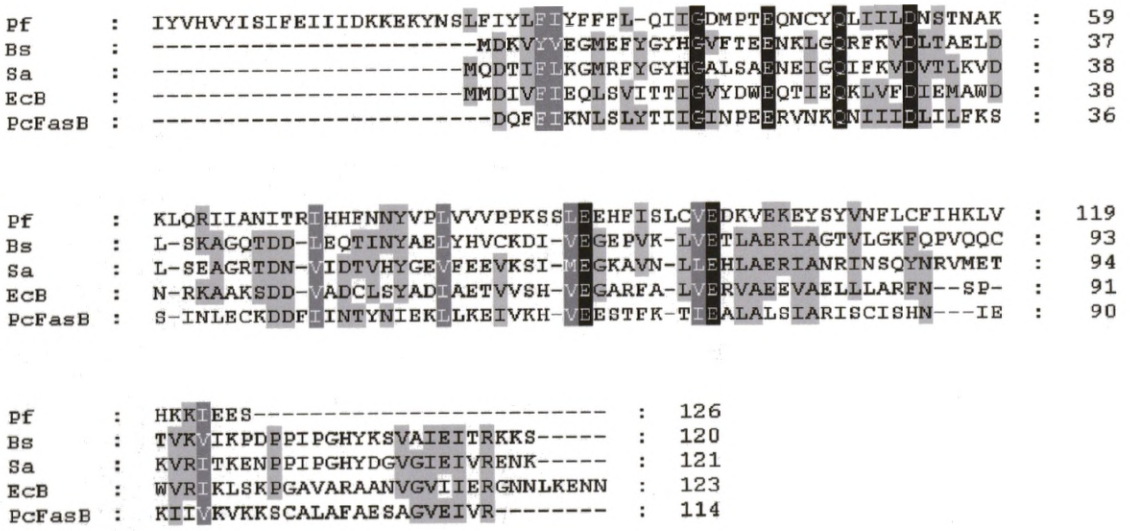


Figure 6.8 Alignment of the revised DHNA-like polypeptide from *P.falciparum* with DHNAs from a number of different species Residues that are conserved between all species, semi-conserved between all species and conserved or semi-conserved between 60% of the species are shaded in black, dark grey and light grey respectively. Key as for Figure 6.4.

Although possibly less striking in terms of its homology to other DHNAs than the original ORF (Figure 6.4), the revised polypeptide was still felt to merit further investigation, as the level of inter-species similarity in DHNAs is not particularly high (1.10, Appendix IV). A new set of primers was designed to amplify upstream regions of the cDNA. Primers Pfdhn5 and Pfdhn6 (Appendix V) pointed upstream and were designed to span introns *x* and *x-1* respectively, while Pfdhn7, Pfdhn8 and Pfdhn9 pointed downstream and were complementary to regions of coding-type sequence in the portion of chromosome 4 upstream of the GxxxxExxxxQxxxxD motif (Pfdhn7 was designed to bind furthest upstream).

1st strand cDNA synthesised from *P.falciparum* strain HB3 RNA as described above was used as template in a PCR reaction (2.7.1) using the outermost Pfdhn5/Pfdhn7 primer combination. An aliquot of the reaction mix was then subjected to agarose gel electrophoresis (2.6.1) and a single band of around 500bp was noted (result not shown). This result again suggested the presence of one or more introns, totalling around 180bp, in addition to *x* and *x-1*. The PCR product was then purified from the reaction mix (2.3.1) and sequenced (2.8) using each of the PCR primers. One additional exon junction was found (Figure 6.9) enabling the identification of one more intron in the putative *dhna* gene, hereby described as *x-2* (153bp) (Figure 6.10). One

additional difference in the cDNA sequence and the genome database sequence was also found, a change from a C to a T at nucleotide position -198, conferring a change in the polypeptide sequence of histidine to tyrosine (result not shown).

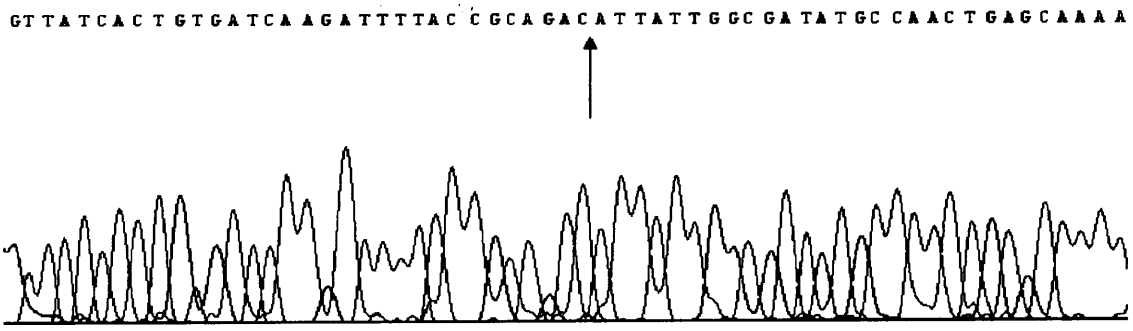


Figure 6.9 Sequence chromatogram indicating the presence of intron *x-2* in the putative *P.falciparum dhna* gene The exon junctions are indicated with an arrow. The position of this intron with respect to the genomic DNA sequence is shown in Figure 6.10.

```
AATTATTCAAAGATAATTATAGATAACTTGGCAGCCATATTATATTCCTATCGTTTAAAT    -541
 N  Y  S  K  I  I  I  D  N  L  A  A  I  L  Y  S  Y  R  L  N     1132

GTAAAATGTCAAAATGATACGTTGTCATGTGACAATgtgtatatatgtaaataatttata    -481
 V  K  C  Q  N  D  T  L  S  C  D  N (V  Y  I  C  K  <)         1149

tgtataaatattatacatacatacatacatacatatatataaatattttatatatatata    -421

tatatatatttttttttttagTGTGCGACGTCGGCTCACTCGGGACAATTAATTTTACCA    -361
                     <  C  A  T  S  A  H  S  G  Q  L  I  L  P      13

GACACTTAAAATTGTTACCTCTATTTACCTCGTGTTTATTAAAACATAATGTTGTGAAG    -301
 D  T  L  K  L  L  P  L  F  T  S  C  L  L  K  H  N  V  V  K      33

AAGGATGCTTTACATGATTTAAAGGTGTACAACTTAATAAAATTGTTGTCGATGCCAATA    -241
 K  D  A  L  H  D  L  K  V  Y  N  L  I  K  L  L  S  M  P  I      53

ATATCTTCTCTTTTGTTTGTATATCCAGTGATGTATTTAATTCATATGAAAGGTCGACAT    -181
 I  S  S  L  L  F  V  Y  P  V  M  Y  L  I  H  M  K  G  R  H      73

AATGAAATTGATTATATGGATGTAGATGACGAGAATTTCCTTCCAAGAGCGATACCGACA    -121
 N  E  I  D  Y  M  D  V  D  D  E  N  F  L  P  R  A  I  P  T      93

AGTGCTGAAAAAATATATTCTAGCGGTATATATTTATTGGATTGTTCAACCCATTTATAT     -61
 S  A  E  K  I  Y  S  S  G  I  Y  L  L  D  C  S  T  H  L  Y     103
```

```
          .         .         .         .         .         .
CTTTATTTTGGTTATCACTGTGATCAAGATTTTACCGCAGACgtaaaaaaaaaaaaaaaa      -1
 L  Y  F  G  Y  H  C  D  Q  D  F  T  A  D                        117

          .         .         .         .         .         .
aaaaatatatatatatatatagatagatagatatatgtgcatgtttatatttccattttt       60

          .         .         .         .         .         .
gagattataattgacaaaaaagaaaaatataattctttatttatttatttatttatttat      120

          .         .         .         .         .         .
ttttttttttttgcagATTATTGGCGATATGCCAACTGAGCAAAATTGTTACCAGTTAATA      180
                I  I  G  D  M  P  T  E  Q  N  C  Y  Q  L  I       132

          .         .         .         .         .         .
ATATTAGATAATAGTACAAATGCAAgtaagtatttttacattatgatgtatgatatatta      240
 I  L  D  N  S  T  N  A                                          140
```

Figure 6.10 Location of intron *x-2* and putative location of intron *x-3* within the putative *P.falciparum* DHNA gene Intron *x-2* is shown between nucleotide positions -18 and 135. A possible location for intron *x-3* is shown between nucleotide positions -504 and -400 respectively. The extreme 5' limit of the cDNA as sequenced at this juncture is highlighted. Amino acid numbering for those exons already identified starts from the cysteine residue encoded by the codon at positions -399 to -397. Amino acid numbering for the upstream ORF starts from the first residue (also the first methionine) in this ORF (not shown). The peptide sequence encoded by that part of the upstream ORF that falls within the boundaries of intron *x-3* is shown in brackets.

The identification of intron *x-2* effectively stitched the three complete exons sequenced from the cDNA with a sizeable, upstream ORF (nucleotide positions -399 to -19). The peptide encoded by this ORF includes no methionine residues within the first 50 aa, nor does the ORF contain any potential 3' intron splicing sites (as defined by e.g. Padgett *et al.*, 1986) upstream of the limit of the known cDNA sequence (although there is a strong candidate site immediately upstream of the ORF) (Figure 6.10); it is therefore reasonable to conclude that, in addition to that part of the ORF already sequenced from the cDNA, the entire 5' portion is also included within the coding sequence of the putative *dhna* gene. This extended the putative polypeptide encoded by this gene to an apparent absolute minimum length of 209 aa, longer than any DHNA domain (see Appendix IV), although shorter than the combined lengths of the FasA and FasB domains of the multifunctional folic acid synthesis proteins of *Pneumocystis carinii* (Volpe *et al.*, 1992) and *S.cerevisiae* (gene *FOL1*, http://genome-www.stanford.edu/Saccharomyces) (see 1.10).

Having established the almost certain existence of a fourth intron within the gene, hereby described as *x-3*, we resolved to examine the sequence upstream of the

likely 3' splicing site (shown at nucleotide position -400 in Figure 6.10) in all three reading frames in order to identify potential 5' splicing sites for intron *x-3*. This led to the discovery of a large, uninterrupted ORF (nearly 3.5 kb) less than 100bp upstream of the presumed 3' splicing site of intron *x-3* (the polypeptide encoded by the 3'-most portion of this ORF is shown in Figure 6.10). A number of potential 5' intron splicing sites enabling in-frame connection of the polypeptide encoded by this ORF with that encoded by the exons uncovered by cDNA sequencing were identified reasonably close to the 3' terminus of the ORF, one of which (that closest to the terminus) is shown in Figure 6.10.

Probing of the SwissProt protein sequence database with the 1362 aa composite polypeptide sequence produced in this manner using a BLASTP (protein vs. protein) search algorithm revealed that this theoretical polypeptide possessed considerable homology, including at its C-terminus, to the sec24 family of protein transport vesicle coat proteins (results not shown). This result strongly suggested that the polypeptide encoded by the gene under scrutiny was not in fact DHNA, but the C-terminal part of a sec24 homologue, and therefore, no further analysis of this gene was undertaken.

6.4 : Discussion

The unsuccessful efforts to isolate the *dhna* gene from the genome of *P.falciparum* illustrate some of the potential pitfalls of both gene identification in the post-genomic era and isolation of poorly conserved genes on the basis of homology. With regards to the first point, the appearance of a seemingly exogenous DHNA sequence in the *P.yoelii* genome sequence database (6.2) is a great cause for concern. This sequence is not found in any of the numerous complete microbial genomes, but shows considerable homology to DHNAs from prokaryotic organisms, most particularly the gamma subdivision of the gram-negative Proteobacteria (results not shown). This group includes the Enterobacteria (*E.coli, Salmonella typhimurium* etc.) and other gastrointestinal pathogens including *Vibrio cholerae*; it is possible that the source of this contamination is a pathogenic bacterium of the rodents in which the parasites used to prepare the sequence library were cultured, although this would represent a serious lapse in experimental hygiene. In any case, while it is to the benefit of the entire scientific community that preliminary genome sequence data should be released into the public domain as it accumulates, this incident clearly illustrates the caution that must be exercised when dealing with such information.

Many of the other problems encountered have stemmed from the paradoxical nature of searching for a poorly conserved gene on the basis of inter-species homology. The use of degenerate primers to PCR-probe the parasite genome proved singularly unsuccessful in this case (the serendipitous isolation of a completely unrelated gene underlines the potential for 'non-specific' binding of degenerate primers however), although this would have been anticipated if the exogenous nature of the apparent DHNA sequence from the *P.yoelii* genome database had been identified at an earlier juncture. The original intention was to use a plasmodial sequence as a template for PCR-probing of the *P.falciparum* genome. While this approach is still likely to be practical in technical terms, this type of PCR-based methodology for isolation of genes in *P.falciparum* is now largely redundant in the face of bioinformatic modes of gene identification, as the *P.falciparum* genome database is, at the time of writing, virtually complete. Searching of all ORFs and predicted peptides in the *P.falciparum* genome database for the GxxxxExxxxQxxxx[D/S] motif, arguably the most conserved set of residues in the DHNAs of different species (Appendix IV) yielded just a single promising candidate DHNA, which upon further analysis has proved to be almost certainly a 'false positive'. Ongoing work probing the *Plasmodium* genome databases with DHNA sequences and a raft of amino acid motifs has failed to yield further suitable candidates (Blagborough A. & Hyde, J., personal communication). Given that, for example, the same databases contain three separate identifiable GTCases (5.2), it seems a stretch of the imagination to believe that *dhna* genes await sequencing in all of the plasmodial genomes. The possibility must therefore be entertained that either plasmodial DHNAs are particularly poorly conserved, even compared with the low level of inter-species homology normally exhibited by this enzyme, or that the catalytic activity provided by DHNA in other species is provided by an entirely different protein in the malaria parasite. As a result, identification of the protein or proteins carrying 'DHNA' activity and their respective genes in *Plasmodium* spp. may require radical alternative experimental strategies. Some discussion of this subject is given in the next chapter.

