Functional, Biochemical and Structural Analyses of two *Plasmodium* Membrane Proteins

by

Amy Marigot Clarke

Thesis submitted for the degree of
Doctor of Philosophy

September 2012

Department of Biology
University of St Andrews

Authors Declaration

1. Candidate's declarations:

I, Amy Marigot Clarke hereby certify that this thesis, which is approximately 70 000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in September 2008 and as a candidate for the degree of Doctor of Philosophy in September 2008; the higher study for which this is a record was carried out in the University of St Andrews between 2008 and 2012.

Date signature of candidate

2. Supervisor's declaration:

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date signature of supervisor

3. Permission for electronic publication:

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright

permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis:

(iii) Embargo on both the printed copy and electronic copy for the same fixed period of two years on the following ground(s):

>publication would preclude future publication;

Date signature of candidate

signature of supervisor

Acknowledgements

"The smallest act of kindness is worth more than the grandest intention." (Oscar Wilde)[1].

Here I have the opportunity to give my thanks to everyone who has contributed to this thesis through their own 'acts of kindness'.

First and foremost I would like to gratefully acknowledge my supervisor Dr. Shiela Unkles for the opportunity to study for a PhD, and for her help on various aspects of work and supervision throughout the four years. I would also like to thank past and present members of the group for their help and support including Jim, Vicki, Eugenia and Naureen. For excellent technical and all round support I'd like to thank Maureen, Lianne, Dave and Harry from the Harold Mitchell Building.

I am very grateful to those who provided me with assistance with laboratory work. A Final Year Honours student, Brian Hoffmann carried out some of the PhoA and GFP assays described in Chapter 5 under my supervision. All of the *Aspergillus* work (Chapter 3), certain nitrite assays and as well as parts of the cloning work presented in this thesis were carried out by Dr. Shiela Unkles.

I wish to thank Prof. Sylke Müller for giving me the opportunity to work in her laboratory on *Plasmodium falciparum*, and for the excellent guidance and encouragement, but most of all for treating me like one of her own. Special thanks must go to Eva, Janet, Ellie and most importantly Andy of the 'Müller lab' who have taught me all I know about *Plasmodium* culture, and for feeding my cells whilst I was away. I am grateful for the general help, support and friendship that was extended to me during my time at the University of Glasgow. For excellent technical support I'd like to thank Anne, Margaret and Mulu from the GBRC.

Thanks also must go to Prof. Mike Barrett and past and present members of the 'Barrett lab', especially Dave, who first inspired me to go down this road. And to Prof.

[1] http://www.goodreads.com/quotes/248952-the-smallest-act-of-kindness-is-worth-more-than-the. Accessed on 9 August 2012.

Lyn-Marie Birkholtz and members of the 'Louw/Birkholtz lab' in South Africa who first introduced me to malaria in the laboratory and who taught me so much in the four months I spent in Pretoria.

An enormous thank you goes to my family for their continual support. I acknowledge that they were always there in times of need. To my parents, Elspeth and Colin (who will always be Mummy and Daddy), I owe a special gratitude. Without their support and understanding it would have been impossible for me to get this far.

Finally I must thank Iain, my love and my best friend. For his encouragement, incredible patience and always being there for me (weather permitting ;). I can't wait for what comes next!

Abstract

Protozoan parasites of the genus *Plasmodium* are the causative agent of malaria. The most severe form of human malaria is caused by *P. falciparum*, responsible for approximately three quarters of a million deaths each year. One major problem in the treatment of malaria is resistance to existing chemotherapies. Consequently, there is an urgent need to identify and validate novel drug targets.

A possible recently identified drug target is the PfNitA protein of *P. falciparum* which contains orthologues in other *Plasmodium* species but is absent from humans. The gene is annotated as a putative formate-nitrite transporter and orthologues are found in a range of prokaryotes as well as the lower eukaryotes algae and fungi. To determine the biological function of the protein, *pfnita* was expressed in *Escherichia coli* strains lacking the endogenous formate and nitrite transporters. In order to analyse the essentiality of the gene a reverse genetics approach was taken and the data discussed. Results indicate that the PfNitA protein is located in the plasma membrane and digestive vacuole of intraerythrocytic parasites suggesting a role in the uptake or excretion of metabolites.

A second complexity with regard to treatment is the lack of a vaccine. A problem in crating a vaccine is antigenic variation. The PIR family of proteins contain a so-called hypervariable domain that has led to the suggestion that the family may play a role in antigenic variation. The objective of the work carried out in this thesis was to investigate the topology and structure of the PcCir2 protein of *Plasmodium chabaudi,* using *E. coli* as the expression host. The topology of Cir2 has been examined by means of reporter fusions and overexpression/purification studies undertaken towards crystallisation. As the PcCir2 amino acid sequence does not show significant homology to other proteins, structural data may provide insights into potential functional or binding domains.

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Overview of Malaria

The genus *Plasmodium* consists of single-celled protozoan parasites, of the Phylum Apicomplexa, which includes the opportunistic human parasites *Toxoplasma gondii* and *Cryptosporidium parvum* and parasites of animals such as *Babesia* spp. and *Theileria* spp. *Plasmodium* infects a range of animals including humans, causing the disease malaria (Roberts and Janovy, 2006). There are five species that give rise to health concerns in humans. Four of these are *Plasmodium falciparum*, *P. vivax*, *P. ovale* and *P. malariae*, while the fifth species, *P. knowlesi*, an Old World monkey malaria causative agent, is known to produce human infections in four Asian countries (IAMAT, 2011[2]). *P. falciparum* is the deadliest of the species and is responsible for the majority of morbidity and mortality caused by malaria. The World Health Organisation (WHO)[3] has estimated that half of the world's population is at risk of infection. Around 243 million of those at risk contract the disease each year and more than three quarters of a million die. The disease is widely spread across the globe, but predominantly affects developing countries (Figure 1.1).

The term malaria was first coined in Italy as *mal aria* and translates to bad air, as the disease was associated with 'bad air' from swamp areas. However, we know now it is not the air from the swamps in which the disease is transmitted, but from the mosquitoes which breed in the stagnant water found in such areas. The transmission of malaria can be increased due to certain factors, for example: climate changes, political unrest and economic conditions (Packard, 2007).

Malarial fevers have long been associated with particular times of the year in endemic countries. In India malaria was referred to as autumnal fever and was especially prevalent after heavy rains. This correlates to the monsoon season that occurs in summer through early autumn, with disease peaking towards the end of the rains. Malaria levels can be affected by other circumstances such as political unrest. Political unrest can see displacement of large populations into refugee camps. Such camps are excellent opportunistic breeding grounds for diseases such as malaria as well as

[2] http://www.iamat.org/pdf/world_malaria_risk_chart.pdf. Accessed on 11 July 2012.
[3] http://www.who.int/malaria/en/. Accessed on 14 February 2012.

cholera and dysentery (Packard, 2007). an inevitable consequence of high density habitation in unsanitary conditions.

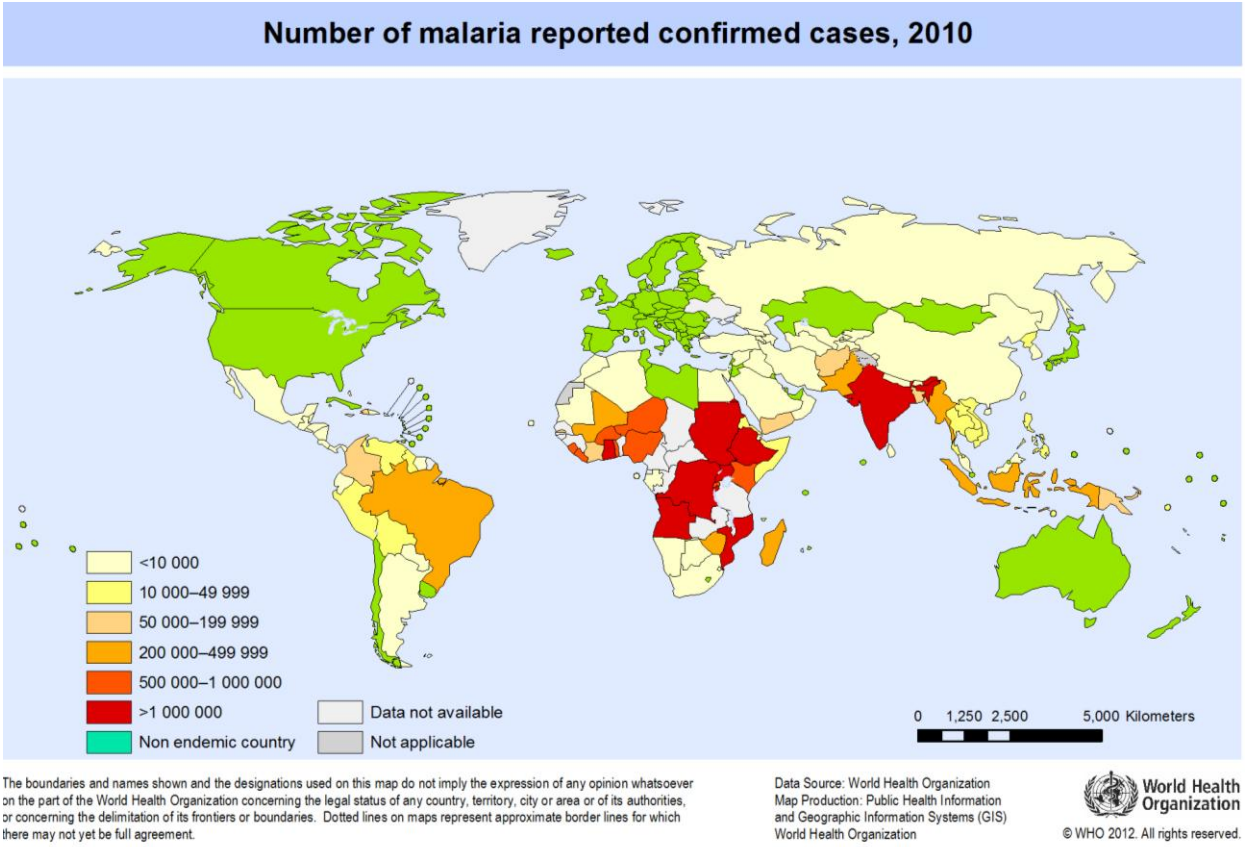


Figure 1.1: Global distribution of malaria. Map displaying the global distribution of malaria in 2010, reproduced from WHO[4]. Regions are highlighted by colours related to the number of cases reported. Areas with no reported data are represented by a pale grey colour and those with no malaria cases are in green.

Malaria typically presents with a periodic fever every 2 or 3 days, the interval depending on the species. Other symptoms may include myalgia, general malaise and headaches. The latter are very mild effects and make it difficult to discern between malaria and, for example, the common cold. If not treated quickly, the infection quickly progresses giving rise to the more serious symptoms of malaria[5]. These include malarial anaemia, lactic acidosis and cerebral infection. The pathogen has evolved ways to avoid the human immune response such as rosetting where infected cells stick to uninfected erythrocytes to avoid removal by the spleen and immune

[4] http://gamapserver.who.int/mapLibrary/Files/Maps/Global_Malaria_ReportedCases_2010.png. Accessed on 11 July 2012.
[5] http://www.who.int/malaria/en/. Accessed on 14 February 2012.

avoidance through antigenic variation (MacKinnon *et al*, 2002). Such immune avoidance mechanisms lengthen the infection that in turn increases the associated morbidity. Clearly malaria is a terrible scourge of humans and a major effort in its elimination is required.

1.2 Life Cycle of Malaria Parasites

It was the collaboration between Sir Patrick Manson, now heralded as the Father of Tropical Medicine, and his young mentee, Sir Ronald Ross, that first elucidated the mode of transmission of malaria. The research work was deemed to be so important that Ross was awarded the Nobel Prize in Physiology and Medicine in 1902 (Ross, 1902). The liver stage of the life cycle was later revealed and much of the research was carried out by Prof. P. C. C. Garnham (Vickerman, 2005). An individual is infected through the bite of an infected female Anopheline mosquito (Figure 1.2, 1). Sporozoites from the mosquito's salivary gland are injected into the skin and make their way to the bloodstream. It has been estimated that 15 – 200 sporozoites are injected with each bite (Vanderberg *et al*, 1977, in Mota and Rodriguez, 2004).

Upon entry into the mammalian host, the sporozoites make use of the circulatory system and are transported to the liver (Figure 1.2, 2). When the parasites reach the liver, they actively force their way through many hepatocytes before coming to rest in a cell where they form a parasitophorus vacuole (PV) and initiate the infection (Silvie *et al*, 2004). The sporozoite undergoes an asexual stage of replication known as schizogony in which up to 30 000 new cells known as merozoites can be produced (Ménard, 2000).

Following liver stage replication, the parasites re-enter the circulation by causing the detachment of the hepatocyte from the liver, followed by the budding of parasite filled vesicles known as merosomes (Sturm *et al*, 2006). The merosome breaks down in the blood and many thousands of merozoites are now free to infect the erythrocytes. Garnhams' work demonstrated the hypnozoite stage of the parasite species *P. ovale* and *P. vivax* (Vickerman, 2005). The hypnozoites form a latent stage of malaria where they lie dormant in liver hepatocytes. These latent parasites are the cause of malaria recrudescence in patients thought to have been cured of the disease.

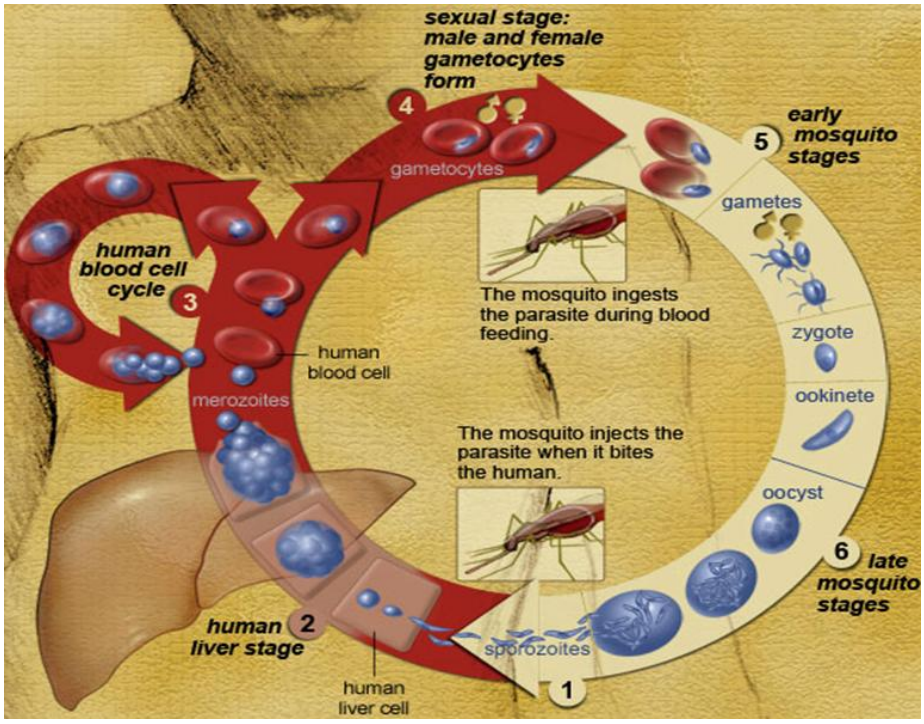


Figure 1.2: Life cycle of the malaria parasite. The figure illustrates the progression of the parasite through the developmental stage in the mosquito, transmission to the human mammalian host and the intraerythrocytic developmental cycle (IDC) which causes the clinical disease malaria. A description of the parasites' life cycle is provided in the following text. Image taken from the National Institutes of Health (USA) website[6].

It takes a malaria parasite approximately 30 seconds (s) to invade a red blood cell (RBC), offering only a limited opportunity for immunological recognition of the merozoite stage parasite. The parasite goes through a stage of intraerythrocytic replication which gives rise to new merozoites (Figure 1.2, 3). The replicative stage sees the parasite start as a ring (due to its distinctive outline) quickly followed by the development of a trophozoite. The trophozoite is particularly metabolically active (Tuteja, 2007). The final stage of the intraerythrocytic developmental cycle (IDC) occurs with replication of the parasites in the schizont stage. The erythrocytic schizogony produces a multinucleate cell (16 – 32 nuclei) followed by division into daughter cells, which engenders the infective merozoites released from the RBCs. The rupture of the RBCs gives rise to the periodic fevers characteristic of malaria infection. Once inside the erythrocytes the parasites cause changes in the RBC membrane (RBCm) (Maier, 2009). Such modifications allow the exchange of materials between the parasite and the extracellular medium. These modifications may also aid the

[6] http://www.niaid.nih.gov/topics/malaria/pages/lifecycle.aspx. Accessed: 11 July 2012.

parasite in its evasion of the host immune system (Sections 1.5 and 1.6, respectively). The parasites may differentiate into male and female gametocytes representing the sexual stage which allows sexual replication in the mosquito (Figure 1.2, 4).

Female macrogametocytes and male microgametocytes are ingested from the human circulation when a female Anopheline mosquito feeds on an infected host. The parasites are ingested with the blood meal into the midgut of the mosquito, in the midgut, the male gametocyte develops into 4-8 flagellated microgametes while the female develops into a spherical macrogamete (Sinden *et al*, 1978). Fertilisation of a macrogamete forms a zygote, which becomes a motile ookinete (Figure 1.2, 5). The ookinete burrows into the midgut epithelium and forms a cyst or oocyst on the basil lamina. Many sporozoites form in the oocyst through asexual multiplication until they are released into the body cavity of the mosquito (Figure 1.2, part 6) (Sinden *et al*, 1978). The sporozoites migrate to the salivary gland of the mosquito where may remain for 1 – 2 months until injection into a new host to continue the cycle (Tuteja, 2007).

1.3 Biochemistry and Cellular Morphology

Although *Plasmodium* species are eukaryotic cells, they contain a number of unusual organelles and biochemical pathways that are absent from the mammalian host. The absence of these pathways in the human cells makes these pathways and organelles attractive as potential novel drug targets. An overview of the organelles present in a trophozoite stage parasite is presented in Figure 1.3.

1.3.1 Apicoplast

The Apicomplexans are the only organisms known to contain an organelle called the apicoplast. The apicoplast is so called from apical complex and chloroplast, and is thought to be a relic of algae formed by secondary endosymbiosis (Lim and McFadden, 2010). The organelle contains its own 35 kb genome coding for 30 proteins, however many of what were once apicoplast genes have been tranferred to the nucleus. Any proteins encoded in the nucleus must be imported to the apicoplast to function. This organelle is surrounded by four membranes and therefore proteins destined for the apicoplast must traverse these membranes. Proteins containing a bipartite signal

extension on the N-terminus are targeted to the apicoplast using a typical chloroplast-entry like peptide that mediates transport (Mullin *et al*, 2006).

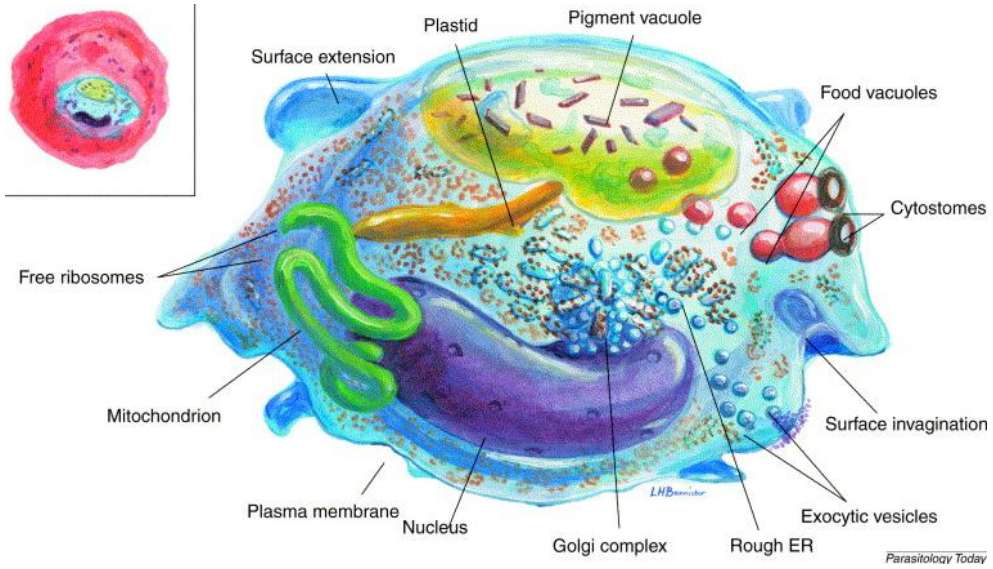


Figure 1.3: Morphology of the malaria parasite. The figure depicts a trophozoite parasite, and has been drawn from electron microscopy images. The image was taken from Bannister *et al* (2000). Of note are the plastid (apicoplast) and pigment vacuole (digestive vacuole) which are discussed below.