CHAPTER 7

CONCLUSIONS

The purification of heterologously expressed fusion polypeptides incorporating C-terminal fragments of the gene product of *Pf gch-1*, the putative *Pf* GTCase (Chapter 3), should provide material which can be assayed for GTCase activity, once protocols for cleaving and separating the parasite-specific domain from the fusion have been finalised. This should enable us to unambiguously assign GTCase activity to this protein (at present, the identification of this species as a GTCase is based on the high level of conservation of the C-terminal half of the protein relative to authenticated GTCases, coupled with the apparent dearth of any other suitable candidates in the now virtually complete *P.falciparum* genome sequence). However, while such studies could confirm that *Pf gch-1* genuinely codes for the malaria parasite GTCase and provide basic kinetic data, detailed analysis of the enzyme function would require the purification of a species including the N-terminal region of the putative *Pf* GTCase, which possibly has a role in physiological regulation of the enzyme's activity, although it appears that, whatever the function of this part of the protein may be in *Plasmodium* spp., it is not subject to stringent sequence requirements at the primary level (see Chapter 5).

The generation of antibodies seemingly able to bind specifically to the putative *Pf* GTCase (3.7) offers other opportunities. Use of the antibody as an affinity ligand on a chromatography column might enable isolation of *Pf* GTCase from parasite extracts. In the absence of full-length heterologously produced protein, native material purified in this way may prove invaluable in terms of determination of enzyme activity. In addition, labelling of the antibodies with a fluorescent tag will hopefully allow investigation of the subcellular localisation of the enzyme; initial studies towards this end suggest some association with the food vacuole (Read, M., personal communication).

Heterologous expression of *Pf gch-1* in non-bacterial systems (Chapter 4) has proved largely unsuccessful; constructs carrying this gene and modified versions thereof in two separate vectors were unable to effect full or partial complementation

of a *Saccharomyces cerevisiae* GTCase knockout mutant, while preliminary expression of *Pf gch-1* in a human cell line failed for obscure reasons. Broader questions as to the usefulness of such systems in the expression of malaria parasite genes remain however. The failure of those constructs carrying the gchmut[3] insert (the version of *Pf gch-1* modified to remove yeast pre-mRNA 3' end-processing signals) to complement the GTCase deficiency in the yeast knockout mutant is particularly interesting. The current state of knowledge on yeast pre-mRNA 3' end-processing suggests that the modifications incorporated into *Pf gch-1* must surely have affected the pattern of mRNA transcript maturation from this gene, but the most pertinent question is to what extent the pattern has changed. It is entirely possible that significant quantities of full length transcript are being produced in one or more systems, but that the polypeptide thus produced is unable to complement the folinic acid auxotrophy, which would raise doubts as to the viability of any future complementation tests with *Pf gch-1* in yeast. Alternatively, the pre-mRNA transcribed from the modified gene may be undergoing premature truncation at a point or points significantly further into the gene, as happened following similar mutagenesis of a yeast expression construct containing the cDNA coding for *P.falciparum* topoisomerase II (Sibley *et al.*, 1997). It may be, however, that the modifications have effected no significant change, or at most a small increase, in transcript length. If this were the case, it would severely bring into question the value of such site-directed mutagenesis in the expression of recalcitrant *P.falciparum* genes in yeast and suggest that complete gene resynthesis may be a more useful tool in such cases. Determination of the pattern of transcripts in this expression system remains a matter of some interest therefore.

Although transient transfection of human cells with *Pf gch-1* was apparently unsuccessful in this case, mammalian cells are worthy of further investigation as a host for heterologous expression of *Pf gch-1* and other parasite genes. The expression vector used (pcDNA4/HisMaxA; Appendix V) is recommended for expression in a broad spectrum of cell types; the existence of *Pf gch-1* derived constructs produced using this vector should therefore provide a spur towards further experimentation.

The observations regarding the variation in *Pf gch-1* transcript levels during the erythrocytic cycle of *P.falciparum* (5.1) fit well with previous related data. This work represents part of a larger project within the UMIST Molecular Parasitology group to characterise stage-specific expression of all the folate-related genes in

antifolate sensitive and resistant strains of the parasite under a variety of physiological conditions. This initiative should provide much food for thought regarding the folate biosynthetic pathway in *P.falciparum.*

The identification of GTCase homologues in several other *Plasmodium* spp. (and subspecies) (5.2) has hinted at the unusual nature of this protein in the malaria parasite. The apparent lack of conservation in the N-terminal and C-terminal extensions of the enzyme (to a particularly unusual degree in the C-terminus) suggests that a novel structure-function relationship is possibly present in malarial GTCases. Data from other parasite genome sequence projects coupled with biochemical analysis of *Pf* GTCase should provide further illumination. In addition, the identification of significant divergence between the GTCase sequences of the two subspecies of *P.yoelii*, *P.y.yoelii* and *P.y.nigeriensis*, is of importance to rodent malaria parasite systematics, while the comparatively G+C rich *P.knowlesi* GTCase gene, when fully characterised, is worthy of consideration as a model for analysis of *Pf* GTCase function by expression in heterologous systems.

The failure to isolate a suitable candidate DHNA gene from amongst the accumulating sequence data of the malaria parasite genome projects (including the near-complete *P.falciparum* database) using conventional BLAST or protein motif searches (Chapter 6) has led to speculation that this enzyme may either be highly diverged in *Plasmodium* compared with other taxa, or replaced by an enzyme of entirely different provenance. In order to identify the protein carrying 'DHNA' activity in the parasite it may therefore be necessary to isolate candidates on structural or functional bases. The ORF (ostensible recognition of folds) genome search algorithm (Aurora & Rose, 1998; Xu *et al.*, 1999) searches for protein homologues by comparing the predicted secondary structure of polypeptides encoded by ORFs within genome databases with that of the probe sequence, a strategy that has proved successful in identifying otherwise unrecognisable homologues of DHPS and TS in the archaeon *Methanococcus janaschii*. The structure of the DHNA of *Staphylococcus aureus* has been solved to high resolution (Hennig *et al.*, 1998); use of this species as a probe in a search for secondary structure homologues within the plasmodial databases may prove successful in identifying DHNA. In addition, 'old-fashioned' attempts to isolate DHNA from parasite extracts based on its ability to bind the closely similar substrate analogue neopterin are also being considered.

As an illustration of the unexpected directions in which scientific research can lead, the positive identification of several introns in a gene originally postulated as a DHNA candidate (6.3) may help to further knowledge of an aspect of parasite biology far removed from folate biosynthesis. The polypeptide encoded by this gene has been identified as a homologue of Sec24p, part of the COPII coat protein complex, found at the surface of vesicles involved in the trafficking of proteins between the endoplasmic reticulum and the Golgi body in *S.cerevisiae* and other organisms (Antonny & Schekman, 2001). Researchers from La Trobe University, Melbourne, Australia have previously identified homologues of two other COPII proteins, Sar1p (Albano *et al.*, 1999) and Sec31p (Adisa *et al.*, 2001) in *P.falciparum*, and presented evidence that these proteins play a role in export of proteins to the surface of the pRBC. We have furnished these experimenters with the sequence generated during our abortive search for the gene encoding DHNA, in the hope that it may aid elucidation of the role of the COPII protein complex in the malaria parasite.

In conclusion, the observations that the malaria parasite seemingly possesses highly divergent GTCase and DHNA molecules, coupled with its bifunctional DHFR-TS, PPPK-DHPS and DHFS-FPGS enzymes (the latter being the only example of its type determined in eukaryotes)(1.8) highlights the unusual nature of folate biosynthesis in this organism, and reinforces the value and further potential of this metabolic pathway as a target for chemotherapeutic intervention.

CHAPTER 8

REFERENCES

Adisa A., Albano F.R., Reeder J., Foley M. & Tilley L. (2001) Evidence for a role for a *Plasmodium falciparum* homologue of Sec31p in the export of proteins to the surface of malaria parasite-infected erythrocytes. *J. Cell Sci.* **114** 3377-3386.

Albano F.R., Berman A., La Greca N., Hibbs A.R., Wickham M., Foley M. & Tilley L. (1999) A homologue of Sar1p localises to a novel trafficking pathway in malaria-infected erythrocytes. *Eur. J. Cell Biol.* **78** 453-462.

Al-Khedery B., Barnwell J.W. & Galinski M.R. (1999) Stage-specific expression of 14-3-3 in asexual blood-stage *Plasmodium*. *Mol. Biochem. Parasitol.* **102** 117-130.

Antonny B. & Schekman R. (2001) ER export: public transportation by the COPII coach. *Curr. Opinion Cell Biol.* **13** 438-443.

Asawamahasakda W. & Yuthavong Y. (1993) The methionine synthesis cycle and salvage of methyltetrahydrofolate from host red cells in the malaria parasite (*Plasmodium falciparum*). *Parasitol.* **107** 1-10.

Auerbach G. & Nar H. (1997) The pathway from GTP to tetrahydrobiopterin: Three-dimensional structures of GTP cyclohydrolase I and 6-pyruvoyl tetrahydropterin synthase. *Biol. Chem.* **378** 185-192.

Auerbach G., Herrmann A., Bracher A., Bader G., Gutlich M., Fischer M., Neukamm M., Garrido-Franco M., Richardson J., Nar H., Huber R. & Bacher A. (2000) Zinc plays a key role in human and bacterial GTP cyclohydrolase I. *Proc. Natl. Acad. Sci. USA* **97** 13567-13572.

Aurora R. & Rose G.D. (1998) Seeking an ancient enzyme in *Methanococcus jannaschii* using ORF, a program based on predicted secondary structure comparisons. *Proc. Natl. Acad. Sci. USA* **95** 2818-2823.

Berendt A.R., Ferguson D.J.P. & Newbold C.I. (1990) Sequestration in *Plasmodium falciparum* malaria; sticky cells and sticky problems. *Parasitol. Today* **6** 247-254.

Beverley S.M., Ellenberger T.E. & Cordingley J.S. (1986) Primary structure of the gene encoding the bifunctional dihydrofolate reductase-thymidylate synthase of *Leishmania major*. *Proc. Natl. Acad. Sci. USA* **83** 2584-2588.

Blackman M.J., Scott-Finnigan T.J., Shai S. & Holder A.A. (1994) Antibodies inhibit the protease-mediated processing of a malaria merozoite surface protein. *J. Exp. Med.* **180** 389-393.

Bognar A.L., Osborne C. & Shane B. (1987) Primary structure of the *Escherichia coli folC* gene and its folylpolyglutamate synthetase-dihydrofolate synthetase product and regulation of expression by an upstream gene. *J. Biol. Chem.* **262** 12337-12343.

Bowman S., Lawson D., Basham D., Brown D., Chillingworth T., Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T., Gentles S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T., Horrocks P., Jagels K., Jassal B., Kyes S., McLean J., Moule S., Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutter S., Skelton J., Squares R., Squares S., Sulston J.E., Whitehead S., Woodward J.R., Newbold C. & Barrell B.G. (1999) The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. *Nature* **400** 532-538.

Bracher A., Fischer M., Eisenreich W., Ritz H., Schramek N., Boyle P., Gentili P., Huber R., Nar H., Auerbach G. & Bacher A. (1999) Histidine 179 mutants of GTP cyclohydrolase I catalyze the formation of 2-amino-5-formylamino-8-ribofuranosylamino-4(3H)-pyrimidinone triphosphate. *J. Biol. Chem.* **274** 16727-16735.

Bracher A., Schramek N. & Bacher A. (2001) Biosynthesis of pteridines. Stopped-flow kinetic analysis of GTP cyclohydrolase I. *Biochem.* **40** 7896-7902.

Brooks D.R., Wang P., Read M., Watkins W.M., Sims P.F.G. & Hyde J.E. (1994) Sequence variation of the hydroxymethyldihydropterin pyrophosphokinase-dihydropteroate synthase gene in lines of the human malaria parasite, *Plasmodium falciparum*, with differing resistance to sulfadoxine. *Eur. J. Biochem.* **224** 397- 405.

Brown G.M. & Williamson J.M. (1987) Biosynthesis of folic acid, riboflavin, thiamine and pantothenic acid. *In: Neidhart, F.C. (Ed) Escherichia coli and Salmonella typhimurium. Cellular and Molecular Biology, Book 1 (pp521-538).* American Society for Microbiology, Washington, USA.

Bruce-Chwatt L.J. (1985) *Essential malariology.* William Heinemann Medical Books Ltd., Oxford, UK.

Butler D., Maurice J. & o'Brien C. (1997) Time to put malaria control on the global agenda. *Nature* **386** 535-536.

Bystroff C., Oatley S.J. & Kraut J. (1990) Crystal structures of *Escherichia coli* dihydrofolate reductase - the NADP+ holoenzyme and the folate-NADP+ ternary complex - substrate binding and a model for the transition state. *Biochem.* **29** 3263-3277.