1.3.2 Digestive vacuole

The parasite, residing in the RBC acquires certain nutrients by the degradation of haemoglobin (Hb) from the ingested erythrocyte cytoplasm. Between 60 – 80 % of the Hb is utilised (Francis *et al*, 1997). The ingestion of Hb peaks during the late trophozoite and schizont stages (Naughton *et al*, 2010). The degradation of Hb occurs in an acidic vacuole known as the food or digestive vacuole (DV). The Hb is hydrolysed into two molecules, globin and haem. The globin is further degraded into small peptides and amino acids for use by the parasites in protein synthesis. The haem, which is toxic to the parasite, is packaged into haemozoin. The haem detoxification process is unique to malaria parasites. In addition, it is thought that the parasites may gain iron through this process (Rosenthal and Meshnick, 1996). The acidic pH (5.0 – 5.4) of the food vacuole is not low enough to ensure full breakdown of the haemoglobin and therefore various enzymes are also required. A cocktail of proteases including aspartic, cysteine, and metallopeptidases are located in the DV and play a major role in the hydrolysis of Hb (Drew *et al*, 2008).

1.3.3 Carbon metabolism

It is known that there is a high demand for glucose during the intraerythrocytic stages of the parasite (Slavic *et al*, 2010). It was previously thought that the parasite lacked a functional tricarboxylic acid (TCA) cycle in the intraerythrocytic form. The main energy gain was thought to be through glycolysis with lactate being the major end-product. It is now accepted that all the known TCA cycle enzymes are present in the genome and are expressed at these stages (Hall *et al*, 2005). The TCA cycle still needs to be fully characterised in *Plasmodium*, however new information regarding the details of the TCA cycle is becoming apparent, and a novel link between glycolysis and the TCA cycle is emerging (Dr. J. Storm and S. Sethia, personal communication). The proposed carbon source for the TCA cycle is not glucose as in most organisms, but rather glutamine. In this regard glutamine is converted to oxoglutarate in the cytoplasm and is transported into the mitochondrion, possibly through the putative malate:oxoglutarate antiporter (Nozawa *et al*, 2011 and Olszewski and Llinas, 2011).

The TCA cycle in *Plasmodium* is believed to consist of two branches, the oxidative branch and the reductive branch. The reductive branch provides acetyl co-enzyme A (Olszewski *et al*, 2010), while the succinyl co-enzyme A produced by the oxidative branch is used for haem biosynthesis (Ginsburg, 2010). Thus, the main source of ATP is through glycolysis. The electron transport chain (ETC) is critical for the parasites survival; for example, the anti-malarial atovaquone is an analogue of ubiquinone and inhibits the cytochrome bc_1 complex in the ETC (Painter *et al*, 2007). The ETC in asexual parasites is not used to generate ATP. The enzyme dihydroorotate dehydrogenase (DHOD) is utilised in the production of pyrimidines and is essential to asexual parasite survival. Ubiquinone produced by the ETC is the electron acceptor for DHOD and is therefore also vital. Thus the fundamental role of the ETC is in pyrimidine biosynthesis (Painter *et al*, 2007).

Glycolytic intermediates also play a role in the production of ribose-5-phosphate for DNA and RNA synthesis. A certain amount of phospho*enol*pyruvate (PEP), instead of being converted to pyruvate and lactate, is transported to the apicoplast where the parasites' pyruvate dehydogenase complex uses PEP to generate acetyl co-enzyme A (Olszewski and Llinas, 2011). The energy requirements and carbon metabolism during the insect stages of the parasite are unknown. However, in the gametocyte stages the

expression of TCA cycle enzymes is upregulated (Hall *et al*, 2005). The sexual stage parasites use amino acids and glucose present in the haemolymph of the insect as carbon sources.

1.4 Chemotherapy, Vaccines and Vector Control

A major problem in the fight against malaria is the lack of effective and safe chemotherapies (White, 2004). Various chemotherapeutic agents have such severe side effects that people are discouraged from using them. For example the cost effective drug mefloquine (Lariam®), in serious cases, can lead to neuropsychiatric behaviour and cardiac disorders (Lariam® Data Sheet, Roche) (Figure 1.4). The safer drug Malarone® (GlaxoSmithKline) is unaffordable to most of those who require it, and is commonly used by travellers to endemic areas as prophylaxis (Figure 1.4). Contributing to the lack of safe and effective drugs is the increasing parasite resistance to the cost effective chloroquine (CQ) and antifolate drugs currently used (Hyde, 2007). Generally people in malaria endemic areas will suffer from frequent clinical attacks. Most will survive these attacks, however, during such illnesses sufferers are unable to attend school or work.

Inappropriate use of chemotherapy is an enormous problem in endemic countries as many people cannot afford the complete course of a drug or even purchase their drugs on the black market. Even if affordable from illegal sources, quality cannot be assured. Due to the high human morbidity and mortality associated with malaria, and the problems that arise from the misuse of existing antimalarial drugs, affordable novel chemotherapies and drug targets need to be developed.

1.4.1 Currently used antimalarial chemotherapies

Artemisinins

Artemesinin comes from the Chinese medicinal herb *Artemisia annua*. Synthesised derivatives include arteether, artemether and artesunate. Artemisinin is a fast acting drug used to treat malaria and is active against all stages of the intraerythrocytic parasite as well as the transmittable gametocyte. Artemisinin prevents the progression of the disease in an individual and the spread of the disease to others. It is a short acting drug and therefore requires to be given in combination with longer acting drugs

(see below). The mechanism of action is unclear, but the endoperoxide bridge is crucial for the antimalarial activity of artemisinin and its derivatives (Valderramos *et al*, 2010). Possibilities for the mode of action of artemesinins include: (i) interaction of the peroxide with haem leading to the production of free radicals and (ii) the inhibition of protein export and endocytic pathways. The prospect of artemisinins inhibiting PfATP6, a sarco/endoplasmic reticulum Ca^{2+}-ATPase (SERCA) has now been discounted (Valderramos *et al*, 2010). Artemisinin was useful for treating infection with parasites resistant to other drugs, unfortunately resistance is emerging (Krishna *et al*, 2006).

Figure 1.4: Anti-malaria chemotherapy. Mefloquine (Lariam) is a commonly used chemotherapy for prophylaxis of malaria though it can have severe side effects. Malarone® is very an expensive drug with a combination of atovaquone and proguanil. Proguanil (Paludrine) is given in combination with chloroquine (Avloclor) as a prophylaxis. A. Clarke personal photographs.

Quinoline Compounds

Chloroquine (CQ) is a well-known antimalarial compound that targets the blood stages of malarial infection. CQ was first developed as an alternative to quinine. CQ was for decades the first line of defence against malaria but due to widespread acquired resistance its use is diminishing. Notwithstanding, CQ is still included in the WHO Essential Medicines[7] list for use in the treatment and *P. vivax* prophylaxis. However, this classification is to be revised in the near future and so its future status is uncertain. CQ is an alkaline drug and accumulates to high concentrations in the food

[7] http://whqlibdoc.who.int/hq/2011/a95053_eng.pdf. Accessed on 11 July 2012.

vacuole of the parasite. The drug causes an increase in vacuolar pH of the parasite and is thought to prevent the detoxification of haem. The toxic haem accumulates in the cell (Martin *et al*, 2009a). Targets other than haem have also been implicated in having a role in the antimalarial action of quinolines. Derivatives of CQ have been manufactured with the aim of avoiding the acquired resistance observed in parasites.

Other quinoline compounds presently prescribed include amodiaquine which can be used in conjunction with artesunate to treat all malarias, or alone to treat *P. vivax*, *P. malariae* and *P. ovale* infections, as well as primaquine which targets the tissue stages of the disease and is used for *P. vivax* and *P. ovale*[7].

Antibiotics

Currently used antibacterials include tetracycline and its derivative doxycycline. Doxycycline is used for prophylaxis, and can be employed for treatment only in combination with another drug such as quinine[7] as the antibacterial is slow acting. Doxycycline inhibits protein synthesis in bacteria, and is thought to elicit its toxic effects on the parasites' apicoplast and mitochondrion and prevents the formation of new merozoites (Goodman *et al*, 2007).

Antifolates

There are two enzymes in the parasites' folate biosynthesis pathway that are selectively targeted by drugs, namely dihydrofolate reductase (DHFR) and dihydropteroate synthase (DHPS). Pyrimethamine and cycloguanil inhibit *Plasmodium* DHFR. Despite humans also containing a DHFR enzyme the drugs bind to the *Plasmodium* enzyme with much higher affinity. The sulfa drugs, sulfadoxine and dapsone inhibit DHPS, an enzyme not present in mammalian cells. The drugs kill the parasite cells by interfering with DNA synthesis (Hyde, 2007). A combination of two drugs (one that inhibits each enzyme) is used, for example Lapdap which combines chloroproguanil and dapsone, to limit the spread of parasite resistance. Such resistance is now widespread and is caused by mutations in the two enzyme encoding genes (Hyde, 2007). WHO[8] recommends that sulfadoxine and pyremethamine are

[8] http://whqlibdoc.who.int/hq/2011/a95053_eng.pdf. Accessed on 11 July 2012.

given together with artesunate for treatment of malaria. Proguanil is prescribed in combination with chloroquine as a prophylaxis (Figure 1.4).

Combination Therapy

It is anticipated that by combining drugs that target different metabolic steps in the parasite, the development of resistance will be retarded. Artemesinin-based Combination Therapies (ACT) are being produced. WHO[8] advises that many of the drugs used to treat malaria are given in combinations.

Coartem®, an artemether-lumefantrine combination is one such drug. Artemether is a derivative of artemisinin, and lumefantrine is related to chloroquine and quinine (Nosten and White, 2007). The combination of artemether and lumefantrine has been added to the WHO list of Essential Medicines[8] for both adults and children for the treatment of malaria. Another combination drug marketed is Malarone® (also known as Malanil) (Figure 1.4). This contains a combination of atovaquone and proguanil. Atovaquone is an analogue of ubiquinone (part of the ETC) and proguanil inhibits dihydrofolate reductase, the two drugs therefore have actions on different metabolic pathways of the cell (Hyde, 2007 and Painter *et al*, 2007).

1.4.2 Progress towards a vaccine

Humans living in malaria endemic areas become clinically immune to malaria. Whilst clinical immunity protects man from severe forms of the disease, individuals still suffer from the milder symptoms of malaria. Given that clinical immunity can and does occur, a vaccine should be feasible (Holder, 2009). However, it is only after many repeated bouts of the disease that immunity develops. If there is an absence of reinfection the protective immunity will diminish.

The requirement for a vaccine in addition to other methods of control is accurately summarised by Butcher (2007):

> ***"In the absence of protective immunity that a vaccine would provide, the populations of endemic areas would be vulnerable to any new outbreak that might occur"[9].***

The process of antigenic variation (Section 1.6) severely impedes the production of a vaccine against malaria as the parasites are capable of producing a multitude of different antigen types that are exposed to the host's immune system. Vaccine targets should therefore be restricted to proteins that are constitutively expressed throughout the life stage (Phillips, 2001) or a multivalent vaccine combining antigens from each of the life cycle stages.

There are three sections of the parasite lifecycle (Section 1.2) that could be targeted with a vaccine (Hill, 2012): 1. The pre-erythrocytic stage: sporozoites entering the blood are prevented from reaching the liver or completing the liver stage of replication. 2. The asexual merozoite: prevention of the multiplication of the parasite in the blood and a reduction of the symptoms. 3. The sexual stage: vaccines that focus on the sexual stage are known as transmission blocking vaccines. These vaccines ideally would kill or inhibit the parasite in the mosquito vector (Hill, 2012). Transmission blocking vaccines prevent the parasite from being passed on to other people, but would provide no protection to those individuals who take the vaccine. Transmission blocking vaccines may give rise to some ethical issues relating to cost and possible side-effects to a vaccine which will not be of benefit to the one taking the vaccine.

Unfortunately *P. falciparum* does not readily infect laboratory animals. Thus, most of the initial work on vaccines is carried out using the rodent malarias, *P. yoelii*, *P. berghei* and *P. chabaudi*, and the simian malaria *P. knowlesi*. For example, the MSP-1 (merozoite surface protein-1), a possible blood stage vaccine, provided rodents with protective immunity (Phillips, 2001). Experiments were carried out using the *P. falciparum* homologue in monkeys that gave protection (Phillips, 2001). MSP-1 is

[9] Directly quoted from Butcher, (2007).

involved in the invasion process; other proteins involved in the invasion process may make effective vaccine candidates too e.g. AMA-1 (apical membrane antigen-1) (Holder, 2009).

1.4.3 Vector control

"Prevention is better than cure". (Erasmus)[10].

As described in Section 1.2.2, transmission of malaria is through a bite from an infected mosquito depositing the parasites in the human body, thus causing the infection. To prevent the spread of malaria, the mosquito vector may be targeted. There are many courses of action involving the eradication of the mosquito vector.

One approach to limit the number of mosquitoes, although perhaps controversial, is through the use of the insecticides. The most controversial and perhaps the most effective pesticide that was used is dichlorodiphenyltrichloroethane (DDT). Due to environmental issues the use of DDT was banned in many western developed countries, as a result of this legislation, the use in developing countries diminished (Tren and Bate, 2001). Adult mosquitoes can be targeted with insecticides within the home as after a mosquito has taken a blood meal it tends to rest on the walls or the eaves of the house. The insecticides are therefore applied to the walls and eaves of the houses. To prevent the mosquito coming in contact with human individuals, people are advised to sleep under bednets. The use of bednets treated with insecticide has been verified to be more effective than those which have not been treated (Lengeler *et al*, 2006). However, the nets necessitate retreatment after about six months.

Mosquitoes breed in stagnant water. To cover all areas of open water would effectively block off access to the breeding sites but also starve larvae of oxygen (Phillips, 2001). This method is satisfactory for permanent bodies of water but is not effective in the rainy season when transient puddles, pools and floods promptly form. Such bodies of water can be much worse in rural areas where drainage is often poor (Dr. E. Matovu, personal communication) (Figure 1.5).

[10] http://www.brainyquote.com/quotes/quotes/d/desiderius148997.html. Accessed on 1 July 2012.

Figure 1.5: Mosquitoes may breed in puddles formed during the rainy season. A. Clarke personal photographs taken in Kampala, Uganda.

Biocontrol is a less well developed mosquito control measure compared to methods such as insecticide spraying, but is still applicable for vector control. Biocontrol of mosquito populations can be achieved using fungi and bacteria. Some biopesticides have been described as "evolution proof" because of the delayed death phenotype they exhibit (Read *et al*, 2009). An "evolution proof" insecticide is described by Read *et al* (2009) as

> ***"an insecticide that kills after the majority of mosquito reproduction has occurred but before malaria parasites are infectious [to mammalian hosts]."***[11]

Possible "evolution proof" insecticides include the entomopathogenic fungi *Metarhizium anisopliae* and *Beauveria bassiana* (Howard *et al*, 2010). The time taken for the fungi to kill the mosquitoes is slow enough to allow the mosquitoes to reproduce, limiting the selection pressure for resistance to develop. These two fungi were tested for their pathogenicity to pyrethroid sensitive and pyrethroid resistant *Anopheles gambiae* mosquitoes, when applied to bed nets. The study by Howard *et al* (2010) demonstrated that the pyrethroid resistant mosquitoes were significantly more susceptible to infection by the fungi than the pyrethroid sensitive strain. The transformation of fungi strains with the genes of various antimicrobial peptides has shown promise in curing malaria in mosquitoes infected with both malaria and the fungi (Fang *et al*, 2011). The metabolites from pathogenic fungi, such as *Culicinomyces clavisporus*, could be used for IRS or to treat bednets (Singh and

[11] Directly quoted from Read *et al*, (2009).

Prakash, 2011), instead of the chemical insecticides that insects are becoming resistant to and that are deemed to be detrimental to the environment and human health.

Larval control measures kill the vectors in the early stages before the insects have the capacity to transmit diseases. A recent study evaluated the effects of the application of *Bacillus sphaericus* (Bs) and *Bacillus thuringiensis var. israelensis* (Bti) to various aquatic habitats in Kenya (Mwangangi *et al*, 2011). The biweekly treatment with Bs/Bti led to the complete eradication of *Anopheline* larvae from the aquatic test areas. The authors do note, however, that the use of biolarvicides should be used in combination with other methods of control to ensure a decrease in the mosquito populations (Mwangangi *et al*, 2011).

Vector control schemes have been effective in the past e.g. in Brazil. However, it has been demonstrated that complacency may develop resulting in the neglect of these schemes leading to a recurrence of malaria (Packard, 2007). Moreover, the collapse of these initiatives, either though lack of funding or a breakdown of the programme, can lead to an even more devastating effect in the now immunologically naïve populations.

1.5 Transporters

Transport proteins are integral membrane proteins allowing the flow or passage of nutrients across biological membranes, for metabolism and use in cellular functions. Other transporters are capable of removing waste metabolites and xenobiotics including drugs from cells, or translocating them into subcellular compartments. Many drugs use transporter mediated mechanisms of uptake (Dobson and Kell, 2008) and such transporters offer the means to selectively target drugs to cells in which they are expressed, presenting the opportunity to restrict toxic effects to target cells. Consequently it is important to study transporter proteins to improve our knowledge of the binding sites, other important sites and interaction with the substrate and mechanisms associated with transport, so we can identify novel drugs with effective uptake or inhibitors of the transporter.

Membrane permeases can be utilised in two ways when relating to novel chemotherapies. One approach is to block the transporter (e.g. Kang *et al*, 2005),

therefore starving the cell of that particular and perhaps vital nutrient. The other method is to use a transporter to carry a drug into the cell. If one has a detailed understanding of a transporter, specific functional groups can be added to a drug that would enhance the uptake by a specific transporter.

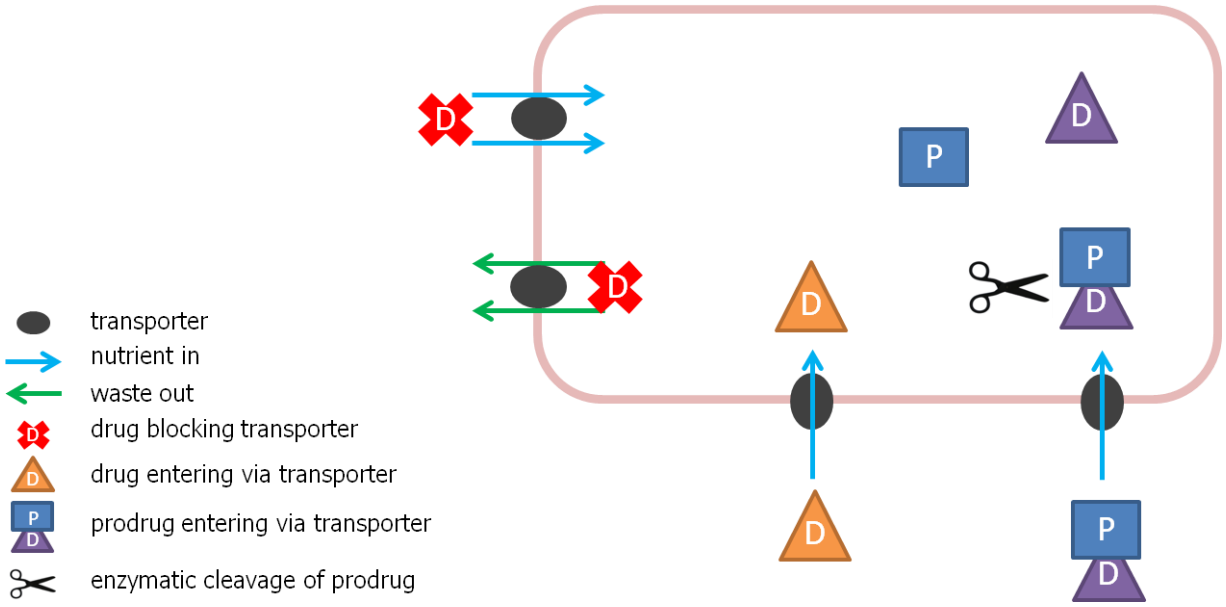


Figure 1.6: Transporters as drug targets. Schematic illustrating the various ways that a compound (D) can be used to target a cell via a transporter. The red cross represents a compound that can be used to block the passage of nutrients into, effectively starving the cell of required compounds (blue arrows). Conversely, a drug may also prevent the efflux of waste products, causing a build up of toxic products (green arrows). A transporter may also be used as a vehicle which delivers a chemotherapeutic compound into a cell either as a drug (orange triangle) or prodrug (blue square) which is cleaved by intracellular enzymes into the active drug (purple triangle). Source: A. Clarke.