Bzik D.J., Li W.B., Horii T. & Inselburg J. (1987) Molecular cloning and sequence analysis of the *Plasmodium falciparum* dihydrofolate reductase-thymidylate synthase gene. *Proc. Natl. Acad. Sci. USA* **84** 8360-8364.

Carucci D.J. (2000) Malaria research in the post-genomic era. *Parasitol. Today* **16** 434-438.

Catteruccia F., Nolan T., Loukeris T.G., Blass C., Savakis C., Kafatos F.C. & Crisanti A. (2000) Stable germline transformation of the malaria mosquito *Anopheles stephensi. Nature* **405** 959-962.

Clyde D.F., McCarthy V.C., Miller R.M. & Hornick R.B. (1973) Immunization of man against sporozoite-induced falciparum malaria. *Am. J. Med. Sci.* **266** 169-177.

Cogswell F.B. (1992) The hypnozoite and relapse in primate malaria. *Clin. Microbiol. Rev.* **5** 26-35.

Collins F.H. & Besansky N.J. (1994) Vector biology and the control of malaria in Africa. *Science* **264** 1874-1875.

Colloc'h N., el Hajji M., Bachet B., l'Hermite G., Schiltz M., Prange T., Castro B. & Mornon J.P. (1997) Crystal Structure of the protein drug urate oxidase-inhibitor complex at 2.05 angstrom resolution. *Nature Struc. Biol.* **4** 947-952.

Colloc'h N., Poupon A. & Mornon J.P. (2000) Sequence and structural features of the T-fold, an original tunnelling building unit. *Proteins-Structure Function and Genetics* **39** 142-154.

Cooke B.M., Wahlgren M. & Coppel R.L. (2000) Falciparum malaria: Sticking up, standing out and out-standing. *Parasitol. Today* **16** 416-420.

Cossins E.A. & Chen L.F. (1997) Folates and one-carbon metabolism in plants and fungi. *Phytochem.* **45** 437-452.

Crewther P.E., Matthew M.L.S.M., Flegg R.H. & Anders R.F. (1996) Protective immune responses to apical membrane antigen 1 of *Plasmodium chabaudi* involve recognition of strain-specific epitopes. *Infection and Immunity* **64** 3310-3317.

Deng H., Callender R. & Dale G.E. (2000) A vibrational structure of 7,8-dihydrobiopterin bound to dihydroneopterin aldolase. *J. Biol. Chem.* **275** 30139-30143.

De Saizieu A., Vankan P. & Van Loon A.P.G.M. (1995) Enzymatic characterization of *Bacillus subtilis* GTP cyclohydrolase-I - evidence for a chemical dephosphorylation of dihydroneopterin triphosphate. *Biochem. J.* **306** 371-377.

Desowitz R.S. (1991) *The Malaria Capers.* W.W.Norton & Company, Inc., New York, USA.

Dorn A., Stoffel R., Matile H., Bubendorf A. & Ridley R.G. (1995) Malarial haemozoin beta-hematin supports heme polymerization in the absence of protein. *Nature* **374** 269-271.

Doyle K. (ed) (1996) *Protocols and applications guide.* 3rd ed. Promega corporation, Madison, Wisconsin, USA.

Escalante A.A., Freeland D.E., Collins W.E. & Lal A.A. (1998) The evolution of primate malaria parasites based on the gene encoding cytochrome b from the linear mitochondrial genome. *Proc. Natl. Acad. Sci. USA* **95** 8124-8129.

Fairlamb A.H. (1989) Novel biochemical pathways in parasitic protozoa. *Parasitol.* **99** S93-S112.

Ferone R. (1973) The enzymic synthesis of dihydropteroate and dihydrofolate by *Plasmodium berghei. J. Protozool.* **20** 459-464.

Fidock D.A., Nomura T., Talley A.K., Cooper R.A., Dzekunov S.M., Ferdig M.T., Ursos L.M.B., Sidhu A.B.S., Naude B., Deitsch K.W., Su X.Z., Wootton J.C., Roepe P.D. & Wellems T.E. (2000) Mutations in the *P.falciparum* digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. *Mol. Cell* **6** 861-871.

Francis S.E., Sullivan D.J. & Goldberg D.E. (1997) Hemoglobin metabolism in the malaria parasite *Plasmodium falciparum. Ann. Rev. Microbiol.* **51** 97-123.

Freund J. (1947) Some aspects of active immunization. *Ann. Rev. Microbiol.* **1** 291-308.

Gamboa de Dominguez, N.D. & Rosenthal, P.J. (1996) Cysteine proteinase inhibitors block early steps in hemoglobin degradation by cultured malaria parasites. *Blood* **87** 4448-4454.

Garcia L.S. & Bruckner D.A. (1993) Malaria and *Babesia* spp. *In: Diagnostic Medical Parasitology (pp113-138).* 2nd ed. American Society for Microbiology, Washington, USA.

Garcia-Sanz J.A. (1997) mRNA Expression. *In: Lefkovits I. (Ed) Immunology Methods Manual, Volume 1 (pp381-482).* Academic Press, Inc., San Diego, California, USA.

Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L., Koonin E.V., Shallom S., Mason T., Yu K., Fujii C., Pederson J., Shen K., Jing J.P., Aston C., Lai Z.W., Schwartz D.C., Pertea M., Salzberg S., Zhou L.X., Sutton G.G., Clayton R., White O., Smith H.O., Fraser C.M., Adams M.D., Venter J.C. & Hoffman S.L. (1998) Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum. Science* **282** 1126-1132.

Garnier J., Osguthorpe D.J. & Robson B. (1978) Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* **120** 97-120.

Ginsburg H., Ward S.A. & Bray P.G. (1999) An integrated model of chloroquine action. *Parasitol. Today* **15** 357-360.

Goidin D., Mamessier A., Staquet M.J., Schmitt D. & Berthier-Vergnes O. (2001) Ribosomal 18S RNA prevails over glyceraldehyde-3-phosphate dehydrogenase and beta-actin genes as internal standard for quantitative comparison of mRNA levels in invasive and noninvasive human melanoma cell subpopulations. *Anal. Biochem.* **295** 17-21.

Goffeau A., Barrell B.G., Bussey H., Davis R.W., Dujon B., Feldmann H., Galibert F., Hoheisel J.D., Jacq C., Johnston M., Louis E.J., Mewes H.W., Murakami Y., Philippsen P., Tettelin H. & Oliver S.G. (1996) Life with 6000 genes. *Science* **274** 546-551.

Graber J.H., Cantor C.R., Mohr S.C. & Smith T.F. (1999) Genomic detection of new yeast pre-mRNA 3'-end-processing signals. *Nucleic Acids Res.* **27** 888-894.

Gray M.W., Burger G. & Lang B.F. (1999) Mitochondrial evolution. *Science* **283** 1476-1481.

Green J.M., Nichols B.P. & Matthews R.G. (1996) Folate biosynthesis, reduction and polyglutamylation. *In: Neidhart, F.C. (Ed) Escherichia coli and Salmonella. Cellular and Molecular Biology (pp665-673).* 2nd ed. American Society for Microbiology, Washington, USA.

Greenwood B.M. (1997) What's new in malaria control? *Annals Trop. Med. Parasitol.* **91** 523-531.

Griffith O.W. & Stuehr D.J. (1995) Nitric oxide synthases - properties and catalytic mechanism. *Ann. Rev. Physiol.* **57** 707-736.

Guo Z.J. & Sherman F. (1996) 3'-end-forming signals of yeast mRNA. *Trends Biochem. Sci.* **21** 477-481.

Hatakeyama K., Inoue Y., Harada T. & Kagamiyama H. (1991) Cloning and sequencing of cDNA encoding rat GTP cyclohydrolase I - the 1st enzyme of the tetrahydrobiopterin biosynthetic pathway. *J. Biol. Chem.* **266** 765-769.

Haußmann C., Rohdich F., Lottspeich F., Eberhardt S., Scheuring J., Mackamul S. & Bacher A. (1997) Dihydroneopterin triphosphate epimerase of *Escherichia coli*: Purification, genetic cloning, and expression. *J. Bacteriol.* **179** 949-951.

Haußmann C., Rohdich F., Schmidt E., Bacher A. & Richter F. (1998) Biosynthesis of pteridines in Escherichia coli - Structural and mechanistic similarity of dihydroneopterin-triphosphate epimerase and dihydroneopterin aldolase. *J. Biol. Chem.* **273** 17418-17424.

Hennig M., D'Arcy A., Hampele I.C., Page M.G.P., Oefner C. & Dale G.E. (1998) Crystal structure and reaction mechanism of 7,8-dihydroneopterin aldolase from *Staphylococcus aureus. Nature Struc. Biol.* **5** 357-362.

Horrocks P. & Newbold C.I. (2000) Intraerythrocytic polyubiquitin expression in *Plasmodium falciparum* is subjected to developmental and heat-shock control. *Mol. Biochem. Parasitol.* **105** 115-125.

Hyde J.E. (1990) The dihydrofolate reductase-thymidylate synthetase gene in the drug resistance of malaria parasites. *Pharmac. Ther.* **48** 45-59.

Hyde J.E. & Holloway S.P. (1993) Isolation of parasite genes using synthetic oligonucleotides. *Methods Mol. Biol.* **21** 303-18.

Ichinose H., Ohye T., Takahashi E., Seki N., Hori T., Segawa M., Nomura Y., Endo K., Tanaka H., Tsuji S., Fujita K. & Nagatsu T. (1994) Hereditary progressive dystonia with marked diurnal fluctuation caused by mutations in the GTP cyclohydrolase-I gene. *Nature Genetics* **8** 236-242.

Inselburg J. & Banyal H.S. (1984) Synthesis of DNA during the asexual cycle of *Plasmodium falciparum* in culture. *Mol. Biochem. Parasitol.* **10** 79-87.

Jomaa H., Wiesner J., Sanderbrand S., Altincicek B., Weidemeyer C., Hintz M., Turbachova I., Eberl M., Zeidler J., Lichtenthaler H.K., Soldati D. & Beck E. (1999) Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. *Science* **285** 1573-1576.

Jones T.H. & Brown G.M. (1967) The biosynthesis of folic acid. VII. Enzymatic synthesis of pteridines from guanosine triphosphate. *J. Biol. Chem.* **242** 3989-3997.

Jonsson A.P. (2001) Mass spectrometry for protein and peptide characterisation. *Cell. Mol. Life Sci.* **58** 868-884.

Kapust R.B. & Waugh D.S. (1999) *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* **8** 1668-1674.

Kaspers B., Gütlich M., Witter K., Lösch U, Goldberg M. & Ziegler I. (1997) Coordinate induction of tetrahydrobiopterin synthesis and nitric oxide synthase activity in chicken macrophages: Upregulation of GTP-cyclohydrolase I activity. *Comp. Biochem. Physiol. B-Biochem. Mol. Biol.* **117** 209-215.

Katzenmeier G., Schmid C., Kellermann J., Lottspeich F. & Bacher A. (1991) Biosynthesis of tetrahydrofolate - sequence of GTP cyclohydrolase-I from *Escherichia coli. Biol. Chem. Hoppe-Seyler* **372** 991- 997.

Killick-Kendrick R. & Peters W. (eds) (1978) *Rodent malaria.* Academic Press Limited, London, UK.

Klayman D.L. (1985) Qinghaosu (Artemisinin) - an antimalarial drug from China. *Science* **228** 1049-1055.

Kozak M. (1984) Compilation and analysis of sequences upstream from the translational start site in eukaryotic messenger-RNAs. *Nucleic Acids Res.* **12** 857-872.

Krishna S., Waller D.W., Terkuile F., Kwiatkowski D., Crawley J., Craddock C.F.C., Nosten F., Chapman D., Brewster D., Holloway P.A. & White N.J. (1994) Lactic acidosis and hypoglycemia in children with severe malaria - pathophysiological and prognostic significance. *Transactions Royal Soc. Trop. Med. Hyg.* **88** 67-73.

Krungkrai J., Yuthavong Y. & Webster H.K. (1985) Guanosine triphosphate cyclohydrolase in *Plasmodium falciparum* and other plasmodium species. *Mol. Biochem. Parasitol.* **17** 265-276.

Krungkrai J., Webster H.K. & Yuthavong Y. (1989) *De novo* and salvage biosynthesis of pteroylpentaglutamates in the human malaria parasite, *Plasmodium falciparum. Mol. Biochem. Parasitol.* **32** 25-37.

Lacks S.A., Greenberg B. & Lopez P. (1995) A cluster of 4 genes encoding enzymes for 5 steps in the folate biosynthetic pathway of *Streptococcus pneumoniae. J. Bact.* **177** 66-74.

Laemmli U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227** 680-685.

Landau I. & Chabaud A. (1994) *Plasmodium* species infecting *Thamnomys rutilans* - a zoological study. *Adv. Parasitol.* **33** 49-90.

Lanzer M., Wertheimer S.P., DeBruin D. & Ravetch J.V. (1993) *Plasmodium*-control of gene expression in malaria parasites. *Exp. Parasitol.* **77** 121-128.

Lee C.S. (1999) PhD Thesis. Department of Biomolecular Sciences. UMIST, Manchester, UK.

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

Lee S.W., Lee H.W., Chung H.J., Kim Y.A., Kim Y.J., Hahn Y., Chung J.H. & Park Y.S. (1999) Identification of the genes encoding enzymes involved in the early biosynthetic pathway of pteridines in *Synechocystis* sp. PCC 6803. *FEMS Microbiol. Letters* **176** 169-176.