Certain drugs that contain added functional groups will enter via a transporter in the form of a prodrug (Kell, 2012). The use of a prodrug relies on some mechanism in the target cell to remove the added motif to convert the drug to the active form. An overview of the means by which transporters can be targeted by chemotherapeutic compounds is illustrated by Figure 1.6.

Initially, the malaria genome project (Gardner *et al*, 2002a) did not identify many putative transporters when compared with the number in other eukaryotes. No ion channels were discovered that displayed significant similarity to other eukaryotic ion channels. 31 % of *Plasmodium* proteins were found to contain at least one putative

transmembrane domain (TMD) but there was a relatively low percentage of polytopic membrane proteins annotated. In fact, only 2.1 % of *Plasmodium* genes are known to code for transporters or channels (Martin *et al*, 2009b). Gardner *et al* (2002a) stated that the low percentage of probable transporter proteins in *P. falciparum* resembles the numbers of transporters present in prokaryotic intracellular organisms. The relative paucity of transport proteins when compared to other organisms implies the transport proteins actually present in the *P. falciparum* genome are important to the organism, perhaps even essential given the likely lack of redundancy.

1.5.1 Permeases targeted in other organisms

Various chemotherapeutic compounds have already been used to target parasite transporters. For example, several drugs used to treat sleeping sickness caused by *Trypanosoma* spp. are carried into the cell by a transporter designated P2 (Barrett and Gilbert, 2006). The transporter recognises an amine moiety present on the anti-trypanosomal agents melarsoprol and pentamidine, used in the late and early stages of trypanosomiasis, respectively. The P2 transporter can therefore be targeted by adding the recognised melamine motif to novel chemotherapeutic compounds (Chollet *et al*, 2009). Ion channels have been successfully targeted by antihelminthic drugs. Ivermectin is a drug that is used to treat onchocerciasis and other nematode infections (Mycek *et al*, 2000). In this regard, ivermectin acts by binding ion channels in the open position, allowing the influx of chloride ions and water. A separate ion channel is targeted in the treatment of schistosomiasis by praziquantel (Mycek *et al*, 2000).

1.5.2 The malaria permeome

As the malaria genome project (Gardner *et al*, 2002a) identified surprisingly few putative transporters based on classical homology searching, a bioinformatic approach using hydropathy plots was taken by Martin *et al*, (2005) to search for further membrane proteins and to assign putative functions to those for which none had been assigned based on simple homology searching. By combining the genome data and computer programs that analyse the amino acid sequence using hydropathy plots, the number of putative transporters and channels recognised in the genome was vastly increased. Most of the membrane proteins discovered by Martin and colleagues (Martin *et al*, 2005) were thought to be transporters, while others were designated as receptors or proteins without a transport function. Of the proteins assigned a

transporter function, some had no amino acid sequence similarity to characterised transporters and therefore such *Plasmodium* proteins could not be given a putative function. Certain transporters have been extensively studied in *Plasmodium* and the characteristics of these are discussed below.

1.5.2.1 Folate transport

Several of the drugs currently used to treat malaria infections act on two enzymes DHFR and DHPS that are involved in the folate synthesis pathway. Forearmed with the knowledge that the folate pathway can be clinically targeted it is important to investigate other aspects of this pathway such as salvage. In this regard Wang *et al* (2007) found that 5-MeTHF (the most abundant form of folate in the serum) was taken up significantly during the trophozoite stage, but poorly during the ring and schizont stages. The influx of folate into a strain resistant to anti-folates was much more rapid than a sensitive strain, demonstrating the folate salvage pathway is of further importance to the development of resistant parasites (Wang *et al*, 2007).

Two *P. falciparum* folate transporters, PfFT1 and PfFT2 have been described by Salcedo-Sora *et al*, (2011). The authors used fluorescence microscopy and report that both transporters were localised to the parasite plasma membrane (PPM) at all stages of the IDC, though intraparasitic staining was also observed. Both PfFT1 and PfFT2 proteins were expressed heterologously in *Xenopus laevis* oocytes or *E. coli* to analyse substrates. Both proteins transported folate and folate precursors. 5-MeTHF was poorly transported by PfFT2 and not by PfFT1 (Salcedo-Sora *et al*, 2011).

1.5.2.2 Nucleoside and nucleobase transport

The protein PfNT1/PfENT1 was the first nucleoside transporter described in malaria parasites (Carter *et al*, 2000 and Parker *et al*, 2000). This permease is an attractive drug target as malaria parasites are unable to synthesise purines *de novo* and therefore rely on uptake from external sources. The human orthologue is sufficiently different adding to the attractiveness of PfNT1 as a drug target. With the publication of the malaria genome (Gardner *et al*, 2002a) a total of four putative nucleoside/base transporters was revealed. PfNT1 is expressed on the PPM. The substrate of the transporter was investigated by expression in *Xenopus* oocytes. However the two

research groups studying PfNT1 came to different conclusions (Carter *et al*, 2000 and Parker *et al*, 2000).

Later studies showed that the transporter could not be knocked out (el Bissati *et al*, 2006 and el Bissati *et al*, 2008) without the addition of high concentrations of hypoxanthine to the medium. The second transporter, PfNT2 was studied by Downie *et al* (2010), the substrate specificity was not determined but the transporter was expressed in the endoplasmic reticulum (ER) membrane, which the authors believe is suggestive of a novel function in the parasites' ER.

1.5.2.3 Glucose transport

Arguably, the best characterised transporter in *P. falciparum* is the hexose transporter (PfHT1). The PfHT1 transporter was first described by Woodrow *et al* (1999). Glucose is essential to the developing intraerythrocytic parasite. The transport of glucose increases by 100-fold in infected RBCs compared with uninfected. Malaria parasites metabolise glucose anaerobically by way of glycolysis for energy production (Section 1.3.3), signifying a high requirement for glucose, especially in the latter stages of the erythrocytic cycle.

The PfHT1 transporter was expressed in *Xenopus* oocytes and shown to transport glucose in this heterologous system. The protein is expressed at high levels during the early ring and early trophozoite stages in *Plasmodium*. Immunofluorescence studies locate the transporter to the PPM confirming the transporter is not exported to the erythrocyte (Woodrow *et al*, 1999). PfHT1 permease can transport fructose and glucose unlike the mammalian glucose transporter, GLUT1 which is only able to transport glucose. The transport of fructose by PfHT1 can be abolished by a single amino acid mutational change (Q196N). The inference is that this mutation is responsible for the substrate specificity of PfHT1 (Joët and Krishna, 2004).

A library of inhibitors that blocked the transport of glucose by PfHT1 expressed in *Xenopus* oocytes, were also able to kill *P. falciparum* and *P. vivax*. A 2010 study by Slavic and colleagues (Slavic *et al*, 2010) was designed to knock out the *pfht1* gene in *P. falciparum*. The researchers found that without an episomal copy of the *pfht1* gene, the native copy could not be deleted, demonstrating that PfHT1 is essential to asexual

stages of *P. falciparum*. Moreover, the *P. berghei* orthologue, PbHT, could not be knocked out in blood stage parasites. A PbHT-GFP fusion protein was made and fluorescence studies confirmed the PPM location of the transporter but also hinted at a DV location. The transporter is expressed throughout the asexual and sexual growth stages (Slavic *et al*, 2010).

1.5.2.4 Glycerol and water transport

The intraerythrocytic parasite has a high demand for glycerol which is required for membrane biogenesis during the later stages of the growth phase as up to 32 new merozoites are made. Glycerol is present in blood plasma and is able to cross the erythrocyte membrane by means of passive diffusion although it is unknown if the alterations made to the RBCm assist with glycerol uptake. A single aquaglyceroporin has been annotated in the *Plasmodium* genome, PfAQP (Hansen *et al*, 2002) and PbAQP (Promeneur *et al*, 2007). Hansen and colleagues (2002) carried out BLAST searches of the plasmoDB database which yielded a single potential aquaporin located on chromosome 11 of the parasite. The *P. berghei* orthologue was discovered by doing a BLAST search of PfAQP in the *P. berghei* genome database (Promeneur *et al*, 2007). The AQP shows most homology to the glycerol facilitator of *E. coli*, and the orthologue from *P. berghei* has 62 % conservation with the *P. falciparum* gene.

Research was carried out on the *P. berghei* and *P. falciparum* orthologues. The protein is responsible for the passage of water and glycerol into the parasite, as well as small solutes such as urea. The *Plasmodium* proteins are the only known family of aquaglyceroporins to transport water and glycerol (in *Plasmodium*) (Newby *et al*, 2008). Utilising fluorescent microscopy techniques, the proteins were localised to the PPM. The *P. berghei* gene was knocked out (Promeneur *et al*, 2007) and it is noted that it is one of the few *Plasmodium* transporters capable of being knocked out. In addition, the authors did not observe morphological changes in the Δ*pbaqp* strains. However, the growth rate of the Δ*pbaqp* parasites was greatly reduced when compared with WT parasites. This implies there is some glycerol gained during glycolysis.

Hedfalk *et al* (2008) cloned the *P. falciparum* aquaporin for functional studies. The crystals produced were of many forms making it hard to gain a 3D crystal structure.

The crystal structure was then solved by Newby *et al* (2008). The latter research group used a codon bias to express the *pfaqp* gene in *E. coli*. This was the first time a full-length malarial membrane protein had been crystallised.

Two motifs common to aquaglyceroporins are present in the *Plasmodium* orthologue. The two, three amino acid motifs differ in one amino acid (each) compared to other orthologues, however when comparing the structural data, the differences were compensated for and the overall structure was preserved (Newby *et al*, 2009). Indeed, mutation of the *Plasmodium* sequence motifs (NLA-NPS) to the conserved motifs (NPA-NPA) does not alter the substrate specificity of the pore. Taken together, these data suggest these motifs do not play a role in substrate specificity (Hansen *et al*, 2002).

1.5.2.5 Apicoplast transporters

There are many transporters present on the apicoplast membranes (Martin *et al*, 2009b). Mullin *et al* (2006) characterised two transporters that display sequence identity with two plant plastid transporters. These transporters designated PfiTPT and PfoTPT were localised to the inner and outer membranes of the apicoplast, respectively. The substrate of these two plastid transporters could not be determined, but they are thought to exchange phosphate for carbon compounds (Mullin *et al*, 2006).

1.5.2.6 Transporters involved in drug resistance

Different transporters are involved in uptake and efflux of individual drugs. This means that resistance can come about through a variety of alterations in different transporters creating heterogeneity in mechanisms of resistance found in laboratory strains used in different studies (e.g. Dd2, 7G8 and HB3). The three transporters implicated in drug resistance are PfCRT (*P. falciparum* chloroquine resistance transporter), PfMDR1 (*P. falciparum* multi drug resistance 1), and PfMRP (*P. falciparum* multi resistance-associated protein). These will be briefly discussed below. On the other hand, the role of a fourth transporter, PfNHE (*P. falciparum* Na^+/H^+ exporter), in drug resistance is unclear (e.g. Bray *et al*, 1999, and, Roepe and Ferdig, 2009).

Thus far, only PfCRT has been experimentally shown to confer resistance to CQ. The resistance is thought to be due to a lysine to threonine mutational alteration (K76T).

The K76T mutation is always observed with some other mutations such as R163S and A220S. In the majority of field isolates the latter mutation was present (Cooper, 2005). It has been proposed the other mutations e.g. R163S and A220S are compensatory, allowing the transporter to carry out its natural but as yet unknown function. When the K76T mutation was reversed in a CQ resistant strain the parasites were susceptible to CQ (Lakshmanan, 2005).

The PfCRT protein was expressed, without putative signal peptides and to a codon bias, in *Xenopus* oocytes, and without these modifications the protein was not expressed at the oocyte surface (Martin *et al*, 2009a). This 2009 study demonstrated that (i) the transport of CQ occurs from a CQ resistant strain, and (ii) no transport occurs in the chloroquine sensitive strain. The resulting data suggests that the mutant PfCRT is used to efflux CQ from the DV to the cytoplasm where the drug cannot exert its toxic effect (Martin *et al*, 2009a). Although the role of PfCRT is yet unknown, the transporter is thought to be necessary for parasite survival as it cannot be knocked out (Waller *et al*, 2003).

PfMDR1 has been shown to have a role in drug resistance, but the mechanism (s) has not been definitively demonstrated. Polymorphisms in these permeases lead to the transport of different antimalarials e.g. quinine and mefloquine (See Section 1.4.1). This polyspecificity implies the transporter can act on various chemotherapies. However the manner in which the transporter acts upon the drugs is unknown (Sanchez *et al*, 2010). When PfMRP was knocked out, the strain displayed a heightened sensitivity to antimalarials such as CQ and artemisinin, and had an accumulation of glutathione (Raj *et al*, 2008). It has been inferred that PfMRP acts as an efflux pump.

1.5.2.7 New permeability pathways

Infected erythrocytes (IE) show an increase in their permeability to various solutes including sugars, anions and cations. This new-found permeability has been ascribed to the so called new permeability pathways (NPPs) induced by the parasite. It is not known if the NPP's are formed from parasite proteins, the host cell proteins or a combination of both. The NPP's allows the passage of a wide variety of substrates into the RBC including sugars, anions and cations. The parasitophorus vacuole membrane

(PVM) is porous and allows the passage of the substrates to reach the PPM. The proposed roles of the NPP include nutrient uptake and end-product removal as well as putative roles in cell regulation (Staines *et al*, 2004).

1.5.3 Formate-nitrite transporter family

Members of the formate-nitrite transporter (FNT) family (transport protein classification 2.A.44) are present in a wide range of prokaryotes and some eukaryotic organisms but are absent from plant and mammalian genomes. As the name suggests, the transporters are thought to facilitate the passage of the structurally similar anions formate and nitrite across membranes. The FNT proteins typically have 6 – 8 TMDs and are between 250 – 400 amino acids in length. Three conserved FNT motifs are present in each orthologous protein. The members of the FNT family have been grouped into 4 clusters (Saier *et al*, 1999). Cluster I contains the formate efflux transporters, e.g. *E. coli* FocA, cluster II the formate uptake transporters e.g. *Methanobacterium formicicum* and *Methanothermobacter thermautotrophicus*, cluster III the nitrite permeases e.g. *E. coli* NirC. Cluster IV contains the *YHL800c* gene from *Saccharomyces cerevisiae*, a member of the transporter family but annotated as an acetate:proton symporter (Paulsen *et al*, 1998). A BLAST[12] (protein-protein BLASTP) search of the *S. cerevisiae* YHL800C protein sequence against the Non-redundant protein sequences (nr) database unearthed orthologues in the important human pathogen *Candida albicans* and other yeast (Appendix 5). *P. falciparum* contains a putative FNT of 309 amino acids. By performing BLASTP searches using the *P. falciparum* predicted protein sequence against other *Plasmodium* spp. and apicomplexan organisms (Appendix 5) it was established that there are orthologues in all malaria species sequenced, and apicomplexan parasites other than *C. parvum*. The *P. falciparum* orthologue PFC0725c (PlasmoDB[13]) has been designated PfNitA.

1.5.3.1 Roles of FNT orthologues

E. coli

The FocA transporter in *E. coli*, which is thought to maintain the intracellular formate pool, belongs to the FNT family of transporters. The protein has been shown to transport the toxic formate analogue hypophosphite, through this process, mutants

[12] http://blast.ncbi.nlm.nih.gov/Blast.cgi. Accessed on 12 July 2012.
[13] http://plasmodb.org/plasmo/. Accessed on 12 July 2012.

unable to transport hypophosphite were selected. The resulting mutations caused an increase in the levels of intracellular formate, and concomitantly less formate was excreted to the growth media (Suppmann and Sawers, 1994).

Formate is an end-product of the anaerobic fermentation of glucose by pyruvate-formate lyase, up to 33 % of the glucose metabolised ends up as formate (Falke *et al*, 2010). To prevent acidification of the cytoplasm by formate it is exported from the cytoplasm where it is metabolised to carbon dioxide by formate dehydrogenase in the periplasm (Suppmann and Sawers, 1994). Conversely, when the pH of the medium becomes acidic (< pH 6.8), formate is transported into the cytoplasm to be metabolised into hydrogen and carbon dioxide by the enzyme formate-hydrogen lyase (Rossman *et al*, 1991 and Sawers, 2005). The closest homologue of FocA is a second putative formate transporter, FocB (Andrews, 1997). The formate transporter proteins can be surmised as playing a role in the cellular redox balance (Sawers, 2005).

A third member of the FNT family is present in *E. coli*, encoded by the *nirc* gene, the product of which is a transporter for nitrite alone, although not the only transporter of nitrite (Jia *et al*, 2009). NirC functions in the import and export of nitrite. The uptake of nitrite by NirC is 10-fold higher than that of NarK and NarU, two nitrite/nitrate transporters (Jia *et al*, 2009). The three proteins form part of the nitrate assimilatory pathway to produce ammonium for incorporation into amino acids, mainly glutamine and glutamate. NarU and NarK proteins are thought to be antiporters and efflux nitrite in exchange for nitrate. *E. coli* NirC re-imports nitrite to the cytoplasm for reduction to ammonium, and subsequent incorporation into amino acids. NirC is expressed only under anaerobic conditions (Clegg *et al*, 2002).

Salmonella

The genome of the enteropathogenic bacterium *Salmonella enterica* (serovar Typhimurium) contains a *nirc* orthologue. It has been suggested that the transporter in this bacterium plays a role in the pathogenicity of the organism (Das *et al,* 2009). The production of nitric oxide (NO) from macrophages was downregulated in wild type (WT) bacteria compared to knock-out *nirc* parasites in the mouse model. The expression of the transporter was upregulated when cells were exposed to nitrite (Das *et al,* 2009). The inference is that the NirC transporter is used to defend the bacterium

from oxidative stress. The NirC protein from *S. typhimurium* was overexpressed in *E. coli* and reconstituted into liposomes (Rycovska *et al*, 2012). The liposomes containing StmNirC (proteoliposomes) were used for electrophysical characterisation to study the transport activities of the protein. The study found that StmNirC acts as a nitrite/proton antiporter, and is also capable of transporting nitrate (Rycovska *et al*, 2012).

S. typhimurium also encodes a FocA protein. StFocA was shown to passively transport formate at physiological pH values. However this transport was abolished at low pH when the protein underwent structural changes (Lü *et al*, 2011). In a more recent study (Lü *et al*, 2012) the FocA permease was found to transport a range of metabolic end-products from mixed acid metabolism. These metabolites included lactate, formate and acetate. In addition to these metabolic end-products, the permease was able to transport pyruvate, nitrite, hypophosphite (*E. coli* section above) and chloride. The authors believe that the FNT proteins "should be defined as broad-range monovalent anion channels" and that the transportation of products is dependent on metabolic processes in the organism (Lü *et al*, 2012).

Clostridia spp.

The FNT orthologue from the anaerobic bacterium *Clostridium botulinum* was highlighted by *in silico* studies as a potential drug target (Koteswara Reddy *et al*, 2010). Czyzewski and Wang (2012) carried out a genetic linkage and phylogenetic analysis on 474 bacterial and archaeal orthologues.

The group mapped the homologous proteins depending on metabolic enzymes present in the study organism. The linkage analysis (Figure 1.7) shows the *foca* genes split into two groups, one with a pyruvate-formate lyase operon (pink -*pflBA*) and a second with a formate dehydrogenase operon (pink -*fdhAB*). The *nirc* genes are grouped with a nitrite operon (blue), and the novel hydrosulphide ion transporters (FNT3) linked to the sulphite reductase operon (green). The areas in grey have no function based on the genetic linkage analysis suggesting these orthologues transport different anionic compounds.

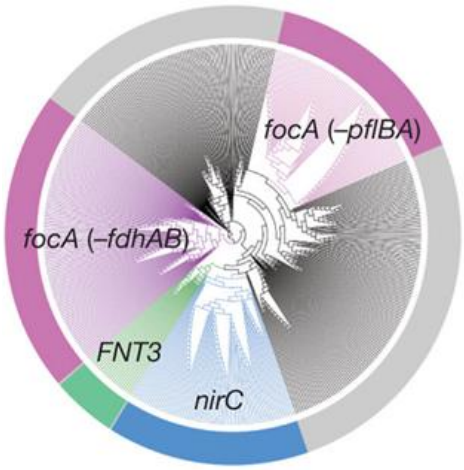


Figure 1.7: Genetic linkage analyses of FNT orthologues. Image reproduced from Czyzewski and Wang (2012).

Electroconductance assays were carried out on the protein orthologue from *C. difficile* in order to determine the substrate (Czyzewski and Wang 2012). The hydrosulphide ion, in quantities sufficient for analysis, causes perturbations in the membrane. Therefore Czyzewski and Wang (2012) used formate as a substrate with the hydrosulphide ion and nitrite as competitors. The hydrosulphide ion channel (HSC) protein was able to transport each of the three substrates demonstrating polyspecificity.

Archaea

Methanogens are formate metabolising archaea. Formate is used as an electron donor in the reduction of carbon dioxide to methane (Schauer *et al*, 1982). *Methanobacterium formicicum* contain two formate dehydrogenase enzymes, and these are co-transcribed with a FNT orthologue gene, *fdhc*. The FdhC protein is 280 amino acids in length and contains 7 putative TMDs. A BLAST[14] search using the FdhC amino acid sequence produced a NirC orthologue from *S. typhimurium* (White and Ferry, 1992). As these archaebacteria can grow on formate as the sole carbon source, it is thought that FdhC is a formate uptake transporter.

[14] http://blast.ncbi.nlm.nih.gov/Blast.cgi. Acessed on 11 July 2012.

Algae

The green algae *Chlamydomonas reinhardtii* has a set of 6 genes coding for FNTs. Three of these may be targeted to the plastid membrane and three to the plasma membrane (Fernandez and Galvan, 2007). The protein encoded by Nar1.1 plays a role in the transport of nitrite into the plastid by regulation of uptake, and perhaps the transport of nitrite *per se*. The uptake of nitrite by chloroplasts of Nar1.1 deletion strains was much less than in the WT cells (Rexach *et al*, 2000). A second FNT gene, Nar1.2 was investigated by Mariscal *et al* (2006). This study demonstrated that the expression of Nar1.1 and Nar1.6 were induced by nitrite, low CO_2 induced Nar1.2 expression. Nar1.2 was also given the identifier *lciA* for low carbon induced. For this reason the group investigated the possibility of bicarbonate as a substrate for the transporter (Mariscal *et al*, 2006). When Nar1.2 was expressed in *Xenopus* oocytes, the uptake of bicarbonate was measured and found to be two-fold higher than in the control oocytes. Nar1.2 transports nitrite at low concentrations as measured by membrane depolarisation (Mariscal *et al*, 2006).

Euglena gracilis

In a study by Deloménie *et al*, (2007) investigating the response to environmental toxins, the FNT protein from the unicellular protist *Euglena gracilis* (EgFth) was upregulated at the mRNA level in cell lines chronically exposed to the toxin cadmium. Cell lines newly exposed to cadmium showed no immediate changes. The implication is that the EgFth protein functions to combat the toxic effects of heavy metals such as cadmium. Conversely however, cells with chronic exposure to the highly toxic pesticide pentachlorophenol showed a down-regulation in *egfth* expression (Deloménie *et al*, 2007).

Yeast

The YHL008c protein of *Saccharomyces cerevisiae* belongs to the FNT family and was annotated as a monocarboxylate transporter (MCT) as it resembles bacterial formate–nitrite transporters (Paulsen *et al,* 1998, in Makuc *et al*, 2002). The proteins' function as a MCT was investigated by Makuc *et al* (2001). The YHL008c protein was found not to be involved in monocarboxylate transport. A study into chloride (Cl^-) homeostasis demonstrated that the uptake of Cl^- was not prevented by deletion of the gene, but was reduced. The authors suggest a role in a Cl^- sensing and activation of a Cl^-

transporter for the YHL008c protein (Makuc *et al*, 2001). A subcellular localisation has not been accurately determined for the YHL008c protein. Nonetheless, studies using GFP tagged proteins suggest it is associated with the vacuole (Paulsen *et al,* 1998, in Jennings and Cui, 2008). The vacuole of *S. cerevisiae* and other yeast has many roles including the storage of solutes and the isolation and degradation of substances harmful to the cell.

Fungi

Aspergillus nidulans contains two transporters for nitrate and nitrite, and a single transporter (NitA) for nitrite uptake (Wang *et al*, 2008 and references therein). The uptake of nitrite was investigated in a mutant strain of *A. nidulans* lacking the two nitrate/nitrite transporter activities. This double mutant strain transported nitrite at a lower rate than the WT cells, however residual uptake was sufficient to sustain growth on medium containing nitrite as the sole nitrogen source. A triple mutant was created which included a mutant in the *nita* gene and the triple mutant was found to be unable to grow on nitrite, unless a high concentration of nitrite was available in the growth media. NitA was thus characterised as a low capacity high affinity nitrite transporter (Wang *et al*, 2008). Formate (formic acid) was assessed as a substrate for *A. nidulans*, but formate was not transported (Dr. S.E. Unkles, personal communication).

Mutagenesis studies have been carried out to establish which amino acids are responsible for substrate binding of the NitA protein to nitrite. The study began by selecting four asparagine residues (N122, N173, N214 and N246) that were highly conserved and present in the TMDs (Symington, 2009). Of these residues, two were unable to be investigated as they prevented growth of the organism. Of the other two one was not directly involved in nitrite transport. It was postulated that this amino acid could constrict the space for nitrite transport in the pore and it was suggested that this residue functions as part of a gate or barrier. The final asparagine residue (N214) is thought to be involved in the binding of nitrite as mutation allowed uptake of only 20 % nitrite when compared to the WT NitA (Symington, 2009 and Unkles *et al*, 2012).

The pathogenic plant fungus *Cylindrocarpon tonkinense* contains an FNT orthologue which is proposed to play a role in the process of denitrification (Kim *et al*, 2010). Denitrification is a poorly understood process in fungi. Under denitrifying conditions,

the FNT orthologue was strongly expressed (Kim *et al*, 2010). This indicates a role for the FNT orthologue in the denitrification process of *C. tonkinense* possibly through the import of formate as an electron acceptor or the efflux of formate produced as a toxic by-product.

1.5.3.2 Structural studies

The FNT orthologues from *A. nidulans* (NitA) and the thermophilic archaebacterium *Thermofilum pendens* (NirC) were expressed in *Pichia pastoris* and *E. coli*, respectively (Beckham *et al*, 2010). The *E. coli* FocA was over expressed in *E. coli* (Falke *et al*, 2009). The quaternary structures of the proteins were studied. Circular dichroism spectrophotometry (CD) provided evidence that all three FNT proteins have a high α-helical content as suggested by the *in silico* predictions. Through size exclusion chromatography the *A. nidulans* NitA was shown to be tetrameric (Beckham *et al*, 2010). Unpublished work using blue native polyacrylamide gel electrophoresis (BN-PAGE) also suggests that *T. pendens* NirC is tetrameric (Dr. S.E. Unkles, personal communication). The *E. coli* FocA protein was confirmed to be pentameric through BN-PAGE and western blotting (Falke *et al*, 2009). The FNT proteins from four bacteria have been crystallised and are given in Table 1.1.

Bacteria strain	pH	Resolution	Reference	PDB NUMBER
Escherichia coli	7.5	2.25 Å	Wang *et al*, 2009a	3KCV
Vibrio cholerae	7.5	2.13 Å	Waight *et al*, 2010	3KLY
Salmonella typhimurium	4.0	2.80 Å	Lü *et al*, 2011	3Q7K
Clostridium difficile	4.5, 7.5, 9.0	3.0, 3.2, 2.2 Å	Czyzewski and Wang, 2012	3TDP, 3TDR, 3TDO

Table 1.1: FNT orthologues with known crystal structures.

Each of the studies confirmed the presence of six TMDs and a cytoplasmic location for the N- and C-termini of the proteins. TMDs 1 – 3 and 4 – 6 were also noted as being 'inverted repeats', while not completely symmetrical. TMDs 2 and 5 are broken into two helices (Figure 1.8). These two TMDs are constrained through hydrogen bonding (Wang *et al*, 2009a). The protein forms a pentamer in the membrane. The overall structure of the FNT protein is thought to resemble that of aquaporins and glycerol facilitators.

The large central pore formed by the five monomers is too large to act as a selectivity filter for a channel or contain a substrate binding site for a transporter (Waight *et al*,

2010). The size of the pore suggests the central pore does not facilitate the transport of formate, and indeed the central pore is filled with lipids from the bilayer. The pore formed by each individual monomer is believed to be the location where the substrate passes through the membrane and is illustrated by Figure 1.8. The protein is suggested to be a channel rather than a transporter as no gate-like structure is present.

(A) (B)

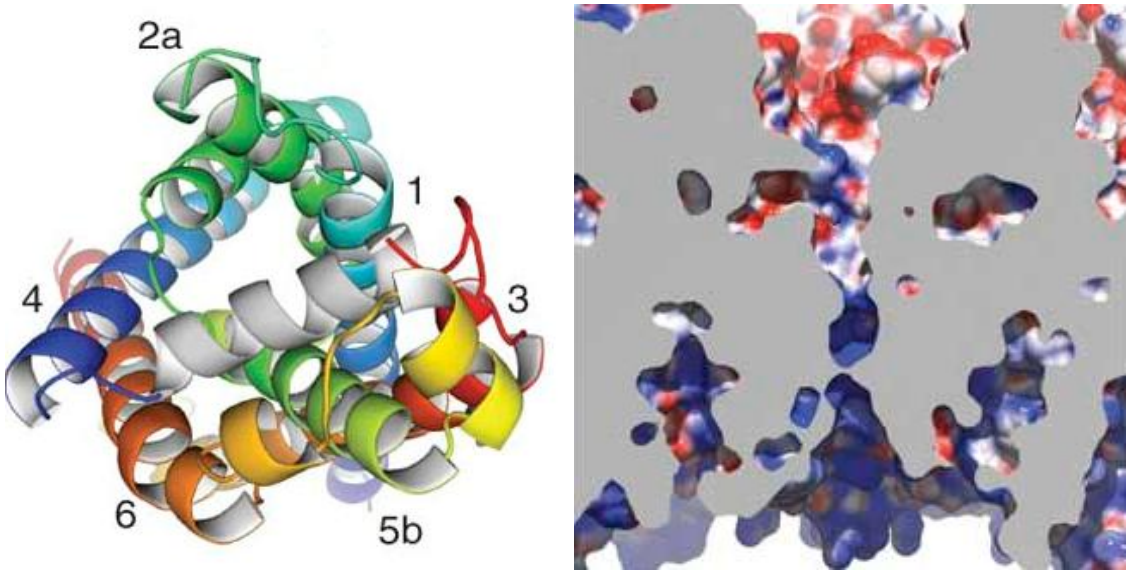


Figure 1.8: 3D Model of FocA monomer. (A) a model of the FocA crystal structure as viewed from the periplasmic side illustrating the central pore where the substrate (formate) passes through. The numbers relate to the transmembrane domains. Parts 2a and 5b are α-helices which are split in two. (b) shows the electrostatic properties of the channel. Red represents a negative charge and is on the periplasmic side, whereas blue is positive and is the cytoplasmic side. Figures taken from Waight *et al* (2010).

The overall charges at the cytoplasmic and periplasmic ends of the pore are positive and negative, respectively. The positive charge of the cytoplasmic side is considered to aid the collection of formate molecules and the negative periplasmic side to repel the positively charged formate (Figure 1.8). The passage of the substrate is facilitated by hydrogen bonding through residues lining the pore, notably His 208 which is conserved between orthologues, hydrogen bonds are also formed from one formate molecule to another (Waight *et al*, 2010). Waight *et al* (2010) hypothesised a role for the so called Ω loop (*V. cholerae* Phe 89 – Thr 134), a flexible and conserved amino acid sequence, in the gating of the protein. Formate must compete with a weaker H

bond formed between the Asn 171 and Thr 90, the displacement of the bonding opens the channel. This point is illustrated in Figure 1.9. Waight *et al* (2010) revealed a water molecule bound to Thr 90 and postulated that water was also a substrate for the *V. cholerae* transporter. Work on proteoliposomes with the *E. coli* FocA protein demonstrated that the transporter was impermeable to water, but allowed the passage of formate (Wang *et al*, 2009a).

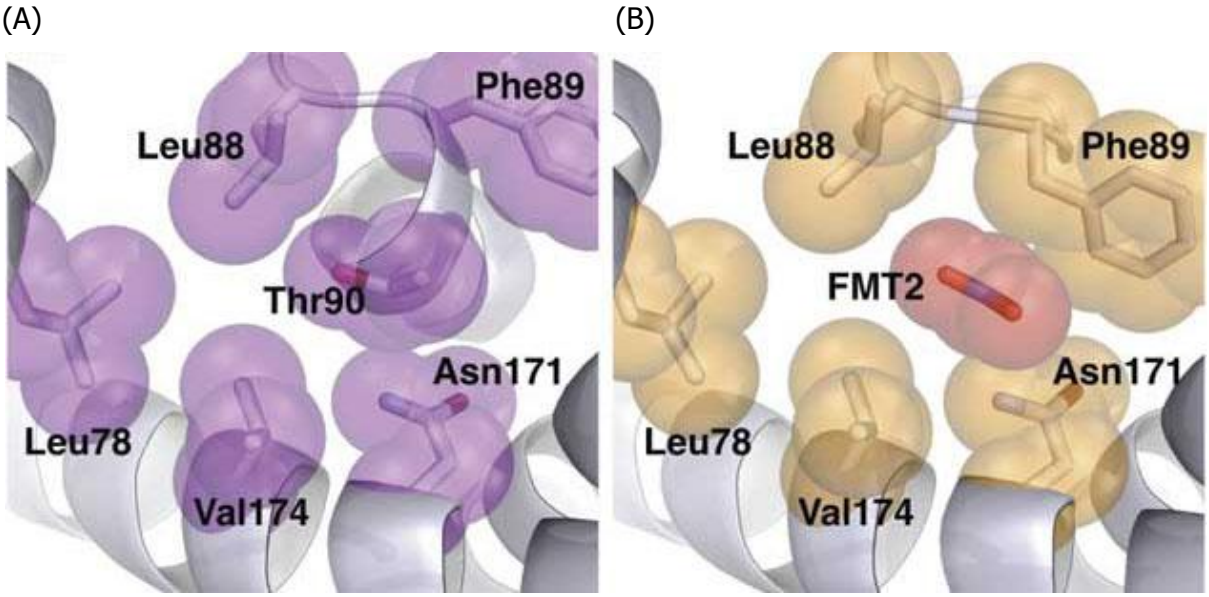


Figure 1.9: FocA selectivity filter. The images here demonstrate the selectivity filter in the central pore of the FocA monomer. (A) shows the pore without formate and is in the closed state. The image in (B) includes a formate (FMT) molecule bound in the pore. FMT2 is hydrogen bonded to His 208, and to a second formate molecule (not visible). To open the pore, a hydrogen bond occurring between Asn 171 and Thr 90 must be broken and replaced by a formate molecule. Images have been reproduced from Waight *et al* (2010).

More recently the structures of the *Salmonella typhimurium* (Lü *et al*, 2011) and *Clostridium difficile* FNT proteins have become available (Czyzewski, and Wang, 2012). Both structures are similar to the *E. coli* and *V. cholerae* structures. The authors demonstrate that the *S. typhimurium* transporter acts in the uptake as well as the export of formate and that the process is dependent on pH. The N-terminal amino acids in StFocA, which in VcFocA were disordered and in EcFocA were truncated, are flexible and allow conformational changes in the protein. The suggested Ω loop gating mechanism was investigated, and Lü *et al* (2011) believe that an interaction between the N-terminal and the Ω loop is responsible for conformational changes. The 2011

study results showed the protein crystallised in three conformational states, open, intermediate and closed (Lü *et al*, 2011). Lü *et al*, (2012) show that the StFocA protein is permeable to anions but not cations, and this is suggestive of positively charged residues in the TMDs. His 209 is a well conserved residue within FNT orthologues and is found in the so called selectivity filter (His 208 in *V. cholerae*). The residue also forms part of an FNT motif (Section 3.2.2 and Saier *et al*, 1999). This residue can become charged. The amino acid was mutated to the non-polar phenylalanine (H208F) and the function of the permease was ablated showing His 208 is required for StFocA function (Lü *et al*, 2012).

The protein crystallised from *C. difficile* is thought to transport the hydrosulphide ion, a novel substrate for a FNT protein (Czyzewski and Wang, 2012). The HSC was crystallised in buffers at three pH values (Table 1.1). No structural differences were observed and in each crystal structure the protein was in a closed state. A selection of mutant HSC proteins were also crystallised, these too were in the closed state. Mutations of residues in the so-called cytoplasmic slit or the periplasmic ring to less bulky residues caused an increase in the transport of formate due to an increase is the pore size of the protein. The authors concluded that the opening of the CdHSC is tightly regulated.

1.6 Antigenic Variation

As noted in section 1.5, the malaria genome project (Gardner *et al*, 2002a) did not result in the annotation a large number of polytopic membrane proteins. The study did however unearth a large number of monotopic membrane proteins postulated to be involved in antigenic variation. This is not unexpected as there are multiple gene families involved (Scherf *et al*, 2008). Moreover, antigenic variation is employed by parasites, including *Giardia lamblia*, *Trypanosoma* spp. and *Plasmodium* spp., and certain bacteria such as *Neisseria* and *Borrelia* (Faguy, 2000), as an immune evasion technique. Antigenic variation acts to prolong the infection of a particular pathogen leading to further opportunities for its transmission to a new host. If the pathogen killed the host in a short time period it could not be transmitted to other hosts. Conversely, if the host immune response killed the pathogen in a short time period it

similarly cannot be transmitted. Figure 1.10 demonstrates an understanding of antigenic variation schematically.

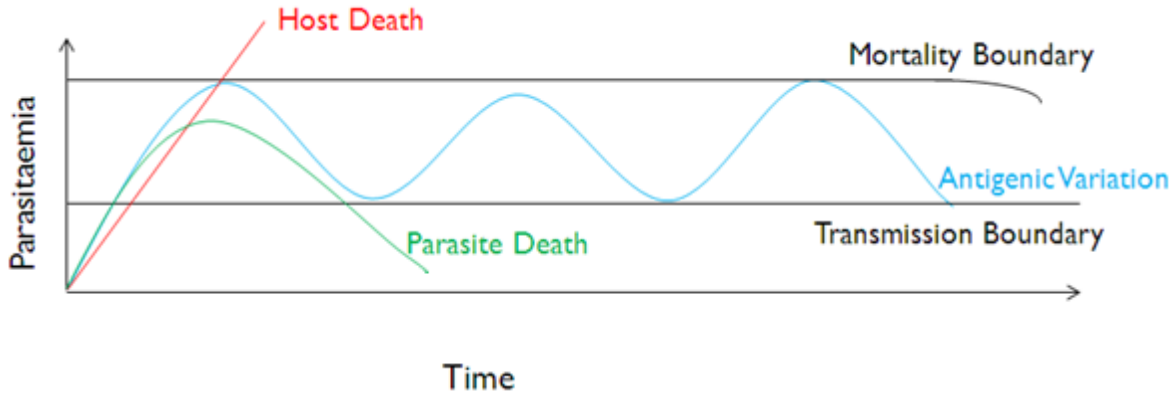


Figure 1.10: Antigenic variation. The red line indicates a pathogen that causes great harm to the host and results in an early death. Killing the host is not a useful attribute for the pathogen as it reduces the amount of time it can be transmitted to new hosts. Conversely the green line represents a pathogen which does not undergo antigenic variation, and with time the immune system is able to overcome the pathogen and eliminate it form the host. The blue line signifies a pathogen that undergoes antigenic variation. The parasite is not eliminated from the host and is able to stay in a number above the transmission boundary, and can therefore be transmitted to many new hosts. Source: A. Clarke.

In *Plasmodium* parasites, antigenic variation is particularly complicated due to the number of variants available to switch between, as well as the fact that mechanisms for switching appear to be diverse. *Plasmodium* has a range of proteins it exports to the surface of the erythrocyte. Many of these proteins are thought to play a role in antigenic variation. These variant surface antigens (VSA's) help the parasite to elude detection by the host immune system. Antigenic variation severely impedes the production of a successful vaccine against malaria as the parasites are capable of producing a multitude of different antigen types exposed to the host's immune system at various levels. These include the RBC surface targeted *var* gene products (PfEMP1 proteins) and various other gene family products expressed at the erythrocytic and other stages of the life cycle. The *var* genes vary between *P. falciparum* strains 3D7, HB3 and Dd2 and show very little overlap (Scherf *et al*, 2008), increasing the difficulty in finding target antigens for vaccine development. However, there are other antigens that may be accessible to the immune system and it is important to identify key epitopes targeted by the immune response and those involved in the most severe forms of malaria.