Lopez P. & Lacks S.A. (1993) A bifunctional protein in the folate biosynthetic pathway of *Streptococcus pneumoniae* with dihydroneopterin aldolase and hydroxymethyldihydropterin pyrophosphokinase activities. *J. Bacteriol.* **175** 2214-2220.

Luker K.E., Francis S.E., Gluzman I.Y. & Goldberg D.E. (1996) Kinetic analysis of plasmepsins I and II, aspartic proteases of the *Plasmodium falciparum* digestive vacuole. *Mol. Biochem. Parasitol.* **79** 71-78.

Mancini R., Saracino F., Buscemi G., Fischer M., Schramek N., Bracher A., Bacher A., Gutlich M. & Carbone M.L.A. (1999) Complementation of the *fol2* deletion in *Saccharomyces cerevisiae* by human and *Escherichia coli* genes encoding GTP cyclohydrolase I. *Biochem. Biophys. Res. Comm.* **255** 521-527.

Mathis J.B. & Brown G.M. (1970) The biosynthesis of folic acid. XI. Purification and properties of dihydroneopterin aldolase. *J. Biol. Chem.* **245** 3015-3025.

McCutchan T.F., Delacruz V.F., Lal A.A., Gunderson J.H., Elwood H.J. & Sogin M.L. (1988) Primary sequences of 2 small subunit ribosomal-RNA genes from *Plasmodium falciparum. Mol. Biochem. Parasitol.* **28** 63-68.

Meining W., Bacher A., Bachmann L., Schmid C., Weinkauf S., Huber R. & Nar H. (1995) Elucidation of crystal packing by x-ray-diffraction and freeze-etching electron-microscopy studies on GTP cyclohydrolase-I of *Escherichia coli. J. Mol. Biol.* **253** 208-218.

Milhous W.K., Weatherly N.F., Bowdre J.H. & Desjardins R.E. (1985) *In vitro* activities of and mechanisms of resistance to antifolate antimalarial drugs. *Antimicrob. Agents Chemother.* **27** 525-530.

Minvielle-Sebastia L., Preker P.J. & Keller W. (1994) RNA14 and RNA15 proteins as components of a yeast pre-messenger-RNA 3'-end processing factor. *Science* **266** 1702-1705.

Mountain H.A., Bystrom A.S., Larsen J.T. & Korch C. (1991) 4 major transcriptional responses in the methionine threonine biosynthetic-pathway of *Saccharomyces cerevisiae. Yeast* **7** 781-803.

Mumberg D., Muller R. & Funk M. (1995) Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene* **156** 119-122.

Nar H., Huber R., Meining W., Schmid C., Weinkauf S. & Bacher A. (1995a) Atomic structure of GTP cyclohydrolase I. *Structure* **3** 459-466.

Nar H., Huber R., Auerbach G., Fischer M., Hosl C., Ritz H., Bracher A., Meining W., Eberhardt S. & Bacher A. (1995b) Active site topology and reaction mechanism of GTP cyclohydrolase I. *Proc. Natl. Acad. Sci. USA* **92** 12120-12125.

Nardese V., Gutlich M., Brambilla A. & Carbone M.L.A. (1996) Disruption of the GTP cyclohydrolase I gene in *Saccharomyces cerevisiae. Biochem. Biophys. Res. Comm.* **218** 273-279.

Nichol C.A., Smith G.K. & Duch D.S. (1985) Biosynthesis and metabolism of tetrahydrobiopterin and molybdopterin. *Ann. Rev. Biochem.* **54** 729-764.

Niederwieser A,. Blau N., Wang M., Joller P., Atares M. & Cardesa-Garcia J. (1984) GTP cyclohydrolase I deficiency, a new enzyme defect causing hyperphenylalaninemia with neopterin, biopterin, dopamine, and serotonin deficiencies and muscular hypotonia. *European J. Pediatrics* **141** 208-214.

Ockenhouse C.F., Sun P.F., Lanar D.E., Wellde B.T., Hall B.T., Kester K., Stoute J.A., Magill A., Krzych U., Farley L., Wirtz RA., Sadoff J.C., Kaslow D.C., Kumar S., Church L.W.P., Crutcher J.M., Wizel B., Hoffman S., Lalvani A., Hill A.V.S., Tine J.A., Guito K.P., de Taisne C., Anders R., Horii T., Paoletti E. & Ballou W.R. (1998) Phase I/IIa safety, immunogenicity, and efficacy trial of NYVAC-Pf7, a pox-vectored, multiantigen, multistage vaccine candidate for *Plasmodium falciparum* malaria. *J. Infect. Diseases* **177** 1664-1673.

Opperdoes F.R. (1995) Carbohydrate and energy metabolism in aerobic protozoa. *In: Marr, J.J. and Muller, M. (Eds) Biochemistry and Molecular Biology of Parasites (pp19-32).* Academic Press Limited, London, UK.

Padgett R.A., Grabowski P.J., Konarska M.M., Seiler S. & Sharp P.A. (1986) Splicing of messenger-RNA precursors. *Ann. Rev. Biochem.* **55** 1119-1150.

Panyim S. & Chalkley R. (1971) The molecular weights of vertebrate histones exploiting a modified sodium dodecyl sulfate electrophoretic method. *J. Biol. Chem.* **246** 7557-60.

Pashley T.V., Volpe F., Pudney M., Hyde J.E., Sims P.F.G. & Delves C.J. (1997) Isolation and molecular characterization of the bifunctional hydroxymethyldihydropterin pyrophosphokinase-dihydropteroate synthase gene from *Toxoplasma gondii. Mol. Biochem. Parasitol.* **86** 37-47.

Patarroyo M.E., Romero P., Torres M.L., Clavijo P., Moreno A., Martinez A., Rodriguez R., Guzman F. & Cabezas E. (1987) Induction of protective immunity against experimental infection with malaria using synthetic peptides. *Nature* **328** 629-632.

Pfleiderer W. (1994) Natural pteridine pigments - pigments found in butterflies wings and insect eyes. *Chimia* **48** 488-489.

Ploom T., Haußmann C., Hof P., Steinbacher S., Bacher A., Richardson J. & Huber R. (1999) Crystal structure of 7,8-dihydroneopterin triphosphate epimerase. *Struc. Folding & Design* **7** 509-516.

Pollack Y., Katzen A.L., Spira D.T & Golenser J. (1982) The genome of *Plasmodium falciparum*.1: DNA base composition. *Nucleic Acids Res.* **10** 539-546.

Polson A., Coetzer T., Kruger J., von Maltzahn E. & Vandermerwe K.J. (1985) Improvements in the isolation of IgY from the yolks of eggs laid by immunized hens *Immunol. Investigations* **14** 323-327.

Reche P., Arrebola R., Olmo A., Santi D.V., Gonzalez-Pacanowska D. & Ruiz-Perez L.M. (1994) Cloning and expression of the dihydrofolate reductase-thymidylate synthase gene from *Trypanosoma cruzi. Mol. Biochem. Parasitol.* **65** 247-258.

Reed M.B., Saliba K.J., Caruana S.R., Kirk K. & Cowman A.F. (2000) Pgh1 modulates sensitivity and resistance to multiple antimalarials in *Plasmodium falciparum. Nature* **403** 906-909.

Rich S.M., Light M.C., Hudson R.R. & Ayala F.J. (1998) Malaria's eve: Evidence of a recent population bottleneck throughout the world populations of *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. USA* **95** 4425-4430.

Roos D.S. (1993) Primary structure of the dihydrofolate reductase-thymidylate synthase gene from *Toxoplasma gondii*. *J. Biol. Chem.* **268** 6269-6280.

Rosenthal P.J. & Meshnick S.R. (1996) Hemoglobin catabolism and iron utilization by malaria parasites. *Mol. Biochem. Parasitol.* **83** 131-139.

Roth E.F., Raventos-Suarez C., Perkins M. & Nagel R.L. (1982) Glutathione stability and oxidative stress in *P.falciparum* infection *in vitro* - responses of normal and G6PD deficient cells. *Biochem. Biophys. Res. Comm.* **109** 355-362.

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

Sambrook J., Fritsch E.F. & Maniatis T. (1989) *Molecular Cloning: a Laboratory Manual.* 2nd ed. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York, USA.

Schneider J., Langermans J.A.M., Gilbert S.C., Blanchard T.J., Twigg S., Naitza S., Hannan C.M., Aidoo M., Crisanti A., Robson K.J., Smith G.L., Hill A.V.S. & Thomas A.W. (2001) A prime-boost immunisation regimen using DNA followed by recombinant modified vaccinia virus Ankara induces strong cellular immune responses against the *Plasmodium falciparum* TRAP antigen in chimpanzees. *Vaccine* **19** 4595-4602.

Schofield L. (1991) On the function of repetitive domains in protein antigens of *Plasmodium* and other eukaryotic parasites. *Parasitol. Today* **7** 99-105.

Schramek N., Bracher A. & Bacher A. (2001) Ring opening is not rate-limiting in the GTP cyclohydrolase I reaction. *J. Biol. Chem.* **276** 2622-2626.

Sedegah M., Hedstrom R., Hobart P. & Hoffman S.L. (1994) Protection against malaria by immunization with plasmid DNA encoding circumsporozoite protein. *Proc. Natl. Acad. Sci. USA* **91** 9866-9870.

Shine J. & Dalgarno L. (1974) The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. USA* **71** 1342-1346.

Sibley C.H., Brophy V.H., Cheesman S., Hamilton K.L., Hankins E.G., Wooden J.M. & Kilbey B. (1997) Yeast as a model system to study drugs effective against apicomplexan proteins. *Methods-a Companion to Methods in Enzymology* **13** 190-207.

Sikorski R.S. & Hieter P. (1989) A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae. Genetics* **122** 19-27.

Sirawaraporn W., Sathitkul T., Sirawaraporn R., Yuthavong Y. & Santi D.V. (1997) Antifolate-resistant mutants of *Plasmodium falciparum* dihydrofolate reductase. *Proc. Natl. Acad. Sci. USA* **94** 1124-1129.

Slock J., Stahly D.P., Han C.Y., Six E.W & Crawford I.P. (1990) An apparent *Bacillus subtilis* folic acid biosynthetic operon containing *pab*, an amphibolic *trpG* gene, a 3rd gene required for synthesis of para-aminobenzoic acid, and the dihydropteroate synthase gene. *J. Bacteriol.* **172** 7211-7226.

Srivastava I.K. & Vaidya A.B. (1999) A mechanism for the synergistic antimalarial action of atovaquone and proguanil. *Antimicrob. Agents Chemother.* **43** 1334-1339.

Steinmetz M.O., Pluss C., Christen U., Wolpensinger B., Lustig A., Werner E.R., Wachter H., Engel A., Aebi U., Pfeilschifter J. & Kammerer R.A. (1998) Rat GTP cyclohydrolase I is a homodecameric protein complex containing high-affinity calcium-binding sites. *J. Mol. Biol.* **279** 189-199.

Stoute J.A., Slaoui M., Heppner D.G., Momin P., Kester K.E., Desmons P., Wellde B.T., Garcon N., Krzych U., Marchand M., Ballou W.R. & Cohen J.D. (1997) A preliminary evaluation of a recombinant circumsporozoite protein vaccine against *Plasmodium falciparum* malaria. *New England J. Med.* **336** 86-91.

Trager W. & Jensen J.B. (1976) Human malaria parasites in continuous culture. *Science* **193** 673-675.

Triglia T. & Cowman A.F. (1994) Primary structure and expression of the dihydropteroate synthetase gene of *Plasmodium falciparum. Proc. Natl. Acad. Sci. USA* **91** 7149-7153.

Triglia T., Wang P., Sims P.F.G., Hyde J.E. & Cowman A.F. (1998) Allelic exchange at the endogenous genomic locus in *Plasmodium falciparum* proves the role of dihydropteroate synthase in sulfadoxine-resistant malaria. *EMBO J.* **17** 3807-3815.

Volkman S.K., Barry A.E., Lyons E.J., Nielsen K.M., Thomas S.M., Choi M., Thakore S.S., Day K.P., Wirth D.F. & Hartl D.L. (2001) Recent origin of *Plasmodium falciparum* from a single progenitor. *Science* **293** 482-484.

Volpe F., Dyer M., Scaife J.G., Darby G., Stammers D.K & Delves C.J. (1992) The multifunctional folic acid synthesis FAS gene of *Pneumocystis carinii* appears to encode dihydropteroate synthase and hydroxymethyldihydropterin pyrophosphokinase. *Gene* **112** 213-218.

Volpe F., Ballantine S.P. & Delves C.J. (1993) The multifunctional folic acid synthesis *fas* gene of *Pneumocystis carinii* encodes dihydroneopterin aldolase, hydroxymethyldihydropterin pyrophosphokinase and dihydropteroate synthase. *Eur. J. Biochem.* **216** 449-458.

Volpe F., Ballantine S.P. & Delves C.J. (1995) 2 domains with amino acid sequence similarity are required for dihydroneopterin aldolase function in the multifunctional folic acid synthesis FAS protein of *Pneumocystis carinii*. *Gene* **160** 41-46.

Wahle E. & Ruegsegger U. (1999) 3'-End processing of pre-mRNA in eukaryotes. *FEMS Microbiol. Rev.* **23** 277-295.

Walter R.D. (1991) Folate metabolism as a target for chemotherapy of malaria. *In: Coombs G.H. & North, M.J. (Eds) Biochemical Protozoology (pp560-568).* Taylor & Francis Limited, London, UK.