The rodent malarias have been key to investigating parasite-host interactions from an immunologic point of view. Given that (i) the understanding of how a protein interacts with the host immune system requires an accessible model, (ii) *P. falciparum* does not readily infect laboratory animals (the *Aotus* monkey and chimpanzees being exceptions (Herrera *el al*, 2002), (iii) experimental research on humans is unethical and (iv) it is difficult to find an *in vivo* experimental model, it is also important to study the molecular mechanisms involved in the gene switch *in vivo* as this may differ from laboratory cultured parasites. Of note however, is the fact that *P. falciparum* seems to be relatively distinct with regard to parasite-host interactions and an ideal model for this parasite does not exist (Waters, 2002).

1.6.1 Variant surface antigens

The *Plasmodium* genome codes many different families of variant surface antigens (VSAs). The importance of VSAs to *Plasmodium* is highlighted by the fact that many of the *Plasmodium* species employ antigenic variation, and the large proportion of the genome committed to it. The most understood studied of the VSAs are the *var* genes, giving rise to the *P. falciparum* Erythrocyte Membrane Protein (PfEMP-1) group. PfEMP-1 surface protein plays a key role in sequestration of the parasite, most likely to avoid removal by purification processes in the spleen. PfEMP-1 isoforms mediate cytoadherence to the endothelial lining in the capillaries through various ligands such as ICAM-1 and CD-36, in a process dependent on the PfEMP1 variant. PfEMP-1 is the cause of cerebral malaria as the IEs adhere to endothelia causing clots to form in the brain and subsequently death (Pasternak and Dzikowski, 2009).

The well-studied PfEMP proteins comprise a collection of variable and conserved domains. It has been suggested that when the protein folds the conserved domains, e.g. the Duffy Binding Like (DBL) domain, adheres to the endothelial cells while it is the variable domains that are exposed to the immune system and therefore targeted. There are 59 *var* genes in the 3D7 strain, and many more pseudogenes the purpose of which is unknown, but may act to increase the gene repertoire (Frank and Deitsch, 2006). Further gene families thought to be involved in antigenic variation in *Plasmodium* include *rif*s, *stevor*s, *Pfmc-2TM*, and, *surfin*s (Scherf *et al*, 2008). Janssen *et al* (2004) highlighted similarities between the *stevor* and *rif* families of *P. falciparum*,

the recently described *vir* genes from *P. vivax* (del Portillo *et al*, 2001) and *bir*/*yir*/*cir* gene families present in the three rodent malarias.

1.6.1.1 *rif*

The *rif* (relative interspersed family) genes or RIFIN proteins are a second gene family thought to play a role in antigenic variation. The genes themselves have a close association with the *var* genes being situated sub-telomerically. The proteins are 27 – 45 kDa in size and there are approximately 200 genes per cell. Using a 'plotsimilarity' program, the *rif* genes are thought to contain two TMDs which are fairly conserved, and have a polymorphic section in the middle of the protein. The suggested orientation of a candidate *rif* gene has the polymorphic or hypervariable region outside the cell (Kyes *et al*, 1999). The results in this 1999 paper also demonstrated that the *rif* genes are transcribed in the early and late trophozoite stages of the intraerythrocytic parasite, i.e. between 18 and 29 hours post invasion (hpi) as well as showing that expression levels differ between parasite strains. Immunoprecipitation and immunofluorescence assays (IFA) revealed that the RIFIN proteins are located on the outside of the RBCs (Kyes *et al*, 1999).

More recently it has been demonstrated that there are two subtypes of *rif* gene, dependent on the subcellular localisation. A-type *rif*s contain a particular motif while B-type *rif*s do not. The A-type is expressed on the erythrocyte surface and the B-type is not exported from the parasite cell (Petter *et al*, 2007). Both RIFIN protein types are translated during the intraerythrocytic and merozoite stages illustrating that expression is not mutually exclusive (Petter *et al*, 2007). At least one B-type RIFIN(s) is expressed in mature gametocyte stages and sporozoites (Wang *et al*, 2009b). Antibobies to RIFINs were present in patients infected with *P. falciparum* which is consistent with these proteins being exposed to the immune system at the surface of infected red cells (Schreiber et al, 2006). IFA showed that RIFINs are present in all of the erythrocytic stages including merozoites. It was demonstrated that A-type and B-type RIFINs were localised to different areas in the merozoites, but not to the cell surface.

1.6.1.2 *stevor*

stevor genes form a distinct gene family postulated to be involved in antigenic variation in *P. falciparum*. They are structurally similar to *rif* genes, and are thought to contain two TMDs separating a hypervariable region with a conserved N-terminus containing a signal sequence. *stevor* genes are temporally expressed after *var* and *rif* genes, and STEVOR proteins are displayed on late stage parasites. It has been shown that STEVOR proteins are exported to the RBC surface through Maurer's Clefts (Niang *et al*, 2009). The conserved N-terminus of the STEVOR protein can be detected by antibodies on the surface of the RBC. This observation may suggest one of two things, (i) the hypervariable region between the two TMDs is in fact intracellular, or (ii) the second TMD does not pass through the cell and acts as an anchor attaching the protein to the membrane. Both possibilities allow the N-terminal to be exposed on the RBC surface (Niang *et al*, 2009).

1.6.2 *Plasmodium* interspersed repeats

The multi-gene family, known as *pir* (*Plasmodium* interspersed repeats) was the first gene family to be shown to be conserved between the human parasite *P. vivax* (*vir*), the simian/human parasite *P. knowlesi* (*kir*), and the rodent malarias *P. berghei* (*bir*), *P. yoelii* (*yir*) and *P. chabaudi* (*cir*). The *pir* grouping represents the largest gene family revealed in *Plasmodium* spp. to date. The *cir* genes of *P. chabaudi* were the first of a major gene family to be identified in rodent malaria. This finding was later followed by the discovery of orthologues in *P. berghei*, and *P. yoelii*, which were revealed though database mining (Janssen *et al*, 2002). These genes also show clear similarities to the *vir* genes described in *P. vivax*. The *vir* gene family was described in the human malaria parasite *P. vivax* by del Portillo *et al* (2001), and is postulated to play a role in antigenic variation and is found at the erythrocyte membrane. The bioinformatic studies by Janssen *et al* (2002 and 2004) demonstrates some similarity between members of the *pir* family and *P. falciparum rif* and *stevor* members, although various aspects of the structural arrangement of the *P. falciparum* gene members indicate they must have diverged earlier. This observation, together with large differences in the amino acid sequence has left some uncertainty that the *rif* and *pir* genes are indeed evolutionarily linked (Cunningham et al, 2010). The PIR proteins of *Plasmodium* species that infect rodents offer great potential to learn about the function of these proteins and will greatly aid malaria research.

1.6.2.1 *pir* gene structure and phylogeny

The three exon/two intron structure of *pir*s from the rodent malarias (Janssen *et al*, 2002) is relatively well conserved in *P. vivax*, although this is not true for all *vir* genes. The second exon is proposed to contain a TMD (only half of *vir* genes do so) and the 3rd exon is highly conserved between sequences and has been postulated to be a cytosolic domain. Data obtained by Janssen *et al* suggests the single *rif* intron was derived from the same common ancestor as the rodent and *P. vivax* genes due to the similarity with the second gene. Other evidence pointing to a common ancestor is the similarity between the *rif* and *pir* sequences. The variable and conserved sequences are in approximately the same positions, as are the positions of the hypervariable regions. The copy number and sub-telomeric gene location are fairly conserved among the *pir* and *rif* genes. There were also "four sequence motifs conserved between most of the *cir* homologues and *rif* introns" (Janssen *et al*, 2004), suggestive of a common ancestor.

P. vivax contains 346 *vir* copies in its genome, including pseudogenes, and this high number is suggestive of variant proteins (Carlton *et al*, 2008). These genes were grouped into twelve subfamilies, A – F and more recently families G – L were described (Fernandez-Becerra *et al*, 2008 and Carlton *et al*, 2008). Subfamily C was most closely related to the *yir*/*bir*/*cir* genes, and two other families, A and D shared structural features with *P. falciparum surfin's* and *Pfmc-2tm* genes the proteins of which are present in Mauer's Clefts (del Portillo *et al*, 2004 and Merino *et al*, 2006).

Fonager *et al* (2007) have grouped the *yir* genes of *P. yoelii* into five subfamilies, according to the chromosome location, alternative splicing events that take place in the gene transcript and presence of signalling or PEXEL motifs (Hiller *et al*, 2004, and Marti *et al*, 2004). These five subfamilies do however only account for half of the *yir* repertoire. Analysis of the *P. berghei* genome shows some shared *pir* groups with *P. yoelii* but also some unique to *P. berghei* (Cunningham *et al*, 2010). The shared groups suggest a conserved evolutionary origin. The presence of the unique groups suggests separate evolution of some *pir*s which could point to different functions in the different parasites. The *P. chabaudi cir* gene multifamily was separated into two subfamilies, with many unassigned *cir*'s that showed sequence divergence (Ebbinghaus and Krücken, 2011, and Lawton *et al*, 2012). Furthermore the *cir* subfamilies were

divided into subgroups, nine for subfamily 1 and seven for subfamily 2 (Ebbinghaus and Krücken, 2011).

Comparatively little is known about the *P. knowlesi kir* genes and their proteins. The genes have been grouped into 4 types dependent on gene structure and are distributed in the chromosome (not positioned sub-telomerically like other *pirs*) (Pain *et al*, 2008).

1.6.2.2 Transcription and translation

Studies on the asexual stages of *P. vivax*, *P. yoelii*, *P. berghei* and *P. chabaudi* indicate differences in the stages in which *pir* genes are expressed. Transcription of *cir* and *bir* genes occurs during the late trophozoite stage, and transcription levels are reduced during the schizont stage (Janssen *et al*, 2002 and Girolamo *et al*, 2008). Interestingly, approximately 9 % of the BIR proteins of *P. berghei* are expressed exclusively in the mosquito stage (Hall *et al*, 2005). Microarray analyses and RNA sequencing demonstrated that different *cir* genes are transcribed during different stages of the IDC, and different *cir*s were expressed in different parasitised cells. Some *cir*s were found to be expressed highly in various populations (Lawton *et al*, 2012).

The *vir* genes are transcribed in waves with transcription occurring in early ring stages and schizonts. Approximately 59 % of *vir*s are transcribed during the IDC (Fernandez-Becerra *et al*, 2008). The *vir* repertoire was examined in natural populations to find out how wide spread the genes are as they account for approximately 10 % of the genome (Merino *et al*, 2006). This study showed that a wide range of *vir* genes are expressed, with the analyses being carried out on the genes present in each subfamily. The VIR proteins are not clonally expressed (Fernandez-Becerra *et al*, 2008). In primary infections, there were IgG antibodies to many VIR proteins suggesting they are not in fact clonally expressed, pointing to a function other than that of (solely) antigenic variation.

The profile for *P. yoelii* is very different as the transcription of *yir* genes is dependent on the host immune response. In immunocompetent mice a different *yir* gene was transcribed upon re-infection, yet in immunocompromised mice *yir* transcription

remained constant (Cunningham *et al*, 2005). In *P. yoelii*, many *yir* genes are transcribed at any one point, albeit at low levels, but not all are translated (Cunningham *et al*, 2005), suggestive of a post-transcriptional effect to modulate the particular gene being expressed. However, in an individual cell only 1 – 3 *yir* genes are transcribed at any one time (Cunningham *et al*, 2009).

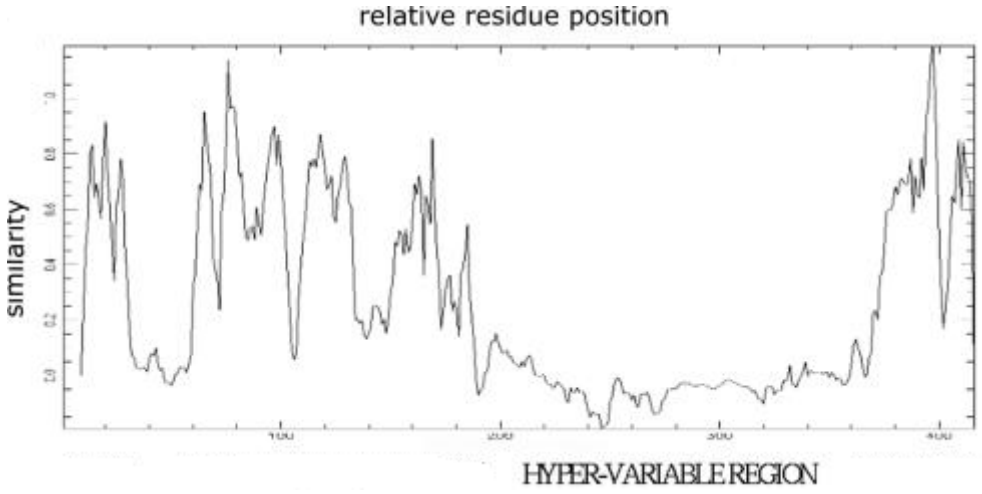


Figure 1.11: PIR family amino acid conservation plot. A scheme to illustrate the conservation of a set of sequences at the amino acid level. The line of similarity drops close to the *x*-axis as the sequences are very divergent at this point and is where the hypervariable region is placed. Image taken from Janssen *et al* (2004).

Alternative splicing has been shown to take place in a subset of the *yir* genes. It is thought that this alternative splicing may regulate the translation of YIR proteins. A conserved promoter motif found in the 5' untranslated region (UTR) has been identified in the *P. yoelii* genome upstream of the *yir* transcripts. The motif was also found to be present in the *P. chabaudi* and *P. berghei* genomes. The 5' UTR motif could act as a 'general recognition element' for the transcription of these genes. However, as the domain is present in all *pir* genes this probably does not play a role in their regulation (Fonager *et al*, 2007).

1.6.2.3 PIR proteins

Given the presence of a TMD and the hypervariable region of the *pir* genes (Figure 1.11), it is generally considered that these proteins play a role in antigenic variation. Therefore it is anticipated that these proteins would be displayed at the surface of the

RBC. B-type RIFINS have been shown to be expressed in merozoites (Mwakalinga *et al*, 2012), therefore it is possible that the proteins would be expressed on the surface of merozoite too. The amino acid sequences from approximately 46 % (160/346) of VIR proteins contain a PEXEL motif (Hiller *et al*, 2004, and Marti *et al*, 2004) and approximately 50 % (171/346) contain a putative TMD (Carlton *et al*, 2008). Consistent with these data, VIR proteins localise to the surface of reticulocytes (Fernandez-Becerra *et al*, 2008 and del Portillo *et al*, 2004). The YIR protein from *P. yoelii* has been shown, using antibodies, to be antigenic and to be present at the surface of IEs (Cunningham *et al*, 2005). The cellular location of the CIR protein family in *P. chabaudi*, a species thought to undergo both rosetting (Mackinnon *et al*, 2002) and antigenic variation has been experimentally determined using western blots. The analysis was directed against the Triton X-100 insoluble fraction of parasites and erythrocyte ghosts. Immunofluorescence analysis also demonstrated that the protein was present on the erythrocyte surface (Janssen *et al*, 2004). The BIR proteins of *P. berghei* were associated with detergent resistant membrane lipid rafts and the PV (Girolamo *et al*, 2008) suggesting the BIR proteins may be molecular chaperones or are translocated outside the parasite by chaperones.

The protein expression of YIR is thought to be predominantly in the schizont stage as the majority of cells observed when stained with YIR antisera were schizonts and less than 2 % were in ring or trophozoite stages (Cunningham *et al*, 2005). YIR proteins were translated in all stages of the intraerythrocytic lifecycle. The proteins translated are not always the same throughout the stages (Cunningham *et al*, 2009). Contrary to the results in other species, a small number of CIR proteins are transcribed at any one time in an *in vivo* model over the 24 h period of investigation (Ebbinghaus and Krücken, 2011). Akin to the other PIR proteins, the amino acid sequence of KIR contains a single predicted TMD at the C-terminal (Jemmely *et al*, 2010). In some *Plasmodium* proteins a sequence motif, known as the PEXEL motif, is involved in carrying proteins beyond the vacuolar membrane, for example in protein targeted to red cell membranes (Hiller *et al*, 2004, and Marti *et al*, 2004). No PEXEL motif is found within the KIR protein sequences. However, another motif, the ZLPS motif, has been postulated to carry KIR proteins to the red cell surface (Pain *et al*, 2008).

The function of these PIR proteins has not yet been demonstrated. However: 9 % of BIRS are expressed in ookinetes, VIR's are non-clonally expressed, B-type RIFINS are expressed intracellularly and expressed in the gametocyte and sporozoite stages. This is an indication that roles other than avoidance of mammalian immunity must be considered for at least a portion of these gene products.

A study was undertaken to establish in which host tissues *cir* transcripts were present (Ebbinghaus and Krücken, 2011). A broad range of tissues were analysed with products present in lung, liver, blood, spleen, kidney and brain tissues. Interestingly for some subgroups, sequences were totally absent from some tissues. For example, in subgroup 5 of subfamily 1, no sequences were observed in liver tissues, where 60 % of the total sequences (all subgroups) were present in the liver. With respect to subgroup 1, no sequences were observed in the brain tissue (Ebbinghaus and Krücken, 2011). The differences in tissue type expression of the subgroups points towards multiple roles for this family of proteins. Lawton *et al* (2012), and Janssen *et al* (2004) identified conserved amino acid motifs within the CIR and PIR sequences, respectively. Lawton *et al* (2012) suggest that the conservation of these motifs put constraints on the gene diversity, suggesting important biological functions for CIR's. The differences observed in cir transcripts between groups A and B, and the temporal expression of *cirs* suggests a different biological function for A and B CIRs and possibly stage specific roles (Lawton *et al*, 2012).

1.6.3 Molecular mechanisms involved in switching

The molecular mechanisms for gene switching are poorly understood, and most work has concentrated on the *var* genes of *P. falciparum*. Each of the *P. falciparum var* genes has a dedicated promoter. Therefore, one gene is switched on while another is inactivated to provide the organism with only one expressed surface antigen (of the *var* family) (Frank and Deitsch, 2006).

The subtelomeric location of these *var*, *rif*, *stevor* and *pir* genes is important for gene diversity due to the increased possibility of multiple recombination events. In the nucleus, *var* genes are clustered together at perinuclear positions. This positioning has been inferred as being important for the generation of new genes due to the closeness of the various *var* genes (Frank and Deitsch, 2006). The *P. vivax* genome encodes

many pseudogenes in the *vir* family (del Portillo *et al*, 2001) and these could be important for recombination and thereby increase the *vir* repertoire. Another mechanism has been attributed to the A+T richness of the *Plasmodium* genome. The A+T rich genome has been postulated to increase the recombination events which may increase the rate of generation of immunological diversity (Winzeler, 2008).

The perinuclear position of the *var* genes is also thought to be important for the switching events. Active genes are thought to be maintained in a specific region of the nucleus, and the relocation of chromosome ends switches the active gene with another (Dzikowski *et al*, 2006). Each *var* gene has its own promoter, and these promoters have been grouped into three sub-types. UpsA is found at the subtelomere, UpsC is found in internal clusters and UpsB a combination of the two locations. These promoter types may have influence the gene expression (Frank and Deitsch, 2006). *Sir2A* and *Sir2B* are *P. falciparum* orthologues of a *S. cerevisiae* gene that plays a role in gene regulation. The *sir*2 gene product binds to the chromosome ends and appears to prevent activation of the *var* gene. As Sir2 binds to the chromosome ends, only the genes driven by the UpsA and UpsC promoter are affected (Frank and Deitsch, 2006). As the Sir2 protein binds to the end of the chromosome, the internal *var* genes do not appear to be affected (Tonkin *et al*, 2009). The influence of PfSir2 on *rif* gene expression was also considered. In this regard, it was demonstrated that in a *pfsir2a* knock-out strain the *rif* genes that were expressed often had close associations with *var* genes. In a *pfsir2a* knock-out, the temporal *rif* expression was altered with *rif*'s being expressed at all stages of the intraerythrocytic lifecycle. A *pfsir2b* knock-out mutant had little effect on *rif* gene expression (Tonkin *et al*, 2009).

1.7 Aims of this Study

(i) The biochemical characterisation of *P. falciparum* NitA based on the development of a set of assays designed for the determination of substrate specificity. This line of enquiry should *a priori* further our understanding of the role of the PfNitA transporter in *Plasmodium*.

(ii) Determination of the importance of the PfNitA protein to asexual parasites and validation as a potential drug target by reverse genetics in *P. falciparum*. In addition, the determination of the localisation of PfNitA within the asexual stage parasite.

(iii) Recombinant expression of *P. chabaudi* Cir2 (17129530), a representative PIR protein in *E. coli* for structural and topological studies, towards determining a role for the protein family.