Wang P., Read M., Sims P.F.G. & Hyde J.E. (1997) Sulfadoxine resistance in the human malaria parasite *Plasmodium falciparum* is determined by mutations in dihydropteroate synthetase and an additional factor associated with folate utilization. *Mol. Microbiol.* **23** 979-986.

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

Watkins W.M., Sixsmith D.G., Chulay J.D. & Spencer H.C. (1985) Antagonism of sulfadoxine and pyrimethamine antimalarial activity *in vitro* by para-aminobenzoic acid, para-aminobenzoylglutamic acid and folic acid. *Mol. Biochem. Parasitol.* **14** 55-61.

Weber J.L. (1987) Analysis of sequences from the extremely A+T rich genome of *Plasmodium falciparum*. *Gene* **52** 103-109.

Weber J.L. (1988) Molecular biology of malaria parasites. *Exp. Parasitol.* **66** 143-170.

Wellems T.E., Walliker D., Smith C.L., Dorosario V.E., Maloy W.L., Howard R.J., Carter R. & McCutchan T.F. (1987) A histidine-rich protein gene marks a linkage group favoured strongly in a genetic cross of *Plasmodium falciparum*. *Cell* **49** 633-642.

Wernsdorfer W.H. & Payne D. (1991) The dynamics of drug resistance in *Plasmodium falciparum. Pharmac. Ther.* **50** 95-121.

White N.J. & Ho M. (1992) The pathophysiology of malaria. *Advances Parasitol.* **31** 83-149.

Wilson R.J.M., Denny P.W., Preiser P.R., Rangachari K., Roberts K., Roy A., Whyte A., Strath M., Moore D.J., Moore P.W. & Williamson D.H. (1996) Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum. J. Mol. Biol.* **261** 155-172.

Wittwer C.T., Ririe K.M., Andrew R.V., David D.A., Gundry R.A. & Balis U.J. The LightCycler(TM) a microvolume multisample fluorimeter with rapid temperature control. *Biotechniques* **22** 176-181.

Wooden J.M., Hartwell L.H., Vasquez B. & Sibley C.H. (1997) Analysis in yeast of antimalaria drugs that target the dihydrofolate reductase of *Plasmodium falciparum. Mol. Biochem. Parasitol.* **85** 25-40.

Xu H.M., Aurora R., Rose G.D. & White R.H. (1999) Identifying two ancient enzymes in Archaea using predicted secondary structure alignment. *Nature Struc. Biol.* **6** 750-754.

Zolg J.W., Macleod A.J., Scaife J.G & Beaudoin R.L. The accumulation of lactic acid and its influence on the growth of *Plasmodium falciparum* in synchronized cultures. *In Vitro J. Tissue Culture Assoc.* **20** 205-215.

APPENDIX I

HOST STRAIN GENOTYPES

E.coli host strains

XL1-Blue

recA1, endA1, gyrA96, thi-1, hsdR17(r_K^-, m_K^+), *supE44, relA1, lac,* [F', $proA^+B^+$, $lacI^qZ\Delta M15$, Tn*10*(Tet^R)]

XL*mutS*

*Δ(mcrA)183, Δ(msrCB-hsdSMR-mrr)173, endA1, supE44, thi-1, gyrA96, relA*1, *lac, mutS::*Tn*10*(Tet^R), [F', $proA^+B^+$, $lacI^qZ\Delta M15$, Tn*5*(Kan^R)]

BL21

E.coli B F^-, *dcm, opmT, hsdS*(r_B^-, m_B^-), *gal*

MSD 2658

F^-, *mcrA, mcrB,* IN(*rrnD-rrnE*)*1,* λ(DE3)

S.cerevisiae host strains

FY1679 Δ*FOL2*::*kanMX4* HP1

*MAT*a, *ura3-52, leu2D1, trp1D63, ΔFOL2*::*kanMX4*

FY1679wt HP1

*MAT*α, *ura3-52, his3D200*

APPENDIX II

SEQUENCES OF *Pf GCH-1* & *Pf* GTCase

The nucleotides and amino acids are numbered on the right. Lower case letters mean 5'/3' non-coding regions. The in-frame start and stop codons are in bold and underlined.

```
                .         .         .         .         .         .
atatatatatatatatatatattatacatttaacttaaaaataatatttttatattat       -61

                .         .         .         .         .         .
ttatacatgtgttaatttttatatttttcattttaatggactggaaatttttttttaaaa       -1

                .         .         .         .         .         .
ATGTATAAATATACGTCAATAAACAAATCTGATAAAATATATGAAACACATAATATGGAA        60
 M  Y  K  Y  T  S  I  N  K  S  D  K  I  Y  E  T  H  N  M  E        20

                .         .         .         .         .         .
GAAAAAAAAAAAAAAGGTAATAATAACAATTTCTCAGGATTATTAAATAATGAAATCGAT        20
 E  K  K  K  K  G  N  N  N  N  F  S  G  L  L  N  N  E  I  D        40

                .         .         .         .         .         .
GATAATAATAAAAAGGAAAAGTTAAAAAATAGTATATCTAAAATGTATAGTAACCATAAG       180
 D  N  N  K  K  E  K  L  K  N  S  I  S  K  M  Y  S  N  H  K        60

                .         .         .         .         .         .
AATAGAGAAAATTTTAACGAATGTGAGAAGGAAGACCTTGTTGTGATAGATGAAAAGGAT       240
 N  R  E  N  F  N  E  C  E  K  E  D  L  V  V  I  D  E  K  D        80

                .         .         .         .         .         .
AATAATAAAAAGAAAAAAAAAAATATGACAAATACATTTGAACAAGATAATAATTATAAT       300
 N  N  K  K  K  K  K  N  M  T  N  T  F  E  Q  D  N  N  Y  N       100

                .         .         .         .         .         .
ATGAATGATAATAAAAGGTTAGGTAGCTTTTTTAAAATAAATGATAAATGTGAATCTATT       360
 M  N  D  N  K  R  L  G  S  F  F  K  I  N  D  K  C  E  S  I       120

                .         .         .         .         .         .
AATGAGAATGTTAATAATATTAATAAACAGTCTCTTAAGGATTCTATACTATTTGACAAT       420
 N  E  N  V  N  N  I  N  K  Q  S  L  K  D  S  I  L  F  D  N       140

                .         .         .         .         .         .
ATAAATGAGAAAGAATATTTTAACGAAACCAAAGAAGAAAATAAGGAAGGAAATAAAAGT       480
 I  N  E  K  E  Y  F  N  E  T  K  E  E  N  K  E  G  N  K  S       160

                .         .         .         .         .         .
AATGATATAGAAAAGATAAATTGTATGAAAGTGAAGAAGAAAACTGTTAAAAAGAATAAA       540
 N  D  I  E  K  I  N  C  M  K  V  K  K  K  T  V  K  K  N  K       180

                .         .         .         .         .         .
AAGAAGATTAATAAAATTATAAACAATAAGAATAAAATATCTAAATCAAATGACATAGAA       600
 K  K  I  N  K  I  I  N  N  K  N  K  I  S  K  S  N  D  I  E       200
```

```
          .         .         .         .         .         .
GAACAAATAATTAATATTAGTAAACATATATATAAAATATTAAATATATCCAAATTACCA      660
 E  Q  I  I  N  I  S  K  H  I  Y  K  I  L  N  I  S  K  L  P      220

          .         .         .         .         .         .
AAATGTGATATATTAAAAAGAACAAATAGAAGATATGCTGAAACATTTTTATATTTAACT      720
 K  C  D  I  L  K  R  T  N  R  R  Y  A  E  T  F  L  Y  L  T      240

          .         .         .         .         .         .
AATGGTTATAATCTGGATATAGAACAAATAATAAAAAGATCTTTATATAAAAGGATGTAT      780
 N  G  Y  N  L  D  I  E  Q  I  I  K  R  S  L  Y  K  R  M  Y      260

          .         .         .         .         .         .
AAAAATAATTCAATAATCAAAGTTACAGGTATACATATATATTCATTATGTAAACACCAT      840
 K  N  N  S  I  I  K  V  T  G  I  H  I  Y  S  L  C  K  H  H      280

          .         .         .         .         .         .
CTTTTACCTTTTGAAGGTACATGTGATATTGAGTATATACCCAATAAATATATTATCGGG      900
 L  L  P  F  E  G  T  C  D  I  E  Y  I  P  N  K  Y  I  I  G      300

          .         .         .         .         .         .
TTATCTAAATTTTCAAGAATAGTTGATGTCTTTTCTAGAAGATTACAATTACAAGAAGAT      960
 L  S  K  F  S  R  I  V  D  V  F  S  R  R  L  Q  L  Q  E  D      320

          .         .         .         .         .         .
TTAACTAACGATATTTGTAACGCTTTAAAAAAATACTTAAAACCATTATATATTAAAGTA     1020
 L  T  N  D  I  C  N  A  L  K  K  Y  L  K  P  L  Y  I  K  V      340

          .         .         .         .         .         .
TCTATTGTAGCTAAACATTTATGTATAAATATGAGGGGAGTTAAAGAGCACGATGCTAAA     1080
 S  I  V  A  K  H  L  C  I  N  M  R  G  V  K  E  H  D  A  K      360

          .         .         .         .         .         .
ACTATAACGTATGCATCTTATAAAGCAGAAAAAGAGAATCCTACAGTTCATTCTTTAAAT     1140
 T  I  T  Y  A  S  Y  K  A  E  K  E  N  P  T  V  H  S  L  N      380

          .         .         .         .         .         .
ATTGACTCTTCTGTGGAAAATTTAAATTAGccatattaaaaggattcaatttaaaaaaaa     1200
 I  D  S  S  V  E  N  L  N  <                                    389

          .         .         .         .         .         .
aaaataaaataaataaaaataatgaataccgtaaaggtttatttataagtatatataaat     1260
```

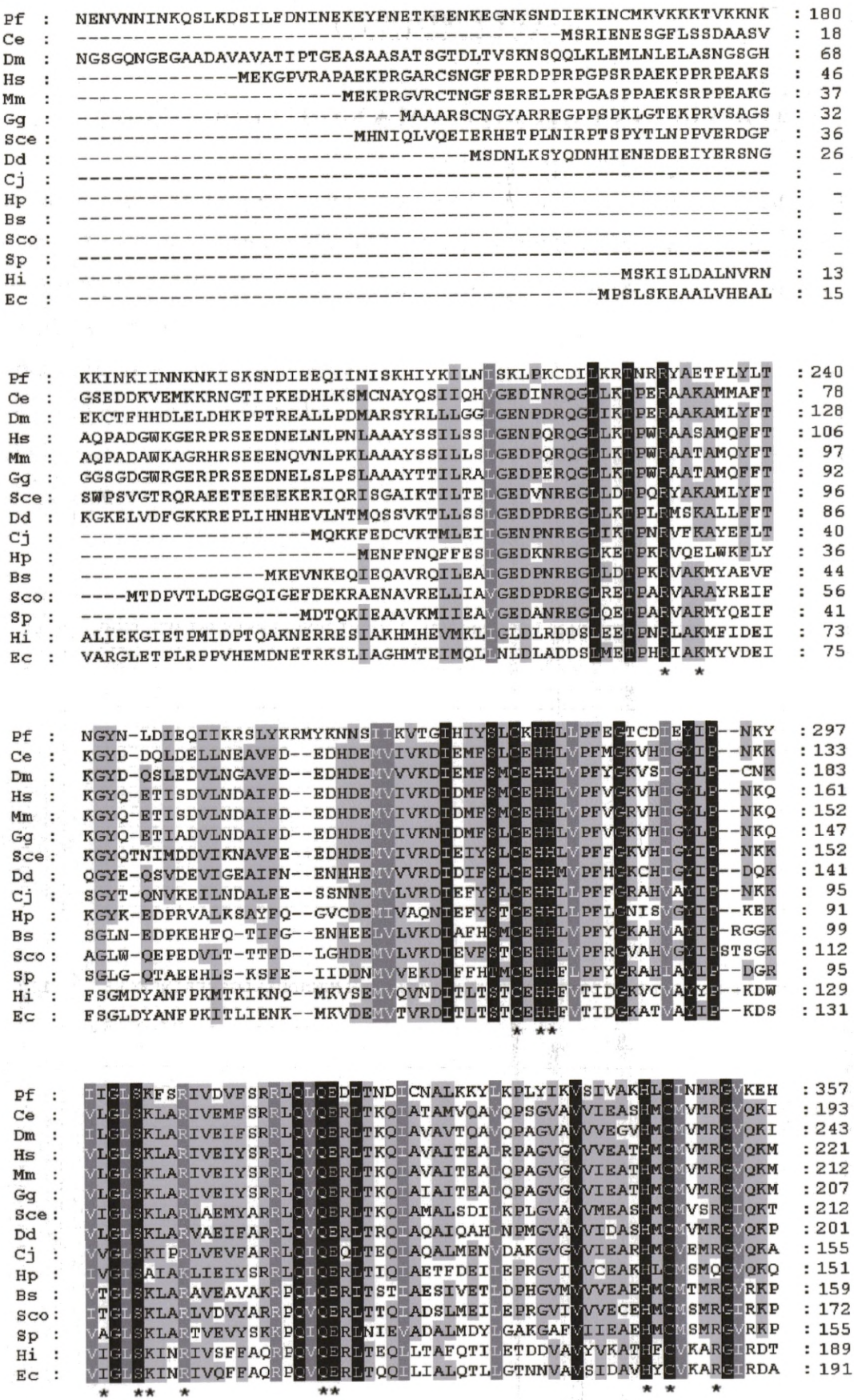
APPENDIX III

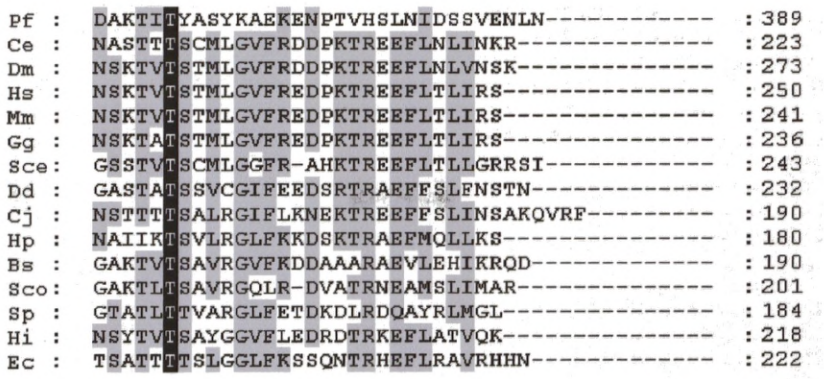
ALIGNMENT OF GTCases FROM A VARIETY OF SPECIES