The results of this study are provided in three chapters. Chapters 3 and 4 contain research carried out into the functional and biochemical characterisation of PfNitA. Chapter 5 describes the production of rPcCir2 protein and its structural analyses. The results presented in these three chapters will be discussed in Chapter 6.

2 Materials and Methods

2.1 Materials

2.1.1 Consumables

Reagents were of high grade and obtained from Sigma-Aldrich (Dorset, UK), Fluka (Gillingham, UK) and BDH (through VWR, Lutterworth, UK). Enzymes were from Promega (Southampton, UK) or NEB (Hitchin, UK) unless stated otherwise.

2.1.2 Bacterial cultures

Bacterial overnight cultures were grown in 4.5 ml Luria Bertani (LB) broth (Appendix 1) with a plasmid selection antibiotic if required. LB agar plates were used for bacterial transformations, and to streak out cultures to produce single colonies. The agar plates were kept at 4 ^{o}C for short term storage. For long term storage 0.5 ml overnight liquid culture was mixed with 0.5 ml fresh LB + glycerol (30 % v/v glycerol) and stored in cryovials at -85 ^{o}C. Liquid cultures were grown in an orbital incubator with agitation at 220 – 250 rpm, and solid cultures in a static incubator, both were grown at 37 ^{o}C unless stated otherwise.

2.1.3 *E. coli* strains

DH5α (*sup*E44 Δ *lac*U169 (φ80*lacZ* Δ M15) *hsd*R17 *rec*A1 *gyr*A96 *thi*-1 *rel*A1) cells were routinely used in the transformation and amplification of plasmids, and the transformation of ligated products.

BL21 (DE3) (*hsdS gal* (λ*cI*ts857 *ind*1 Sam7 *nin*5 *lac*UV5-T7 gene *1*)) cells were used for protein expression, and are compatible with vectors using the T7 and *tac* promoters for recombinant expression of proteins.

C43 (DE3) cells are a mutant strain of BL21 (DE3) cells that were found to be particularly useful for the expression of membrane proteins as they resulted in a lower toxicity to that observed with the BL21 (DE3) strain. The strain has two uncharacterised mutations (Miroux and Walker, 1996).

Arctic Express™ (DE3) (Stratagene, Stockport, UK) (B F– *ompT hsdS*(rB– mB–) *dcm*+ Tetr *gal* λ(DE3) *endA* Hte [*cpn10cpn60* Gentr]) cells have been specially designed for the expression of proteins at lower temperatures. Protein production at low temperatures is achieved by the co-expression of the two chaperones Cpn10 and

Cpn60 present on a pACYC-based plasmid that encodes a gentamycin resistance gene for selection.

OneShot® (Invitrogen) (F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lacZ*ΔM15 Δ*lacX*74 *rec*A1 *ara*D139 Δ(*araleu*) 7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G) cells are resistant to killing by the *ccdB* gene product. These were used for the propagation of the Gateway vectors.

JCB4018 (RK4353 Δ *napA-B* Δ*narZ::Ω* Δ*narK::tet narU::kan*) and **JCB4250** (RK4353 Δ*napA-B* Δ*narZ::Ω* Δ*narK::tet narU::kan* Δ*nirC::cam* Δ*nrfA::chl*) are kind gifts from Prof. J. Cole, University of Birmingham, UK, for the PfNitA functional complementation assay. JCB4250 is a knock-out for the *E. coli* NirC nitrite transporter (Jia *et al*, 2009).

MC4100 (F$^-$ *araD139* Δ *(aarfF-lac)U169 ptsF25 deoC1 relA1 flbB530 rpsL150λ*$^-$), **REK701** (like MC4100 but Δ*focA*) and **RM201** (MC4100, Δ(*pfl-25*) Ω(*pfl*::*cat* pACYC184)) are kind gifts from Prof. G. Sawers, Martin Luther University, Germany, for the PfNitA functional complementation assay. REK 710 is a loss of function mutant for the *E. coli* Foc A formate transporter (Suppmann and Sawers, 1994).

CC118 (Δ(ara-leu), *araD*, Δ*lacX*74, *galE*, *galK*, *phoA*20, *thi*-1, *rpsE*, *rpoB*, *argE*(Am), *recA*,) cells were used in the alkaline phosphatase assay, and were a kind gift from Dr. D. Daley, Stockholm University, Stockholm, Sweden.

2.1.4 Plasmids

Plasmids were cloned and amplified in DH5α competent cells, with the exception of the Gateway Vectors which were amplified in the specifically designed OneShot® Top 10 competent cells.

pTTQ18 and **pTTQ10H** are protein expression plasmids encoding 6 and 10 Histidine residues, respectively, as a tag for purification. Both encode an ampicillin resistance gene as a selectable marker, tac promoter and are approximately 4500 bp in size. IPTG is used to induce protein expression (Stark, 1987).

pEHISTEV and **pEHISGFPTEV** are protein expression plasmids encoding a N-terminal 6 x His tag and a TEV protease site to cleave the tag from the expressed protein. The plasmids were kind gifts from Dr. H. Lui, University of St Andrews. The pEHISGFPTEV encoding a N-terminal green fluorescent protein (GFP) reporter gene. These plasmids encode a kanamycin resistance gene as a selectable marker, and a T7 promoter, IPTG

is used to induce protein expression. pEHISTEV is approximately 5500 bp and pEHISGFPTEV approximately 6500 bp in length (Lui and Naismith, 2009).

pV5M and **pV5MGPD** are around 7700 bp in length and encode an ampicillin resistance gene as a selectable marker for selection in *E. coli* and the *argB* gene for selection in *A. nidulans*. The V5M plasmid uses the NrtA promoter and V5MGPD the GPD promoter. Both plasmids are used in fungal transformations. At the 3' end of the insert there is a sequence encoding a V5 viral protein epitope for detection in western blots using the horseradish peroxidase (HRP) conjugated anti-V5 antibody (Invitrogen).

pCC1 and **pCC4** were used to create *Plasmodium falciparum* knock out constructs. The plasmids were a kind gift from Dr. T.K. Smith, University of St Andrews. pCC1 encodes the *hdhfr* (human dihydrofolate reductase) gene and confers resistance to Walter Reed Institute anti-DHFR antifolate 99210 (WR 99210), and pCC4 encodes the *blasticidin S deaminase* gene conferring resistance to blasticidin for positive selection. Both plasmids contain the cytosine deaminase gene as a negative selectable marker utilised during the transfection procedure. There is an ampicillin resistance gene present for positive selection during cloning. These plasmids create KO's via double cross over recombination (Duraisingh *et al*, 2002 and Maier *et al*, 2006).

pHA-1 has a C-terminal PhoA reporter gene. pHA-1 was a kind gift from Dr. D. Daley (Stockholm University), and encodes a 'stuffer' gene, *yedZ*. It encodes an ampicillin resistance gene and is approximately 8000 bp. (Whitley *et al*, 1994).

pGFPe has a C-terminal GFP reporter gene and is a modified version of the pET28 (a+) plasmid. As it is made from the pET28 (a+) plasmid it encodes an 8x His tag, the kanamycin resistance gene and the T7 promoter. The plasmid is approximately 7000 bp in length. pGFPe was also a kind gift from Dr. D. Daley, the original was created by Waldo *et al* (1999).

2.1.4.1 Gateway® plasmids

The Gateway® (Invitrogen) vectors are a set of plasmids designed for use in *P. falciparum* (Tonkin *et al*, 2004 and van Dooren *et al*, 2005). They comprise a set of vectors in which a promoter is encoded, the gene of interest and a tag gene such as GFP.

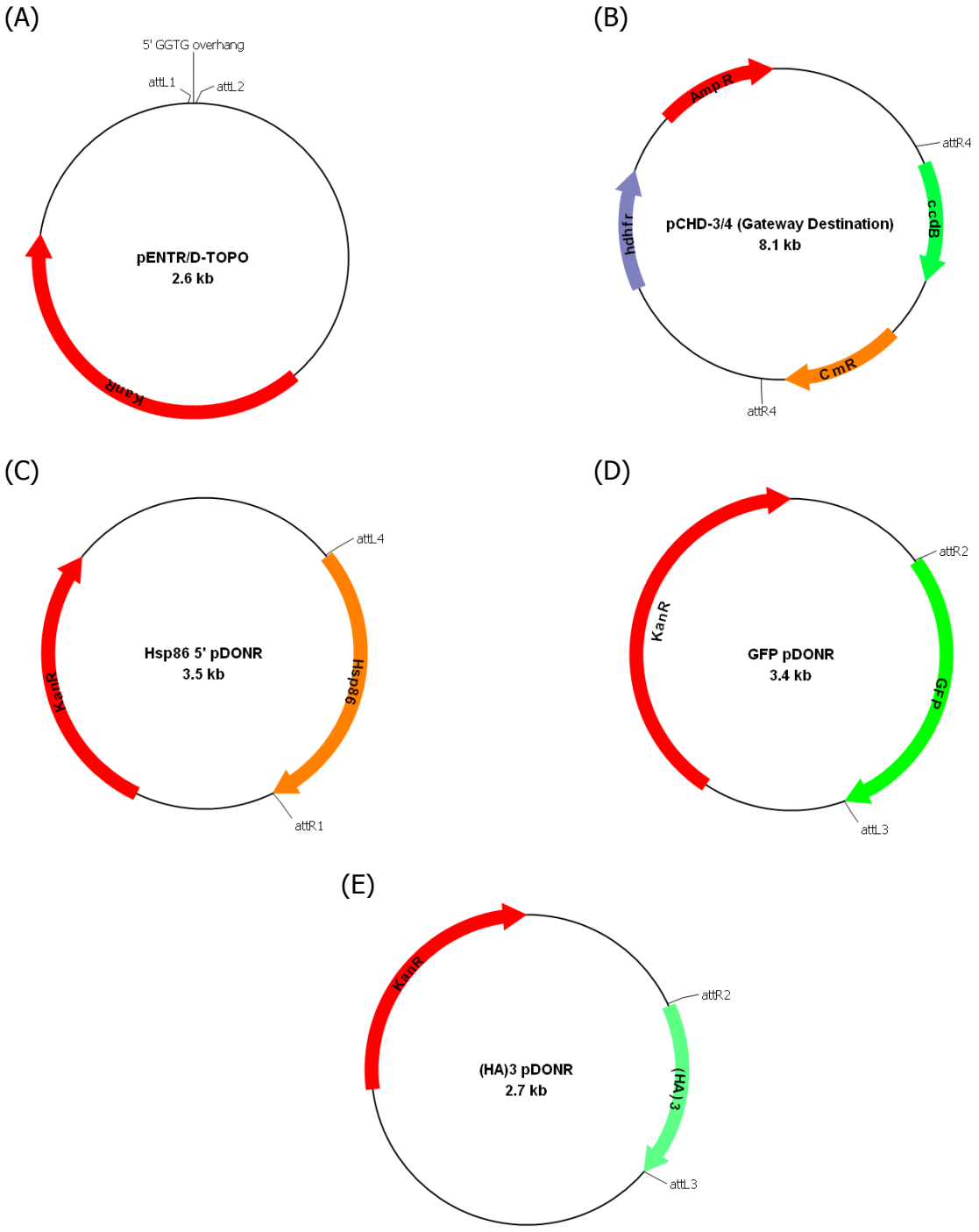


Figure 2.1: Gateway® vectors. (A) the entry vector into which the gene is cloned highlighting the 5' GGTG overhang for directional cloning. (B) the destination vector. Between the *att* sites is a chloramphenicol resistance gene and a *ccdB* gene which are lost after correct recombination upon production of the final plasmid vectors. (C) Hsp-86 promoter. (D) GFP tag and (E) HA tag. In each vector the att sites used for Gateway cloning are given, as well as the DSM's used during bacterial cloning.

By using the Gateway® cloning method a gene of interest can have either an N-terminal of C-terminal fusion tag. The Gateway® vectors used in this thesis were a kind gift from Prof. S. Müller and are detailed in Figure 2.1.

pENTR/D-TOPO is an entry vector used for the cloning of the gene of interest after PCR amplification.

pCHD3/4 is the destination vector. Plasmids encoding the genes of interest i.e. the above entry vectors are combined with the destination plasmid in a LR reaction (section 2.2.10)

pDON-R GFP, **pDON-R (HA)$_3$** and **pDON-R HSP-86** are entry vectors, these vectors encode 3' or 5' tags.

2.1.5 Synthetic genes

The *P. falciparum nita* gene was synthesised according to a codon bias required for expression in *Aspergillus nidulans* (*pfnita_opt*) and *E. coli* (*pfnita_optEc*). The *P. chabaudi cir*2 gene was synthesised according to the codon bias required for expression in *E. coli* (*pccir*2). Both genes were purchased from GenScript (NJ, USA). Sequences of the synthetic genes are given in Appendix 6.

2.1.6 Antibiotics

Antibiotics were made up in 1 ml aliquots at 100 mg/ml and stored at -20 ^{o}C. Table 2.1 gives the antibiotics used in this thesis and working concentrations. Carbenicillin (www.carbenicillindirect.com) was used instead of ampicillin (Melford, Suffolk, UK) in protein expression trials and for extended growth conditions as the ampicillin begins to decompose after 16 h. All other antibiotics were obtained from Sigma.

Antibiotic	Working concentration
Kanamycin	50 µg/ml
Carbenicillin	100 µg/ml
Ampicillin	100 µg/ml
Gentamycin	20 µg/ml

Table 2.1: Working concentrations of each of the antibiotics used in this thesis.

2.2 Molecular Biology

2.2.1 Polymerase chain reaction

Polymerase chain reaction (PCR) was used to amplify gene products of interest for use in cloning and hybridisation. Oligonucleotide primers were obtained from VH Bio (Gateshead, UK) and Eurogentec (Fawley, UK) and are given in Appendix 2. PCR's were carried out using the enzymes: Vent Thermo (NEB), Phusion (Finnzymes, Espoo, Finland) and Kapa HiFi (Kapa Biosystems, Woburn, MA, USA) polymerases and the buffers provided. Each of these enzymes is a high-fidelity enzyme with proof reading activity. When amplifying products from *Plasmodium falciparum* genomic DNA (gDNA), Kapa HiFi polymerase was used with the GC Buffer provided or Phusion polymerse. The reactions were carried out in a G-Storm GS1 thermocycler. PCR products were sequenced after they had been cloned into their respective plasmids. Sequencing was outsourced to the Dundee Sequencing Service (Dundee University, UK) or Macrogen (Seoul, South Korea), and carried out using specific PCR primers used in the PCR reaction or the T7 primer. Sequences were analysed using Sequencher (Gene Codes Corporation, Ann Arbor, MI, USA).

2.2.2 Mutagenesis PCR

Mutagenesis PCR was carried out to generate a single amino acid mutation, the technique followed was that of splice overlap extension (Warrens *et al*, 1997). Splice overlap extension involves two rounds of PCR. The so called hybrid primers are used to introduce the mutation of interest with a set of flanking primers, which are also used to create the full-length gene. The process is illustrated schematically in Figure 2.2. The first round of PCR was carried out using a hybrid antisense primer (3) and a flanking sense primer (1) to obtain the 5' fragment of the gene (PCR 2). The 3' fragment of the gene was amplified using a hybrid sense primer (2) and a flanking antisense primer (4) (PCR 1). The 5' (Product 1) and 3' (Product 2) gene fragments overlap by 22 base pairs and both include the mutation. These two gene fragments were used to construct the full-length gene which was amplified using the two flanking primers (1 and 4). The final PCR (3) produces one full-length gene containing a single, desired mutation (Product 3).

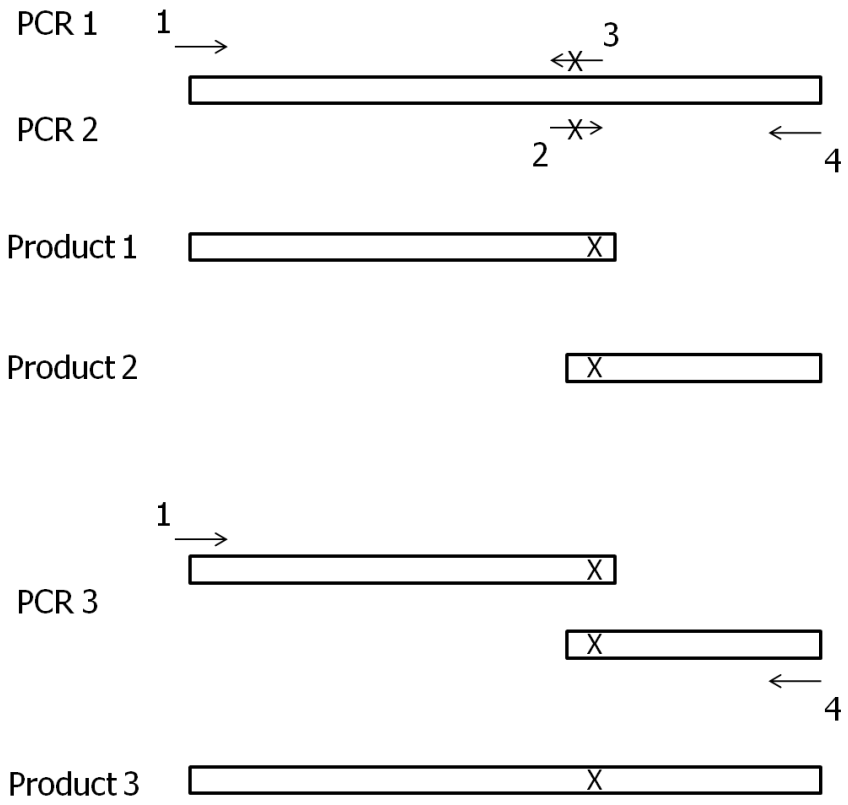


Figure 2.2: Mutagenesis by splice overlap extension. Two PCRs are carried out using the gene as the template and primer sets 1 + 3 (PCR 1) and 2 + 4 (PCR 2). Hybrid primers 2 and 3 contain the desired mutation (X). Both PCR products overlap by 22 bp where the mutation was created. PCR products 1 and 2 containing the mutation are purified and serve as the template for the next round of PCR using the flanking primer set 1 + 4 (PCR 3). The 22 bp overlap allows annealing of the two PCR products. The final PCR product 3 codes for the full-length protein, and contains the single mutation. Source: A. Clarke.

2.2.3 Agarose gel electrophoresis

DNA products were run on a 1 % (w/v) agarose gel for visualisation and quantification, for smaller DNA fragments a higher percentage agarose gel (1.5 – 2 % w/v) was used. Agarose was made in 1 x TAE buffer (40 mM Tris acetate, 1 mM EDTA), and ethidium bromide was added to a final concentration of 0.5 µg/ml. For large fragments xylene cyanol (30 % glycerol, 0.25 % xylene cyanol in distilled water (DW)) and for small fragments bromophenol blue (30 % glycerol, 0.25 % bromophenol blue in DW), were used as a loading and tracking dye. A molecular weight marker was added, Hyperladder I (Bioline, London, UK) or Keyzyme 100 bp plus DNA ladder (Anachem,

Luton, UK). Gels were viewed over a UV transilluminator and the image recorded using the Herolab E.A.S.Y Photographic Suite (Wiesloch, Germany).

2.2.4 PCR product purification

PCR products were either purified directly from the reaction, or if spurious bands were present extracted from agarose gel.

Purification from PCR reactions was carried out using PCR purification kits from: Invisorb (Berlin, Germany), Marligen Biosciences (MD, USA), Omega Biotek (GA, USA), and GenScript. The PCR product was purified using the instructions and solutions provided with each of the kits. DNA was eluted in the elution buffer provided or TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0).

Fragments were extracted from an agarose gel using the manufacturer's protocol in a Qiagen (Crawley, UK) Gel Extraction Kit. Briefly, the DNA band was excised from the gel, incubated with buffer and silica beads at 50 ^{o}C, and washed to remove salts and other contaminants from the gel. The DNA was eluted from the beads by incubation in TE buffer.

A 1/10 elution volume of the purified PCR product was run on an agarose gel to confirm the correct size of the product, and for quantification a Nanodrop, ND-1000 spectrophotometer (ThermoScientific, Northumberland, UK) was used.

2.2.5 Plasmid isolation

Plasmids were isolated from overnight culture following the instructions provided in a Qiagen miniprep kit. The plasmids were eluted in 30 – 50 µl TE pH 8.0 and stored at -20 ^{o}C. If large concentrations of plasmid were required, for example for malaria transfections, a Qiagen maxi prep was used. 1 µl of plasmid was retained for digestion to confirm the size and presence of insert by restriction analysis, if an insert was present. Plasmids without insert were also cut to verify they were of the expected size (see section 2.2.7).

2.2.6 Ethanol precipitation of DNA

Precipitation of DNA was carried out to concentrate a sample of DNA. 3 volumes of 100 % ethanol was added to a purified DNA sample, followed by 1/10 volumes of 3 M sodium acetate (pH 5.2). To facilitate DNA precipitation, 1 µl of 20 µg/µl glycogen (Invitrogen) was added. The DNA suspension was incubated at -20 ^{o}C for 24 h. The sample was centrifuged for 20 mins at maximum speed on a benchtop centrifuge, 4 ^{o}C washed with 200 µl of 70 % ethanol, and centrifuged as before, for 10 mins. The sample was taken to a sterile tissue hood, the supernatant removed, and the precipitated DNA allowed to air dry to remove the ethanol. The DNA was resuspended in 30 µl of filter sterilised MilliQ (MQ) water and stored at 4 ^{o}C for use in malaria transformations (see section 2.5.5).

2.2.7 Restriction endonuclease digestion

Up to 1 µg of DNA was digested with 10 – 20 units (U) of a specific Type II restriction endonuclease for 2 – 3 h, with the exception of *Sac*II where 30 U were used per µg DNA. When digesting with two enzymes using different buffer, the buffer was chosen as directed from the manufacturer, or a sequential digest was carried out using the lower salt containing buffer first followed by addition of salt to the required concentration of the second enzyme. After digestion the reaction was treated with 5 units Shrimp Alkaline Phosphatase (Promega) or Arctic Alkaline Phosphatase (NEB) in the buffer provided. The DNA concentrations after digestion were checked by running 1/10 volume on a gel to visualise the DNA and quantified using a Nanodrop spectrophotometer.

2.2.8 Drop dialysis of DNA

Drop dialysis was carried out on all purified PCR, digestion products and plasmid isolations to remove excess salts which may interfere with the subsequent reactions. Dialysis was carried out using HAWP mixed cellulose ester membranes from Millipore (Hertfordshire, UK) with a pore size of 0.025 µm. The product to be dialysed was pipetted on the membrane floating on autoclaved DW. The sample was left for 30 minutes (mins) before being removed from the dialysis membrane.

2.2.9 DNA ligation

Ligations were carried out using a Rapid DNA Ligation Kit (Fermentas, York, UK) according to the manufacturer's instructions; digested plasmid and digested insert were mixed at a ratio of 1:3, and incubated with T4 ligase at room temperature for 1 h. 2.5 U of ligase were used in a 10 µl reaction.

2.2.10 Gateway® cloning

The Gateway Cloning system was used to prepare knock-in constructs for *Plasmodium* transfections. The cloning reactions were carried out as described in the Invitrogen manual and are described below. Figure 2.3 displays the reaction schematically.

The gene of interest was first amplified by PCR using primers specific to the gene with the forward primer containing a 5' CACC sequence for the cloning into the pENTR™/D-TOPO® vector. The PCR (0.5 – 4 µl) product was combined with 1 µl pENTR™/D-TOPO® at a 0.5:2 – 2:1 molar ratio, and 1 µl of a pre-defined salt solution, and made up to 6 µl with MQ water and incubated at 22 ^{o}C for 5 mins. If the PCR product was 1 kb 1 – 5 ng was recommended and if 2 kb 5 – 10 ng recommended. Two microlitres of the ligation reaction was transformed into chemically competent OneShot® Top 10 DH5α cells. The remainder of the ligation reaction was stored at 4 ^{o}C.

The Gateway® cloning reaction combines the desired PCR product (in the pENTR™/D-TOPO® vector) with a promoter (5' element) and an N- or C-terminal tag such as green fluorescent protein (GFP) or haemaggluntinin ($(HA)_3$) (3' element). The reaction is illustrated schematically in Figure 2.4.

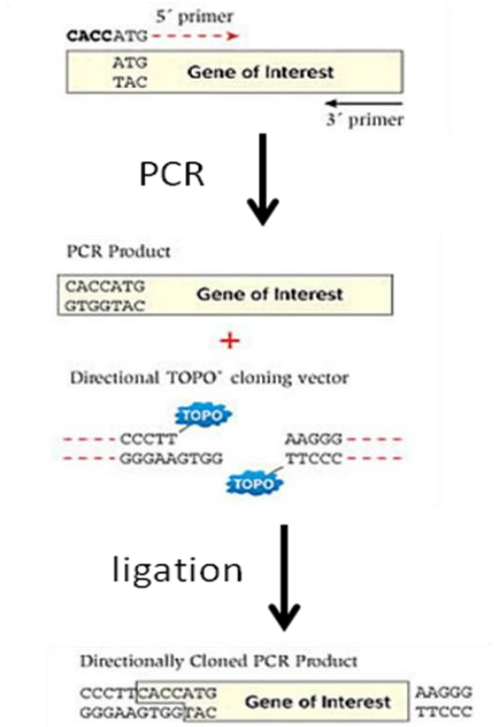


Figure 2.3: Topo cloning reaction. An overview of the pENTR™/D-TOPO® vector cloning. The PCR product contains a CACC sequence, which allows directional cloning to the GTCC overhang in the vector. Image is adapted from Invitrogen[15].

The three plasmids containing the tag, PCR product and promoter were diluted and used at 10 fmoles, the volume not exceeding 7 µl. The destination vector was concentrated to 20 fmoles in 1 µl, the destination vector used is pCHD-3/4 (Figure 2.1) which has been specially designed for use in *Plasmodium*. The total reaction volume was made up to 8 µl with TE buffer (pH 8.0), 2 µl of LR Clonase™ II Plus (Invitrogen) added to the plasmid mixture, and the sample centrifuged briefly. The reaction was incubated at 22 °C for 16 h. After the 16 h incubation, 1 µl of 2 µg/µl proteinase K was added to stop the reaction. The sample was transformed into chemically competent OneShot® DH5α cells.

[15] http://products.invitrogen.com/ivgn/product/K240020

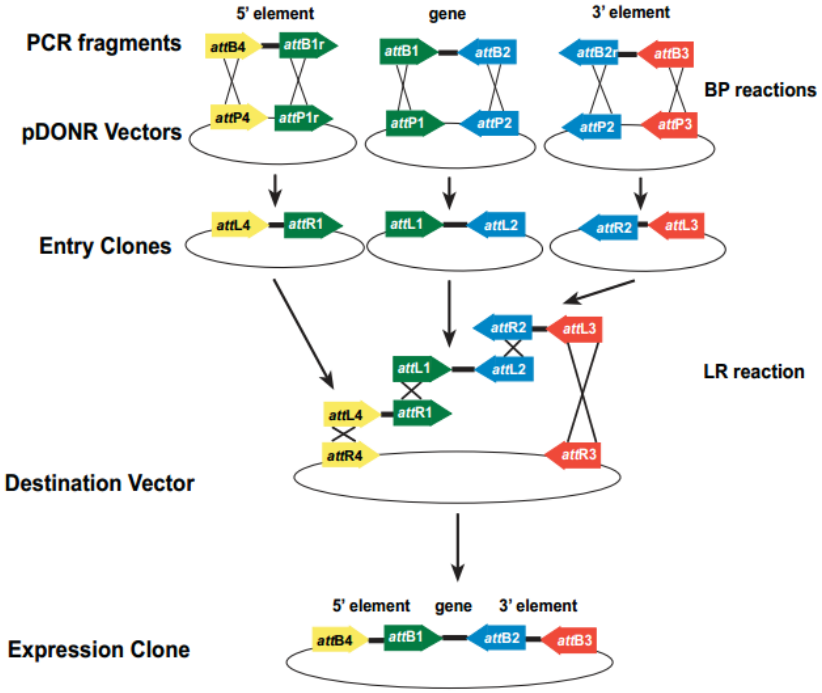


Figure 2.4: MultiSite Gateway® 3-Fragment reaction. Overview of the Gateway® cloning reaction. The entry clones were pre-made using the BP reaction and contain the Hsp-86 promoter (5' element) and GFP or $(HA)_3$ tags (3' element). The gene specific entry clone was created by the TOPO reaction described above. The LR reaction was carried out to bring each of the Entry Clones together with the Destination Vector to create the final construct. Image is taken from the MultiSite Gateway® manual (Version H, 31 October 2010) (Invitrogen)[16].

2.2.11 Preparation of competent cells

Competent cells were prepared according to the method by Inoue *et al* (1990). *E.coli* were streaked on LB agar and grown overnight at 37 °C. A single well defined colony was used to inoculate LB broth and grown overnight. One millilitre of overnight culture was used to inoculate 100 ml of SOB medium (Appendix 1). The freshly inoculated culture was grown at 18 – 20 °C with rotary shaking at 180 rpm until an OD_{680} of approximately 0.6 was reached. The flask was left to stand on ice for 10 mins, and the cells harvested by centrifugation at 1000 x *g* for 15 mins at 4 °C. The cell pellet was resuspended in 40 ml cold Transformation Buffer (10 mM PIPES, 15 mM calcium chloride, 250 mM potassium chloride, 55 mM manganese chloride, pH 6.7), and incubated on ice for 10 mins, before further centrifugation as before. The resulting pellet was resuspended in 4 ml cold Transformation Buffer, 0.3 ml DMSO added, and

[16] http://tools.invitrogen.com/content/sfs/manuals/multisite_gateway_man.pdf

the cells mixed gently and incubated on ice for a further 10 mins. The cells were aliquoted into pre-chilled (to -85 °C) microfuge tubes and stored at -85 °C.

2.2.12 Bacterial transformation

Fifty microlitres of competent cells were added to 50 – 100 ng plasmid DNA or full ligation reaction, and incubated on ice for 20 mins. The cells were 'heat shocked' at 42 °C for 90 s (DH5α) or 45 s (all other strains), followed by 2 mins on ice. Two hundred microlitres of prewarmed SOC medium (Appendix 1) was added and mixed gently, followed by a recovery period in a 37 °C incubator with agitation (220 rpm) for 1 h or 2 h if cells were to be plated on LB + kanamycin. If a low number of transformants was expected the tubes were centrifuged briefly using a bench top centrifuge and the pellet resuspended in 100 µl (of the same medium) before being plated on LB agar containing the selection antibiotic, otherwise 100 µl of the transformation was plated.

2.2.13 Colony cracking

To check transformed bacterial colonies for the correct plasmid, colony cracking was used. A single colony was resuspended in 10 µl DW, and 10 µl Cracking Buffer (200 mM sodium hydroxide, 0.5 % SDS, 10 g sucrose, and a few crystals of bromocresol green) added. The lysed colony was run on a 0.8 % agarose gel alongside a positive control plasmid of about the expected size, and a negative control of vector without an insert for comparison.

2.2.14 Colony PCR

Colony PCR's were carried out to confirm the presence of an insert in a vector after a transformation. The colony was suspended in the volume of PCR, generally 10 µl. The PCR was carried out in the same way that it was amplified from plasmid or genomic DNA, or by using a taq polymerase; MyTaq™ HS DNA Polymerase (Bioline) and 2x Thermo-Start HP ReddyMix (Thermo Scientific).

2.2.15 Southern blotting

Two methods were used to do Southern blots. The first was carried out in the University of Glasgow, the second at the University of St Andrews.

Method 1: Vacuum transfer with an alkaline phosphatase probe

Genomic DNA (gDNA) (1 – 2 µg) was digested using 1 µl of the desired restriction enzyme/s. The total digestion was loaded on an 0.8 % agarose gel and electrophoresed at 20 V overnight, the first and last lanes of the gel were left bank. The following day the gel was stained using Sybr Safe® (Invitrogen) and the gel photographed alongside a fluorescent ruler so that migration distances could be measured. The gel was washed gently in MQ water and incubated with agitation, in 300 ml of 0.25 M hydrochloric acid for 10 – 15 mins, the gel was washed twice in MQ water, followed by incubation in 300 ml denaturation buffer (1.5 M sodium chloride, 0.5 M sodium hydroxide) for 30 mins. Again the gel was washed twice in MQ water, incubated in neutralisation buffer (3 M sodium chloride, 0.5 M Tris, pH 7.0) for 30 mins, and washed once in MQ water. A vacuum blotting device (VacuGene XL, GE Healthcare, Chalfont St. Giles, UK) was set up to transfer the gDNA to a Hybond-N^+ membrane (GE Healthcare), 20 X SSC (3 M sodium chloride, 0.3 M sodium citrate, pH 7.0) was poured over the gel, and was topped up as needed to ensure the gel was covered all the time. The vacuum pressure was set between 50 – 60 mbar and applied 1 h 30 mins. The position of the lanes on the gel were marked onto the membrane, and the membrane air dried and crosslinked using a UV CX-2000 crosslinker (UVP, Cambridge, UK) at 700 x 100 µJ/cm^2. The membrane was placed in a very clean hybridisation bottle and incubated with 10 ml hybridisation buffer (5 x SSC, 0.1 % (w/v) dextran sulphate, 1:20 liquid block (GE Healthcare)) with rotation for 15 mins. The probe was added to the bottle and incubated with rotation at 55 ^oC overnight.

The probe was made using the instructions found in the AlkPhos Direct Labelling and Detection System kit (GE Healthcare). Briefly, The DNA for the probe was diluted to 10 ng/µl and mixed with reaction buffer, labelling reagent, and cross linker. The final DNA concentration was chosen so that there was 5 – 10 ng DNA probe per ml hybridisation buffer. The reaction was incubated at 30 mins at 37 ^oC. Half (16 µl) of probe was added to the hybridisation bottle, the other half was stored in 50 % (v/v) glycerol at -20 ^oC until required.

The following day the membrane was washed 3 times for 15 mins with 50 ml primary wash buffer (2 M urea, 150 mM sodium chloride, 1 mM magnesium chloride, 50 mM sodium dihydrogen phosphate pH 7.0, 0.1% (w/v) SDS, 0.2 % (w/v) blocking reagent)

at 55 °C with rotation. After which, the membrane was washed twice for 10 mins in 100 ml secondary wash buffer (50 mM Tris, 100 mM sodium chloride, pH 10.0) at room temperature with shaking. The buffer was drained and the blot covered in 2 – 3 ml CDP – star™ detection reagent (GE Healthcare) for 2 – 5 mins. The detection solution was drained and the membrane sealed in plastic and taped in place in a film cassette. A piece of Hyperfilm™ ECL high performance chemiluminescence film (GE Healthcare) was placed on top of the membrane and the first film developed after 1 h. Blots were stripped by incubation with 0.5 % SDS at 60 °C with agitation for 2 h.

Method 2: Capillary transfer with a radiolabelled probe

The initial steps of the second Southern blotting method are almost identical to the previous method, differences are stated below. gDNA was digested overnight and size fractioned on an agarose gel, and stained with ethidium bromide (0.5 µg/ml). The capillary transfer was set up by pouring 20 x SSC into a Pyrex dish and by creating a bridge using a perspex insert covered in two strips of 3MM (Whatman, Kent, UK) filter paper wetted in 20 x SSC acting as a wick. The gel was placed onto the filter paper followed by a Hybond-N^+ membrane placed over the gel with care taken to remove any bubbles. On top of the membrane, 3 wet (in 20 x SSC) and 3 dry pieces of 3mm filter paper were added, followed by a large stack of paper towels and a weight. The apparatus set up is illustrated in Figure 2.5. The DNA was transferred overnight (8 – 18 h). The following morning the membrane was rinsed in 2 x SSC, dried and crosslinked as before in a Spectrolinker XL-1500 crosslinker (Spectronics Corporation).

The membrane was incubated in approximately 50 ml hybridisation buffer (5x SSPE (175.3 g/L sodium chloride, 27.6 g/L sodium dihydrogen phosphate, 7.4 g/L EDTA, pH 7.4 using sodium hydroxide), 9 % polyethylene glycol 6000, 0.5 % skimmed milk, 1% SDS, 0.1 % sodium pyrophosphate, 0.2 mg/ml herring sperm DNA) and prehybridised at 65 °C in a hybridisation oven with gentle agitation.

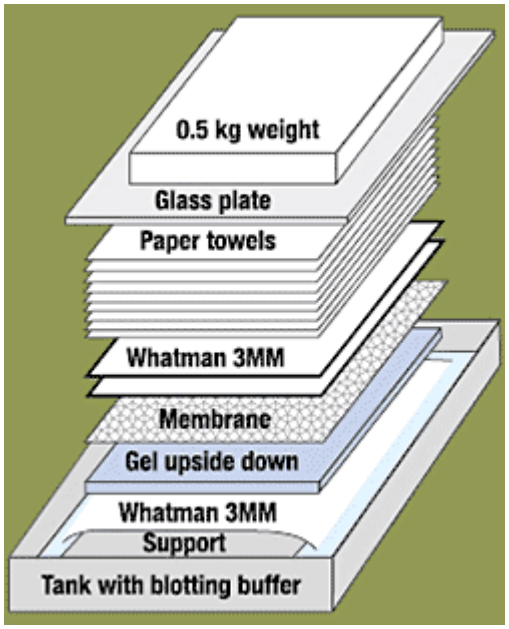


Figure 2.5: Southern blotting by capillary transfer. This figure demonstrates the set up a the Southern blot apparatus to transfer the DNA from agarose gel to the Hybond-N^+ membrane. Image was taken from the Thermo Scientific/Fermentas website[17].

The radiolabelled probe was made using a random primer labelling method. DNA (50 ng) was made up to 50 µl with DW and boiled for 2 mins, and placed on ice to cool. 5 µl 10x dNTP's (minus dCTP) and 10 µl 5x labelling buffer (Promega) were added and mixed, after which 0.5 µl Klenow fragment (10 U/µl) was added and mixed gently. 2.5 µCi of ^{32}P dCTP (Perkin Elmer) 6000 Ci/mmol was added, and the whole reaction mixture incubated at 37 ^{o}C for 30 mins. To the reaction mixture, approximately 5 µl dextran blue (in TE buffer) was added. Labelled probe was separated from unincorporated nucleotides by running through a Sephadex G50 column. The eluted probe was denatured by boiling for a further 10 mins followed by quenching to prevent re-annealing. Once cool, the probe was added to the membrane and incubated overnight and 65 ^{o}C.

The following day the buffer containing the probe was removed and the membrane washed in 5x wash buffer (5x SSC. 0.1 % SDS, 0.1% sodium pyrophosphate), washed twice in 3x wash buffer (3x SSC. 0.1 % SDS, 0.1% sodium pyrophosphate), followed by one wash in 1x wash buffer (1x SSC. 0.1 % SDS, 0.1% sodium pyrophosphate). A final rinse on 1x SSC was carried out and the membrane was sealed in plastic bagging

[17] http://www.fermentas.com/templates/files/tiny_mce/support_images/blotting.gif

which was placed into a cassette. A phosphor storage screen (Imaging Screen K, BioRad, Hertfordshire, UK) was placed over the membrane and stored overnight at room temperature. The phosphor screen was placed in a Molecular Imager FX (BioRad) and the Quantity One computer program (The Discovery Series, 1998, BioRad) used to visualise the blot.

2.3 Biochemical Techniques

2.3.1 Small scale protein expression

Small scale expression trials were set up to optimise protein expression conditions. For membrane proteins a 50 ml culture volume was grown in 250 ml conical flasks. For expression of protein domains 10 ml culture volume was used in 25 ml glass Universals. A 1:50 volume of overnight culture was used to inoculate the expression trial medium, and the appropriate antibiotic was added to maintain the expression plasmid. Cells were grown at 37 ^{o}C, 250 rpm for around 2 h until an OD_{600} of between 0.4 and 0.6, at which point inducer (IPTG or L-arabinose) was added at varying concentrations to induce protein expression. Different temperatures (after the initial 37 ^{o}C incubation), different medium (a list of the various media tested is given in Appendix 1) and expression/incubation times to optimise the expression conditions were all tested.

For expression using Arctic Express™ cells the procedure was slightly different, and was carried out according to the manufacturers' instructions. Briefly, a single colony was grown overnight in LB broth containing the plasmid selection antibiotic and gentamycin to maintain the plasmids carried by the cells. The overnight culture was used to inoculate LB without antibiotics, and the culture grown for 3 h at 30 ^{o}C. The culture was induced with IPTG, the temperature reduced to between 10 and 13 ^{o}C, and expression allowed to continue for 24 h.

To assess expression, either crude membrane extracts (see section 2.3.3) or cell lysate was applied to a denaturing Nu-PAGE gel (see section 2.3.5). NuPage® loading buffer was added to the samples, cell lysate was heated to 70 ^{o}C before loading on a gel. Membrane preparations were incubated at room temperature for 30 mins, aggregation of hydrophobic proteins has been known to occur after boiling of samples (Thomas and

McNamee, 1990). The protein concentration in the membrane was determined using a BCA protein assay (Pierce through Thermo Scientific), and a predetermined protein standard curve prepared with bovine serum albumin (BSA).