Residues that are conserved between all species, semi-conserved between all species and conserved or semi-conserved between 60% of the species listed are shaded in black, dark grey and light grey respectively. Those residues marked with asterisks correspond to active site residues in the *E.coli* GTCase as described by Nar *et al.* (1995b). The key is as follows; Pf, *Plasmodium falciparum*; Ce, *Caenorhabditis elegans*; Dm, *Drosophila melanogaster*; Hs, *Homo sapiens*; Mm, *Mus musculus*; Gg, *Gallus gallus*; Sce, *Saccharomyces cerevisiae*; Dd, *Dictyostelium discoideum*; Cj, *Campylobacter jejuni*; Hp, *Helicobacter pylori*; Bs, *Bacillus subtilis*; Sco, *Streptomyces coelicolor*; Sp, *Streptococcus pneumoniae*; Hi, *Haemophilus influenzae*; Ec, *Escherichia coli.*

```
Pf  :  MYKYTSINKSDKIYETHNMEEKKKKGNNNNFSGLLNNEIDDNNKKEKLKNSISKMYSNHK  :  60
Ce  :  ------------------------------------------------------------  :   -
Dm  :  ------------------------------------------------------------  :   -
Hs  :  ------------------------------------------------------------  :   -
Mm  :  ------------------------------------------------------------  :   -
Gg  :  ------------------------------------------------------------  :   -
Sce :  ------------------------------------------------------------  :   -
Dd  :  ------------------------------------------------------------  :   -
Cj  :  ------------------------------------------------------------  :   -
Hp  :  ------------------------------------------------------------  :   -
Bs  :  ------------------------------------------------------------  :   -
Sco :  ------------------------------------------------------------  :   -
Sp  :  ------------------------------------------------------------  :   -
Hi  :  ------------------------------------------------------------  :   -
Ec  :  ------------------------------------------------------------  :   -

Pf  :  NRENFNECEKEDLVVIDEKDNNKKKKKNMTNTFEQDNNYNMNDNKRLGSFFKINDKCESI  : 120
Ce  :  ------------------------------------------------------------  :   -
Dm  :  ----------------------------------------------------MKPQTSEQ  :   8
Hs  :  ------------------------------------------------------------  :   -
Mm  :  ------------------------------------------------------------  :   -
Gg  :  ------------------------------------------------------------  :   -
Sce :  ------------------------------------------------------------  :   -
Dd  :  ------------------------------------------------------------  :   -
Cj  :  ------------------------------------------------------------  :   -
Hp  :  ------------------------------------------------------------  :   -
Bs  :  ------------------------------------------------------------  :   -
Sco :  ------------------------------------------------------------  :   -
Sp  :  ------------------------------------------------------------  :   -
Hi  :  ------------------------------------------------------------  :   -
Ec  :  ------------------------------------------------------------  :   -
```

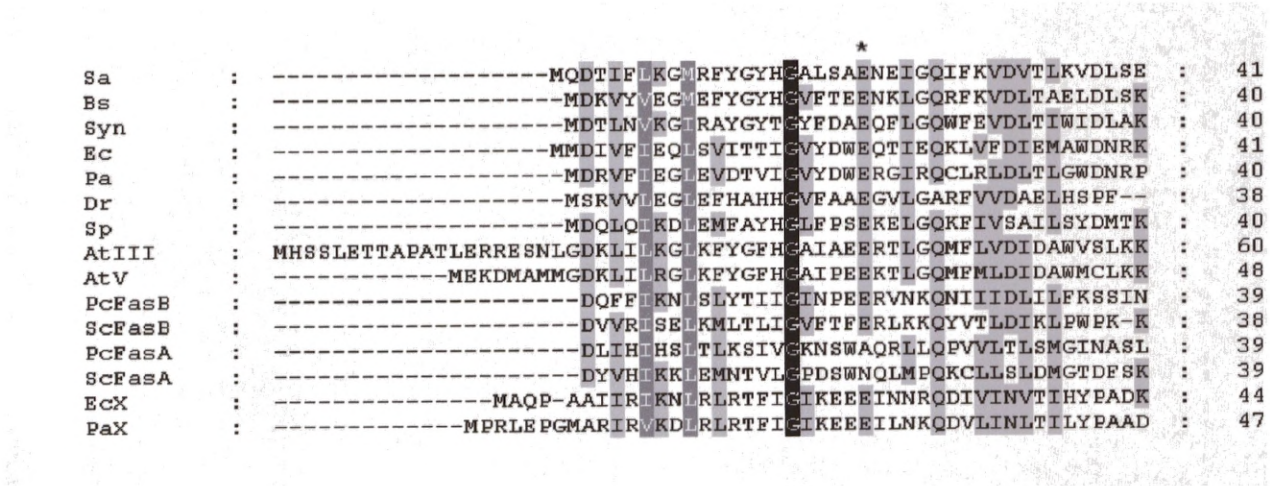


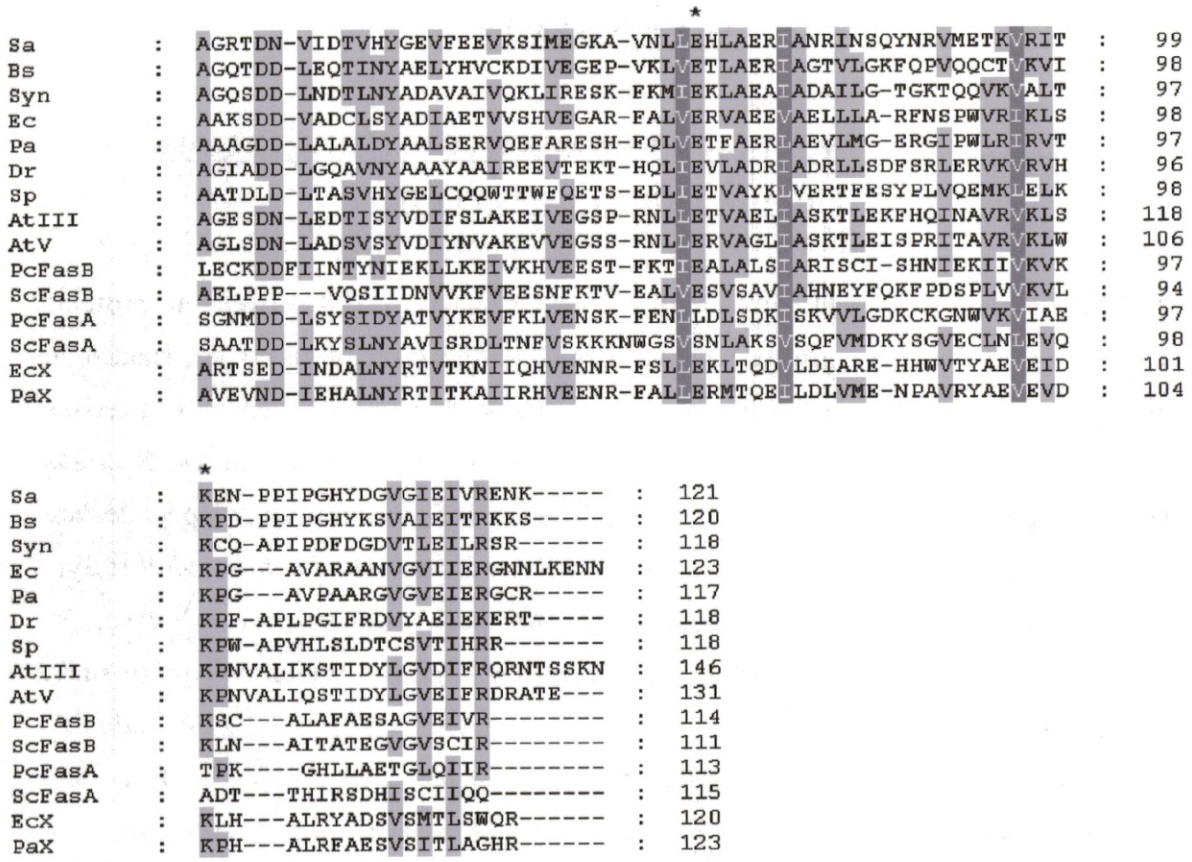

Pf : DAKTITYASYKAEKENPTVHSLNIDSSVENLN---------- :389
Ce : NASTTTSCMLGVFRDDPKTREEFLNLINKR---------- :223
Dm : NSKTVTSTMLGVFRDDPKTREEFLNLVNSK---------- :273
Hs : NSKTVTSTMLGVFREDPKTREEFLTLIRS---------- :250
Mm : NSKTVTSTMLGVFREDPKTREEFLTLIRS---------- :241
Gg : NSKTATSTMLGVFREDPKTREEFLTLIRS---------- :236
Sce: GSSTVTSCMLGGFR-AHKTREEFLTLLGRRSI---------- :243
Dd : GASTATSSVCGIFEEDSRTRAEFFSLFNSTN---------- :232
Cj : NSTTTTSALRGIFLKNEKTREEFFSLINSAKQVRF---------- :190
Hp : NAIIKTSVLRGLFKKDSKTRAEFMQLLKS---------- :180
Bs : GAKTVTSAVRGVFKDDAAARAEVLEHIKRQD---------- :190
Sco: GAKTLTSAVRGQLR-DVATRNEAMSLIMAR---------- :201
Sp : GTATLTTVARGLFETDKDLRDQAYRLMGL---------- :184
Hi : NSYTVTSAYGGVFLEDRDTRKEFLATVQK---------- :218
Ec : TSATTTTSLGGLFKSSQNTRHEFLRAVRHHN---------- :222

APPENDIX IV

ALIGNMENT OF DHNAs & DHNA-RELATED POLYPEPTIDES FROM A VARIETY OF SPECIES

Residues that are conserved between all proteins, semi-conserved between all proteins and conserved or semi-conserved between 60% of the proteins listed are shaded in black, dark grey and light grey respectively. Those residues marked with asterisks correspond to residues that actively participate in substrate binding in the *S.aureus* DHNA as described by Hennig *et al.* (1998). The key is as follows (all polypeptides are DHNAs unless otherwise stated); Sa, *Staphylococcus aureus*; Bs, *Bacillus subtilis*; Syn, *Synechocystis* sp. PCC 6803; Ec, *Escherichia coli*; Pa, *Pseudomonas aeruginosa*; Dr, *Deinococcus radiodurans*; Sp, *Streptococcus pneumoniae*, N-terminal domain of sulD; AtIII, *Arabidopsis thaliana*, chromosome III encoded; AtV, *Arabidopsis thaliana*, chromosome V encoded; PcFasB, *Pneumocystis carinii*, fasB domain of folic acid synthesis protein; ScFasB, *Saccharomyces cerevisiae*, fasB domain of folic acid synthesis protein; PcFasA, *P.carinii*, fasA domain of folic acid synthesis protein; ScFasA, *S.cerevisiae*, fasA domain of folic acid synthesis protein; EcX, *E.coli* dihydroneopterin triphosphate epimerase (DHNTE); PaX, *P.aeruginosa* DHNTE.




*
Sa : AGRTDN-VIDTVHYGEVFEEVKSIMEGKA-VNLLEHLAERLANRINSQYNRVMETKVRIT : 99
Bs : AGQTDD-LEQTINYAELYHVCKDIVEGEP-VKLVETLAERTAGTVLGKFQPVQQCTVKVI : 98
Syn : AGQSDD-LNDTLNYADAVAIVQKLIRESK-FKMIEKLAEATADAILG-TGKTQQVKVALT : 97
Ec : AAKSDD-VADCLSYADIAETVVSHVEGAR-FALVERVAEEVAELLLA-RFNSPWVRIKLS : 98
Pa : AAAGDD-LALALDYAALSERVQEFARESH-FQLVETFAERLAEVLMG-ERGIPWLRIRVT : 97
Dr : AGIADD-LGQAVNYAAAYAAIREEVTEKT-HQLIEVLADRLADRLLSDFSRLERVKVRVH : 96
Sp : AATDLD-LTASVHYGELCQQWTTWFQETS-EDLIETVAYKLVERTFESYPLVQEMKLELK : 98
AtIII : AGESDN-LEDTISYVDIFSLAKEIVEGSP-RNLLETVAELLASKTLEKFHQINAVRVKLS : 118
AtV : AGLSDN-LADSVSYVDIYNVAKEVVEGSS-RNLLERVAGLLASKTLEISPRITAVRVKLW : 106
PcFasB : LECKDDFIINTYNIEKLLKEIVKHVEEST-FKTIEALALSLARISCI-SHNIEKIIVKVK : 97
ScFasB : AELPPP---VQSIIDNVVKFVEESNFKTV-EALVESVSAVLAHNEYFQKFPDSPLVVKVL : 94
PcFasA : SGNMDD-LSYSIDYATVYKEVFKLVENSK-FENLLDLSDKLSKVVLGDKCKGNWVKVIAE : 97
ScFasA : SAATDD-LKYSLNYAVISRDLTNFVSKKKNWGSVSNLAKSVSQFVMDKYSGVECLNLEVQ : 98
EcX : ARTSED-INDALNYRTVTKNIIQHVENNR-FSLLEKLTQDVLDIARE-HHWVTYAEVEID : 101
PaX : AVEVND-IEHALNYRTITKAIIRHVEENR-FALLERMTQELLDLVME-NPAVRYAEVEVD : 104
*
Sa : KEN-PPIPGHYDGVGIEIVRENK----- : 121
Bs : KPD-PPIPGHYKSVAIEITRKKS----- : 120
Syn : KCQ-APIPDFDGDVTLEILRSR------ : 118
Ec : KPG---AVARAANVGVIIERGNNLKENN : 123
Pa : KPG---AVPAARGVGVEIERGCR----- : 117
Dr : KPF-APLPGIFRDVYAEIEKERT----- : 118
Sp : KPW-APVHLSLDTCSVTIHRR------- : 118
AtIII : KPNVALIKSTIDYLGVDIFRQRNTSSKN : 146
AtV : KPNVALIQSTIDYLGVEIFRDRATE--- : 131
PcFasB : KSC---ALAFAESAGVEIVR-------- : 114
ScFasB : KLN---AITATEGVGVSCIR-------- : 111
PcFasA : TPK----GHLLAETGLQIIR-------- : 113
ScFasA : ADT---THIRSDHISCIIQQ-------- : 115
EcX : KLH---ALRYADSVSMTLSWQR------ : 120
PaX : KPH---ALRFAESVSITLAGHR------ : 123

APPENDIX V

OLIGONUCLEOTIDES

All primers are listed 5' to 3'. For nucleotide and amino acid sequence of *Pf gch-1* see Appendix II.

***Pf gch-1* specific primers used e.g. in PCR and sequencing**

Primer cyclo10

TT GAA GGT ACA TGT GAT ATT GAG

(sense primer corresponding to aa residues 284-291)

Primer cyclo11

T AAA GCG TTA CAA ATA TCG TTA G

(antisense primer, reverse complement of aa residues 322-330)

Primer cyclo12

CT AAA TCA AAT GAC ATA GAA GAA C

(sense primer corresponding to aa residues 194-202)

Primer cyclo13

G TTC TTC TAT GTC ATT TGA TTT AG

(antisense primer, reverse complement of aa residues 194-202)

Primer cyclo14

CC ATT ATA TAT TAA AGT ATC TAT TGT AGC

(sense primer corresponding to aa residues 335-344)

Primer cyclo16

G AAT GAT AAT AAA AGG TTA GGT AGC

(sense primer corresponding to aa residues 101-109)

Primer cyclo17

GCT ACC TAA CCT TTT ATT ATC ATT C

(antisense, reverse complement of aa residues 101-109)

Primer cyclo18

A CGT TAT AGT TTT AGC ATC GTG C

(antisense primer, reverse complement of aa residues 356-364)

Primer cyclo19

A ATA TAC GTC AAT AAA CAA ATC TG

(sense primer corresponding to aa residues 3-11)

Primer cyclo20

TT TTA TCA GAT TTG TTT ATT GAC G

(antisense primer, reverse complement of aa residues 5-13)

Primer cyclo21

A AGG TAA TAA TAA CAA TTT CTC AGG

(sense primer corresponding to aa residues 25-32)

Primer cyclo22

G TAT ATC TAA AAT GTA TAG TAA CCA TAA G

(sense primer corresponding to aa residues 51-60)

Primer cyclo23

C AAA TAG TAT AGA ATC CTT AAG AGA CTG

(antisense primer, reverse complement of aa residues 130-139)

Primer cyclo24

AT CCT GAG AAA TTG TTA TTA TTA CC

(antisense primer, reverse complement of aa residues 26-33)

Primer cyclo25

CTT ATG GTT ACT ATA CAT TTT AGA TAT AC

(antisense primer, reverse complement of aa residues 51-60)

Primer cyclo26

T ATA TCA CAT TTT GGT AAT TTG G

(antisense primer, reverse complement of aa residues 217-225)

Primer cyclo27

TTG TTC TAT ATC CAG ATT ATA ACC

(antisense primer, reverse complement of aa residues 242-249)

Primer cyclo28

GT ATA CCT GTA ACT TTG ATT ATT G

(antisense primer, reverse complement of aa residues 264-272)

Primer cyclo29

TAC TCA ATA TCA CAT GTA CCT TC

(antisense primer, reverse complement of aa residues 285-292)

Primer cyclo30

A TAA CCC GAT AAT ATA TTT ATT GG

(antisense primer, reverse complement of aa residues 294-302)

Primer cyclo31

T TGT AAT CTT CTA GAA AAG ACA TC

(antisense primer, reverse complement of aa residues 309-317)

Primers designed for amplification of inserts for pGEX-6P-2 from a *Pf gch-1* cDNA clone in pBluescript

Primer gtpgex1

TT T**GA ATT** CAA ATG TAT AAA TAT ACG T

(sense primer corresponding to aa residues 1-6, plus some upstream non-coding sequence, some bases mismatched to the original sequence (underlined) to give an *Eco*RI site (in bold))

Primer gtpgex2

AT A**GA ATT** CCT AAA TCA AAT GAC ATA GAA G

(sense primer corresponding to aa residues 191-201, some bases mismatched to the original sequence (underlined) to give an *Eco*RI site (in bold))

Primer gtpgex3

TT TTT A**GC GGC CGC** AAC AGT TTT CTT CTT CAC TTT C

(antisense primer, reverse complement of aa residues 169-181, some bases mismatched to the original sequence (underlined) to give a *Not*I site (in bold))

Primer gtpgex4

TC CTT T**GC GGC CGC** CTA ATT TAA ATT TTC CAC AGA AG

(antisense primer, reverse complement of aa residues 383-STOP, plus some downstream non-coding sequence, some bases mismatched to the original sequence (underlined) to give a *Not*I site (in bold))

Primers designed for amplification of inserts for pMAL-c2G from a *Pf gch-1* cDNA clone in pBluescript

Primer gtpmal1

AAA **GGA TCC** AAA TCA AAT GAC ATA GAA GAA

(sense primer corresponding to aa residues 192-201, some bases mismatched to the original sequence (underlined) to give a *Bam*HI site (in bold))

Primer gtpmal2

AAA **GGA TCC** TTA TAT AAA AGG ATG TAT AAA AAT AAT TC

(sense primer corresponding to aa residues 252-264, some bases mismatched to the original sequence (underlined) to give a *Bam*HI site (in bold))

Primer gtpmal3

TT TT**A AGC TTG** CTA ATT TAA ATT TTC CAC AGA AG

(antisense primer, reverse complement of aa residues 383-STOP, plus some downstream non-coding sequence, some bases mismatched to the original sequence (underlined) to give a *Hind*III site (in bold))

Primers designed for mutagenesis of pMET-gch

Primer kpnsrf

AGG GAA CAA AAG CTG GGT **GCC CGG GCC** CCC CTC GAG GTC G

(primer corresponding to the multiple cloning site of pMETplus, some bases mismatched to the original sequence (underlined) to convert a unique *Kpn*I site to an *Srf*I site (in bold))

Primer srfkpn

AGG GAA CAA AAG CTG **GGT ACC** GGG CCC CCC CTC GAG GTC G

(primer corresponding to the modified multiple cloning site of pMETplus single mutants, some bases mismatched to the original sequence (underlined) to convert a unique *Srf*I site to a *Kpn*I site (in bold))

Primer gtpmut1

T ATA AAT GAG AAG GAG TAT TTT AAC GAA ACG AAG GAA GAA AAT AAG G

(sense primer corresponding to aa residues 140-156, some bases mismatched to the original sequence (underlined) to incorporate desired point mutations into *Pf gch-1*)

Primer gtpmut2

T AGT AAA CAT ATC TAC AAG ATC CTA AAC ATA TCC AAA TTA C

(sense primer corresponding to aa residues 206-220, some bases mismatched to the original sequence (underlined) to incorporate desired point mutations into *Pf gch-1*)

Primer gtpmut3

CA AAA TGT GAT ATA CTG AAG AGA ACA AAT AGA AGA TAT G

(sense primer corresponding to aa residues 220-233, some bases mismatched to the original sequence (underlined) to incorporate desired point mutations into *Pf gch-1*)

Primers designed for amplification of the C-terminal insert for pMETplus from pMET-gchmut[3]

Primer gtpmet1

AAG **ACT AGT** ATG TCT AAA TCA AAT GAC ATA GAA G

(sense primer corresponding to aa residues 190-201, some bases mismatched to the original sequence (underlined) to give a *Spe*I site (in bold) and a START codon (highlighted))

Primer gtpmet2

CT **AGA GCT CGT** TGG CTA ATT TAA ATT TTC CAC AGA AG

(antisense primer, reverse complement of aa residues 383-STOP, plus some downstream non-coding sequence, some bases mismatched to the original sequence (underlined) to give a *Sac*I site (in bold))

Primers designed for amplification of inserts for p424GPD from pMET-gchmut[3]

Primer gtpgpd1

ATA **GGA TCC** *ACT AGT* AAA ATG TAT AAA TAT ACG TC

(sense primer corresponding to aa residues 1-6, plus some upstream non-coding sequence, one base mismatched to the original sequence (underlined) to give a *Bam*HI site (in bold.) Amplification products produced by this primer retain a *Spe*I site (italicised) encoded by the original template sequence)

Primer gtpgpd2

AAA **GGA TCC** AAA ATG AAT GAC ATA GAA GAA C

(sense primer corresponding to aa residues 192-202, some bases mismatched to the original sequence (underlined) to give a *Bam*HI site (in bold) and a START codon (highlighted))

Primer gtpgpd3

CT AGA **ACT CGA GGG** CTA ATT TAA ATT TTC CAC AGA AG

(antisense primer, reverse complement of aa residues 383-STOP, plus some downstream non-coding sequence, some bases mismatched to the original sequence (underlined) to give a *Xho*I site (in bold))

Primer gtpgpd4

GC TCT **AGA ATT CGT** TGG CTA ATT TAA ATT TTC CAC AG

(antisense primer, reverse complement of aa residues 384-STOP, plus some downstream non-coding sequence, some bases mismatched to the original sequence (underlined) to give an *Eco*RI site (in bold))

Primers designed for amplification of *FOL2* derived inserts from *S.cerevisiae* genomic DNA

Primer fol2A

C AAA CAT **ACT AGT** ATG CAT AAC ATC CAA TTA GTG C

(sense primer corresponding to aa residues 1-8, plus some upstream non-coding sequence, some bases mismatched to the original sequence (underlined) to give a *Spe*I site (in bold))

Primer fol2B

CGC TTT **GAG CTC** TTT TTT TCA AAT ACT TCT TCT TC

(antisense primer, reverse complement of aa residues 239-STOP, plus some downstream non-coding sequence, some bases mismatched to the original sequence (underlined) to give a *Sac*I site (in bold))

Primer fol2C

CGC TTT TA**T CGA** TTT TTT TCA AAT ACT TCT TCT TC

(antisense primer, reverse complement of aa residues 239-STOP, plus some downstream non-coding sequence, some bases mismatched to the original sequence (underlined) to give a *Cla*I site (in bold))

Primers used in RT-PCR

Primer ssu1

CAT TCG TAT TCA GAT GTC AGA GGT G

(sense primer specific to the *P.falciparum* asexual stage SSU rRNA (McCutchan *et al.,* 1988), designed for gene-specific PCR from of 1st strand cDNA)

Primer ssu2

CGT TCG TTA TCG GAA TTA ACC AGA C

(antisense primer specific to the *P.falciparum* asexual stage SSU rRNA (McCutchan *et al.,* 1988), designed for gene-specific PCR from of 1st strand cDNA)

Primer ssu3

GCT TAC TAG GCA TTC CTC GTT CAA G

(antisense primer specific to the *P.falciparum* asexual stage SSU rRNA (McCutchan *et al.,* 1988), designed to prime synthesis of 1st strand cDNA)

Primer dhfs17

GG TTC ATT CAT AAA AAT GGT ATC TTG C

(antisense primer specific to *Pf dhfs/fpgs* (Lee *et al.,* 2000), reverse complement of part of 3' UTR)

Primer dhfs20

CAC AAT GAA ACG GCA ATA GAT AG

(sense primer specific to *Pf dhfs/fpgs* (Lee *et al.,* 2000), corresponding to gene tract upstream of intron 1)

Primers designed for analysis of the putative *P.yoelii gch-1* gene

Primer PyGTP1

ATA ATA AAA GGG GAA AGA ATA GAC

(sense primer, upstream of possible frameshift in genome database)

Primer PyGTP2

G CAA TAA ATG ATG TTT ACA TAA AG

(antisense primer, downstream of possible frameshift in genome database)

Primer PyGTP3

TTA ACT GAA GGA TAT AAA ATG AGT G

(sense primer corresponding to putative aa residues residues 156-164 (numbering as for full length protein))

Primer PyGTP4

CAT ATA TCA CCC ATT TTC ACA CTT G

(antisense primer, reverse complement of part of putative 3' UTR)

Primer PyGTP5

CCA TAT ATT TCT AAG TAT GCA TGT ATG

(sense primer corresponding to part of putative 5' UTR)

Primer PyGTP6

GGT CAG CTT TTA TTT GTC CAA TTG

(sense primer corresponding to part of putative 5' UTR, binding downstream of PyGTP5)

Primer PyGTP7

TTT TAC ACT CAT TTT ATA TCC TCC AG

(antisense primer, reverse complement of aa residues 157-165 (numbering as for full length protein))

Degenerate primers designed for attempted isolation of the gene encoding DHNA from the *P.falciparum genome*

R = A+G, W = A+T, Y = C+T, M = A+C, K = T+G

Primer dhn1

ATG GAT ATT RTA TTT WTA ARA GGW YT

(degenerate sense primer specific to putative *P.yoelii dhna* gene, corresponding to putative aa residues 1-9, corrected for typical *P.falciparum* codon usage)

Primer dhn2

GTA TTA GAT WTA GAA ATG GCW ACW GA

(degenerate sense primer specific to putative *P.yoelii dhna* gene, corresponding to putative aa residues 29-37, corrected for typical *P.falciparum* codon usage)

Primer dhn3

GC TAY TTT TTC WGC TAM WGT YTC

(degenerate antisense primer specific to putative *P.yoelii dhna* gene, reverse complement of putative aa residues 74-80, corrected for typical *P.falciparum* codon usage)

Primer dhn4

CC WCK TTC WAT TAA WAC WCC WAC

(degenerate antisense primer specific to putative *P.yoelii dhna* gene, reverse complement of putative aa residues 108-115, corrected for typical *P.falciparum* codon usage)

Primers designed for analysis of the putative *P.falciparum dhna* gene

Primer Pfdhn1

C GAT ATG CCA ACT GAG CAA AAT TG

(sense primer corresponding to putative aa residues 38-46)

Primer Pfdhn2

CCA ACT GAG CAA AAT TGT TAC CAG T

(sense primer corresponding to putative aa residues 41-49)

Primer Pfdhn3

TT ATG AAC CAG CTT ATG AAT AAA AC

(antisense primer, reverse complement of putative ORF in downstream EST)

Primer Pfdhn4

TTT TCT TTA AGT TCT TGG ATA TTT C

(antisense primer, reverse complement of putative 3' UTR in downstream EST)

Primer Pfdhn5

CAA TGA CGA TTT TGG TGG GAC CAC

(cDNA specific antisense primer bridging intron *x*)

Primer Pfdhn6

ACG TTG TAA TTT TTT TGC ATT TGT AC

(cDNA specific antisense primer bridging intron *x-1*)

Primer Pfdhn7

CG GGA CAA TTA ATT TTA CCA GAC AC

(sense primer corresponding to putative upstream region of ORF)

Primer Pfdhn8

T GTA TAT CCA GTG ATG TAT TTA ATT C

(sense primer corresponding to putative upstream region of ORF, binding downstream of *Pfdhn7*)

Primer Pfdhn9

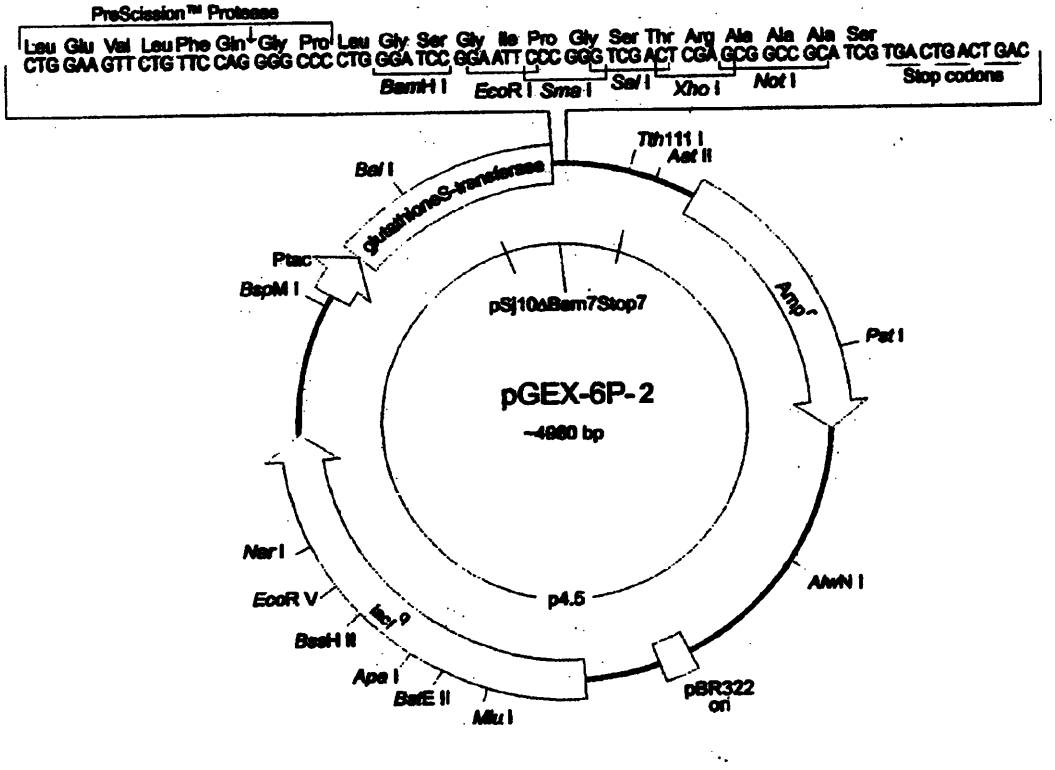
CCA AGA GCG ATA CCG ACA AGT GC

(sense primer corresponding to putative upstream region of ORF, binding downstream of *Pfdhn8*)

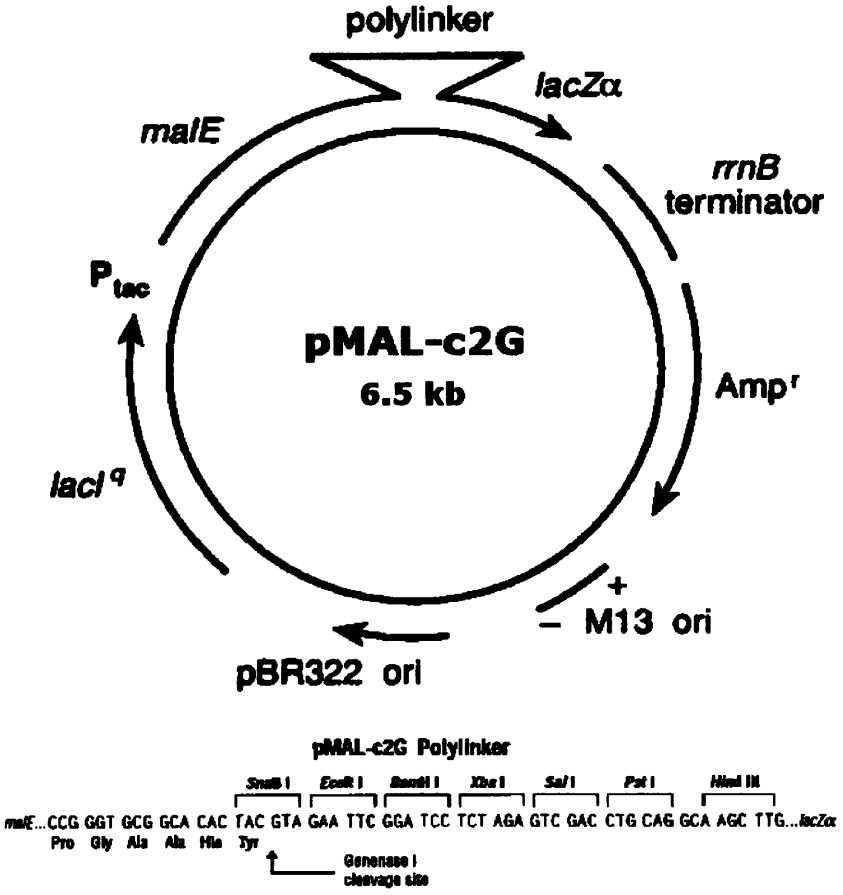
APPENDIX VI

VECTOR MAPS

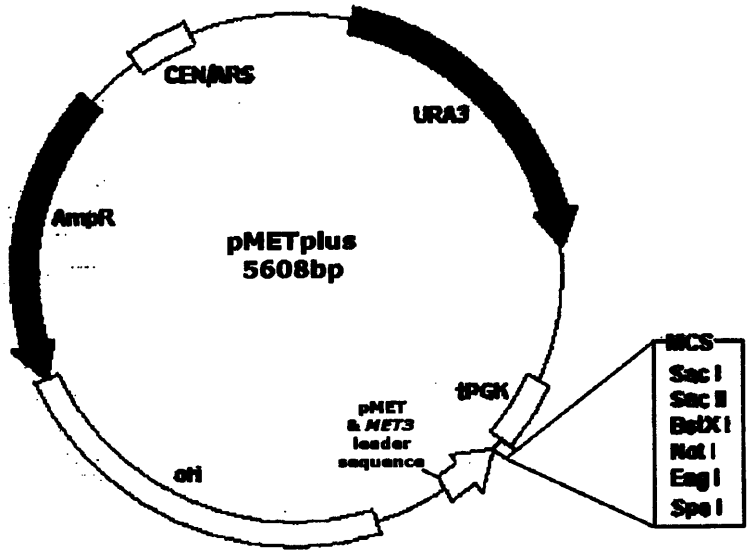
pGEX-6P-2 expression vector (Pharmacia): this vector contains the strong tac promoter for expression in *E.coli* and expresses gene product as a fusion with a cleavable N-terminal glutathione S-transferase (GST) tag allowing rapid detection of expressed product and affinity purification.



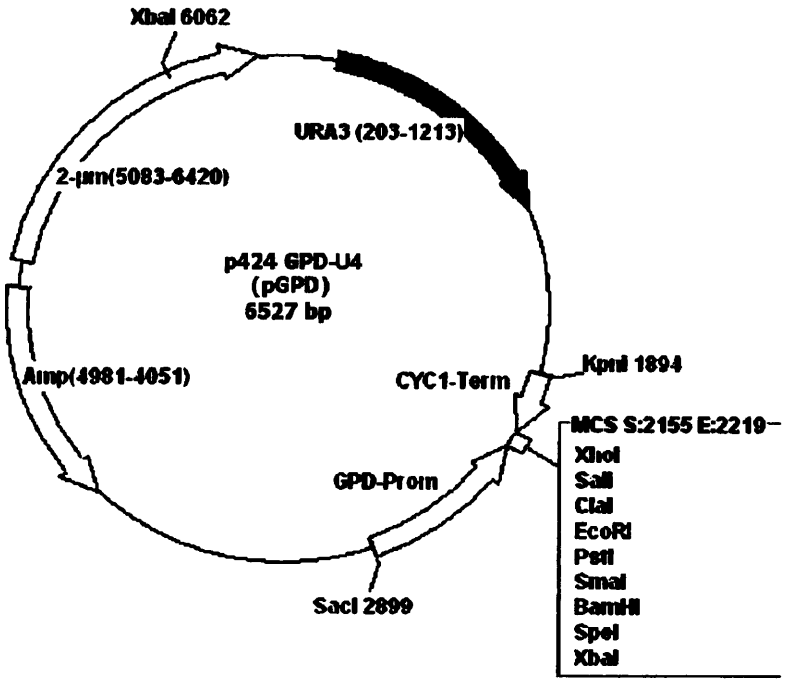
pMAL-c2G expression vector (New England Biolabs): this vector contains the strong tac promoter for expression in *E.coli* and expresses gene product as a fusion with a cleavable N-terminal maltose binding protein (MBP) tag allowing rapid detection of expressed product and affinity purification.



pMETplus expression vector: this is a single-copy modified vector derived from the centromeric *E.coli/S.cerevisiae* shuttle vector pRS306 (Sikorski *et al.,* 1989). The pMET vector, in which the multiple cloning site is flanked by a sequence containing most of the *S.cerevisiae MET3* (methionine transulfurase) promoter (Mountain *et al.,* 1991), was provided by S.G.Oliver (Yeast Genetics Group, UMIST). B.Verrall of the UMIST Molecular Parasitology Group modified this to include the terminator sequence of the *S.cerevisiae PGK* (phosphoglycerate kinase) gene and the leader sequence of *MET3*.



p424 GPD-U4 (pGPD) expression vector: this is a multicopy *E.coli/S.cerevisiae* shuttle vector derived from the pRS series (Sikorski *et al.,* 1989), provided by J.Cortese & C.V.Plowe (University of Maryland). It contains the *S.cerevisiae GPD1* (glyceraldehyde-3-phosphate dehydrogenase) promoter and *CYC1* (cytochrome-c-oxidase) terminator sequences (Mumberg *et al.,* 1995).



pcDNA4/HisMaxA (Invitrogen): this vector contains the strong CMV (human cytomegalovirus) promoter for expression in mammalian cell lines and expresses gene product as a fusion with a cleavable N-terminal tag including a polyhistidine tract allowing affinity purification and the Express epitope for easy detection of expressed product.