2.3.2 Large scale protein expression

Large scale expression was carried out in 6 x 5 L litre conical dimpled flasks filled with 2 L of the desired medium. Cells were grown to an OD_{600} of 0.4 – 0.5 and induced with the optimised concentration of IPTG, and supplemented with 20 ml 0.4 % histidine (filter sterilised) if a His-tag was fused to the desired protein. The flasks were incubated for the optimised expression time and temperature. Cells were harvested by centrifugation at 10 000 x *g* for 15 mins at 4 ^{o}C. The weight of the cell pellet was recorded and the pellet stored at -85 ^{o}C after resupension in 20 mM Tris, 0.5 mM EDTA, 10 % glycerol pH 7.5 to a final volume of 100 ml and stored at -85 ^{o}C.

2.3.3 Preparation of membranes

2.3.3.1 *E. coli* membranes using homogenisation

To prepare crude membranes, the Water Lysis Method was utilised. Induced culture (50 ml) was centrifuged at 1400 x *g*. The pellets were resuspended in 10 ml 0.2 M Tris-HCl pH 8.0, transferred to an Oakridge tube and placed on a Luckham 100 rotatest shaker for 20 mins at room temperature. Cell lysis was achieved by adding 4.85 ml sucrose buffer (1 M sucrose, 0.2 M Tris-HCl, 1 mM EDTA) at T_0. At $T_{1.5\ mins}$ 65 µl 10 mg/ml lysozyme made fresh in sucrose buffer was added, at $T_{2\ mins}$ 9.6 ml DW was added. The tube was again placed on the rotatest shaker for 20 mins at room temperature followed by centrifugation at 38700 x *g* at 4 ^{o}C for 15 mins. The periplasmic fraction in the supernatant was discarded. The pellet was resuspended in 30 ml DW using a dounce homogeniser, and left to stand at room temperature for 30 mins before being centrifuged as before. This supernatant fraction contains the cytoplasmic contents. The pellet was washed in 30 ml phosphate buffer (0.1 M sodium phosphate pH 7.2, 1 mM β-mercaptoethanol), by homogenisation followed by centrifugation as before. The wash was repeated, and the pellet containing the washed membranes resuspended in 150 – 300 µl phosphate buffer using a small dounce homogeniser. Membranes were stored at -20 ^{o}C.

2.3.3.2 *E. coli* large scale membrane preparation

To prepare membranes from which protein was to be isolated, a second method was used (Hope, 2005). The bacterial pellet from a large scale expression culture was passed twice through a cell disruptor (Constant Systems) at 25 Kpsi. The cell lysate was centrifuged at 10 000 x *g* for 30 mins at 4 °C to remove cell debris. The supernatant was aspirated and centrifuged at 131 000 x *g* for 2 h at 4 °C. The pellet containing the cell membranes was resuspended in the phosphate buffer (0.1 M sodium phosphate pH 7.2, 1 mM β-mercaptoethanol) using a small dounce homogeniser. Membranes were aliquoted in 0.5 ml volumes to microfuge tubes and snap frozen using liquid nitrogen. The tubes were strored at -85 °C.

2.3.3.3 *Aspergillus* crude plasma membrane preparation

A. nidulans 'germlings' (50 mg) (recently germinated conidia), previously prepared by SE Unkles, were thawed from -85 °C. These were mixed with 300 µl cold extraction buffer (10 mM sodium phosphate, 200 mM sodium chloride, 10 % glycerol, pH 7.0) and 300 µl glass beads. Cells were disrupted in a FastPrep-24 at 4.5 m/s for 20 s, and placed on ice. The mixture was centrifuged at 12000 x *g* for 1 min to remove the beads, and the supernatant was transferred to a fresh microfuge tube and centrifuged at 38700 x *g* for 45 mins at 4 °C. The pellet containing the *Aspergillus* plasma membrane was resuspended in 50 µl cold extraction buffer and stored at -85 °C.

2.3.4 Separation of insoluble and soluble proteins from *E. coli*

To determine if an expressed domain was soluble, the proteins were separated into the soluble and insoluble fractions. Two methods were used: (i) sonication, and (ii) homogenisation, nevertheless, the protocol used throughout was the same. 10 ml of culture was harvested and resuspended in 1 ml 50 mM Tris-HCl, 2 mM EDTA. Lysozyme was added to a final concentration of 100 µg/ml and 100 µl 1% triton-X 100 was added and mixed gently. This was incubated at 30 °C for 15 mins and placed on ice. For sonication the mixture was sonicated at a high output setting for a few seconds, and placed on ice. For homogenisation the mixture was homogenised in a small dounce homogeniser until the mixture lost viscosity. When the mixture was less viscous it was centrifuged at 12 000 x *g* for 15 mins. The supernatant containing soluble proteins was transferred to a fresh microfuge tube and the pellet resuspended

in the total original volume. Samples of these fractions were run on a NuPAGE (see Section 2.3.5.2) gel to determine which fraction contained the desired protein.

2.3.5 Protein gels

2.3.5.1 SDS PAGE

Polyacrylamide gel electrophoresis is used to separate proteins dependant on molecular weight. Gels consisted of a running gel and a stacking gel and were run in Tris-Glycine SDS running buffer (20 mM Tris, 192 mM glycine, 0.1 % (w/v) SDS, pH 8.3). The percentages of running gels used in this thesis were 7.5 % and 12.5 % and are given in Table 2.2. Running gel was mixed and poured into a 1 mm plastic cassette (Invitrogen), once set stacking gel was poured into the cassette and the appropriate comb was inserted. The gels were electrophoresed at 200 V for approximately 40 mins. Following electrophoresis, the gel was removed from the cassette and stained using Coomassie Brilliant Blue (CBB) (0.5 % coomassie blue in 227 ml methanol, 227 ml DW, 46 ml acetic acid) for approximately 1 h and repeatedly washed in DW with gentle agitation.

	7.5 % running gel	12.5 % running gel	5 % stacking gel
Buffer 1	4.95 ml	3.85 ml	
Buffer 2			2.5 ml
acrylamide	1.65 ml	2.75 ml	0.5 ml
TEMED	8 µl	8 µl	8 µl
APS (20 % (w/v))	35 µl	35 µl	25 µl

Table 2.2: Composition of SDS-PAGE gels. Buffer 1: 375 mM Tris, 0.1 % (w/v) SDS, pH 8.9, Buffer 2: 122 mM Tris, 0.1 % (w/v) SDS, pH 6.7. TEMED – Tetramethylethylenediamine.

2.3.5.2 NuPAGE®

The NuPAGE (Invitrogen) gels have been formulated to work with InVision™ His-tag In-Gel Stain (Invitrogen) (hereto referred to as His-tag stain). Gels were stained for the His-tag fused to the target protein without the need of western blotting. After the gel had run, the proteins were fixed in destain solution (900 ml ethanol, 900 ml DW, 200 ml acetic acid) for at least 1 h and washed twice in water before application of the His-stain. The His-tag stain (20 ml) was applied to the gel and incubated at RT on a rotatest shaker for 1 h, and washed twice in 20 mM phosphate buffer (20 mM sodium phosphate, pH 7.8) for 10 mins to remove excess stain and minimise any background.

The stain was visualised under a UV transilluminator and an image recorded. The gel was finally stained using SimplyBlue™ (Invitrogen) and latterly Instant Blue (Expedeon, Harston, UK) coomassie stains.

2.3.6 Western blotting

Protein samples of the same concentration were run on an SDS-PAGE gel, and soaked in transfer buffer (25 mM Tris-HCl pH 8.3, 192 mM glycine, 20 % methanol, 0.1 % SDS) for 20 mins. PVDF (GE Healthcare) or nitrocellulose membrane (Protran™, Whatman) and 3MM paper was cut to the same size as the gel. The PVDF membrane was rinsed in methanol and soaked in transfer buffer along with the 3MM paper. The blotting equipment (Mini Protean II System, BioRad) was assembled with the protein gel adjacent to the PVDF membrane. The blotting equipment was placed in an ice bath, and voltage applied at 100 V for 50 mins to transfer the protein to the membrane. The membrane was washed in TBST (0.9 % w/v sodium chloride, 10 mM Tris pH 7.4, 0.1 % v/v Tween 20) twice for 15 mins and incubated in TBS (0.9 % w/v sodium chloride, 10 mM Tris pH 7.4) with 5 % blocking agent (Amersham). Incubation with the block was carried out for 1 – 4 h at RT or overnight at 4 ^{o}C. The membrane was washed in TBST three times for 10 mins and incubated in TBS with 0.5 % blocking agent and appropriate antibody dilution. The antibody was incubated with the membrane for 1 – 4 h at RT or overnight at 4 ^{o}C. The membrane was washed four times in TBST, as a general rule twice for 5 mins and twice for 20 mins. If a secondary antibody was required the process was repeated. ECL Plus detection solution (Amersham) was prepared and 2 ml used to cover the membrane for 5 mins. The membrane was placed in plastic 'bagging' and exposed using a LAS-3000 Imager (Fujifilm, Bedfordshire, UK); blots were recorded and saved as a computer file. Latterly, membranes were exposed on High Performance Chemiluminescence films (GE Healthcare). The antibodies used and concentrations used at are given in Table 2.3.

Antibody	Conjugate	Raised in	Dilution WB	Dilution IFA	Source
Primary antibodies					
anti-V5	HRP		1:5000		Invitrogen
anti-GFP	HRP		1:5000		Santa Cruz
anti-GFP		mouse	1:5000	1:500	BD Biosciences
anti-PhoA	HRP		1:10000		Abcam
anti-His	HRP		1:5000		Invitrogen
anti-PfCRT		rabbit		1:600	Dr. P. Bray
anti-PfHT		goat		1:500	Prof. S. Krishna
anti-HA		mouse	1:2000	1:250 – 1:750	Roche
anti-BCKDH		rabbit	1: 10000		Eurogentec
Secondary antibodies					
anti-mouse	HRP		1:10000		Promega
anti-rabbit	HRP		1:10000		Promega
anti-mouse	alexa 488			1:2000	Invitrogen
anti-rabbit	alexa 594			1:2000	Invitrogen
anti-goat	alexa 594			1: 2000	Invitrogen

Table 2.3: Working concentrations of each of the antibodies used in this thesis.

2.3.7 Solubilisation of membrane proteins

Small scale solubilisation trails were set up to find the best detergent to extract the protein from the membrane. Detergents were trialled at 0.5, 1, 2 and 4 % (w/v) for their ability to extract protein from the *E. coli* membrane. Total membrane protein was added to a buffer solution (20 mM Tris (pH 8.0 at 4 ^{o}C), 10 mM imidazole, 800 mM NaCl, 10 % glycerol) containing detergent so the final protein concentration was 10 mg/ml. This protein-detergent mixture was incubated for 1 hour or overnight in an ice bath and stored at 4 ^{o}C, followed by centrifugation at 47800 x *g* for 1 h at 4 ^{o}C. The supernatant containing the solubilised proteins was transferred to a fresh microfuge tube and the pellet resuspended in the total original volume. Samples were run on a NuPAGE gel to analysed the solubilised proteins.

Large scale solubilisation involved incubating the membrane overnight in the desired detergent at the optimum concentration in solubilisation buffer (20 mM Tris (pH 8.0 at 4 ^{o}C), 10 mM imidazole, 800 mM NaCl, 10 % glycerol). Membrane extract was added to the solubilisation buffer to a final protein concentration of 10 mg/ml, the detergent was added to the mixture slowly and mixed well. The mixture was placed in an ice bucket and stored overnight at 4 ^{o}C. The following day the mixture was centrifuged at 47800 x *g* for 30 mins at 4 ^{o}C. The solubilised proteins were present in the supernatant and were purified.

2.3.8 Purification of solubilised membrane protein

The BD Talon IMAC resin to be used in the purification process was washed. 0.5 ml was taken and washed in 50 ml DW and centrifuged at 335 x g for 3 mins at 4 ^{o}C twice, and once in wash buffer (20 mM Tris (pH 8.0 at 4 ^{o}C), 10 mM imidazole, 400 mM NaCl, 5 % glycerol, [desired concentration of detergent]), and centrifuged at before. Solubilised protein was applied to Co^{2+} metal affinity resin (BD Talon, Clontech). A Co^{2+} column was chosen over Ni^{2+}. Co^{2+} binds less protein and the binding affinity is reduced compared to Ni^{2+}. However, the binding is more specific using Co^{2+} and the number of contaminating proteins eluted is reduced. The cell lysate and resin was incubated on ice to allow binding of the protein to the beads for 4 h. The resin was washed twice in wash buffer and centrifuged as before. The beads with protein bound were transferred to a 5 ml syringe blocked with 3mm paper and elution buffer (20 mM Tris (pH 8.0 at 4 ^{o}C), 200 mM imidazole, 400 mM NaCl, [desired concentration of detergent]) applied. Fractions were collected in 1.5 ml microfuge tubes, 0.25 ml at a time. After the elution buffer had run through the column, the syringe was blocked with a stopper and the fractions analysed for protein content using a Nanodrop spectrophotometer and OD_{280} for protein.

Fractions containing the protein were combined and concentrated to 100 µl using Vivaspin (Sartorius, Surrey, UK) centrifugation devices with molecular weight cut-offs stated in each of the chapters. Purified protein was stored at -85 ^{o}C. An overview of the purification procedure is illustrated in Figure 2.6.

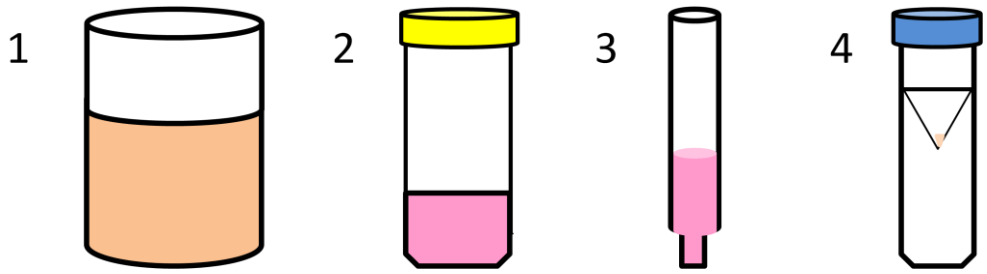


Figure 2.6. Protein purification protocol. Following cell disruption and membrane preparation, the cell lysate was solubilised in detergent overnight (1). The solubilised protein was bound to a Co^{2+} metal affinity column (2). The resin was washed and the protein was eluted from the column using imidazole (3). The resultant eluate was applied to a NuPAGE gel and concentrated (4). Source: A. Clarke.

2.3.9 Purification of proteins from inclusion bodies

Cells were harvested by centrifugation at 2000 x *g* for 5 mins. The supernatant was removed and cells resuspended in an equal volume of lysis buffer (50 mM Tris-HCl pH 8.0, 25 % sucrose (w/v), 1 mM EDTA) and 1/10 volume lysozyme (1 mg/ml in lysis buffer) added. The cell suspension was incubated on ice for 30 mins and frozen at -20 ^{o}C and thawed in water at room temperature. One tenth volume DNase solution (100 mg/ml in 100 mM $MgCl_2$, 10 mM $MnCl_2$) was mixed in and the cells incubated at room temperature until the suspension became less viscous. An equal volume of detergent buffer (0.2 M sodium chloride, 1 % deoxycholic acid (w/v) 1 % nonidet P-40 (v/v)) was added to the cell solution. The solution was centrifuged at 5000 x *g* for 10 mins. The supernatant was removed carefully so as not to dislodge the white inclusion body pellet or membrane proteins which formed an upper jelly-like layer. To remove the membrane proteins, the sample was washed in 1 % Triton X-100/1 mM EDTA until the jelly like phase was no longer be observed. The remaining pellet was analysed by SDS-PAGE. This method was adapted from a method available from 'www.molecularinfo.com'[18].

2.4 Structural and Functional Determinations

2.4.1 Formate transport

2.4.1.1 Radioactive tracer assay

To measure the direct uptake of formate by cells a tracer assay was used with ^{14}C formic acid (reference activity, 7.8 MBq). Twenty millilitres of TGYEP medium was inoculated with 0.2 % overnight, aerobic LB culture from a single colony, and supplemented with 0.5 % glucose and the appropriate antibiotic. Cultures were incubated in a static incubator at 37 ^{o}C for 22 h. The OD_{600} of the culture was recorded, the cells pelleted and resuspended in 20 ml assay buffer (Section 3.4.2.2). To the resuspended culture, a specific concentration (Section 3.4.2.2) of a sodium formate and ^{14}C formic acid mixture was added (450 µl 20 mM sodium formate and 10 µl formic acid (reference activity, 7.8 MBq)), the cultures were placed in a water bath set at 37 ^{o}C. At various time points a 1 ml sample volume was filtered through a

[18] http://www.molecularinfo.com/MTM/G/G2/G2-2/G2-2-2.html. Accessed on 10 August 2012.

vacuum filtration device with a 0.45 µm cellulose membrane filter (pre-wetted with 15 ml assay buffer (without ^{14}C formic acid). The membrane filters were washed three times with 15 ml assay buffer (without ^{14}C formic acid) to remove traces of ^{14}C formic acid from the membrane and bacteria surfaces. The washed membrane was placed in a scintillation vial and covered with 2.5 ml liquid scintillation cocktail (Optiphase HiSafe 3, Perkin Elmer, MA, USA). Two 5 µl and one 10 µl aliquot of the formate/fomic acid mixture were placed in scintillation vials and covered in 2.5 ml liquid scintillation cocktail to act as standards.

2.4.1.2 Hypophosphite assay

The method for characterising the uptake of formate, through its toxic analogue hypophosphite, was adapted from one described by Suppmann and Sawers (1994). A single colony was picked and grown in 5 ml medium (Appendix 2) with and without 100 mM hypophosphite in a static incubator to create an anaerobic environment, at 37 ^{o}C for 22 h. After the growth period a small aliquot was taken from each culture and the OD_{600} noted to monitor bacterial cell density.

2.4.1.3 Indirect assay using a β-galactosidase reporter

The indirect method for investigating formate uptake was an adaptation of the one used by Falke *et al* (2009). A β-galactosidase gene (*lacZ*) has been fused downstream of the *fdhF* promoter, the promoter is active only when formate is found in the cell. A single colony was grown aerobically in 5 ml LB + antibiotic overnight, 400 µl of overnight culture was used to inoculate 22 ml TGYEP medium pH 6.7 and supplemented with 50 mM formate. Cells were grown for 4 h at 37 ^{o}C in a static incubator to provide anaerobic conditions, and IPTG added to induce protein expression. 500 µl samples were taken after 2, 3, and 4 h to carry out a β-galactosidase assay. The OD_{600} of the 500 µl culture sample was recorded using a nanodrop spectrophotometer. The cells were centrifuged at top speed for 2 mins on a bench top centrifuge, and the supernatant discarded. Cells were resuspended in 75 µl lysis buffer (250 mM Tris pH 7.5, 0.2 % Triton X-100, and 4 mg/ml sodium deoxycholate) and 225 µl buffer A (100 mM sodium dihydrogen phosphate, 10 mM potassium chloride, 1 mM magnesium sulphate, 50 mM ß -mercaptoethanol) was added. The sample was incubated at 37 ^{o}C for 5 mins. Seventy five microlitres of ONPG (*ortho*-Nitrophenyl-β-galactoside) (4 mg/ml in 100 mM sodium dihydrogen

phosphate, pH 7.5) was added to initiate the reaction. The time for the reaction to occur was set to 5 mins to prevent any confusion in the time of the colour change. After 5 mins 150 µl of 1 M sodium carbonate was added to stop the reaction. The cell debris was collected by centrifugation as before, and the OD_{420} of the supernatant recoded using a nanodrop spectrophotometer. The β-galactosidase activity was calculated in Miller units using the following equation:

$$Miller\ units = \frac{OD_{420} \times 1000}{time(\min) \times OD_{600} \times vol.cells(ml)}$$

2.4.2 Nitrite transport

A nitrite transport assay was carried out and measured the depletion of nitrite from the medium. An overnight LB aerobic culture from a single colony was used to inoculate 20 ml cultures of JC minimal salts medium (JC for Jeff Cole) and supplemented with 20 mM potassium nitrate, 0.4 % glucose and appropriate antibiotic (Cole *et al*, 1974 and Jia *et al*, 2009). Cultures were transferred to a static 37 ^{o}C incubator. Strains carrying plasmids were incubated for 4 h prior to induction with 1 mM IPTG, and incubated for a further 22 h.

To prepare a nitrite standard curve 0 – 100 µM sodium nitrite solutions in 50 mM MOPS buffer (pH 8.5) were prepared and incubated at 37 ^{o}C for 30 mins. 40 µl of each solution was added to a 96 well plate in triplicate. 40 µl 1% sulphanilamide (in 25 % HCl) followed by 40 µl 0.2 5 N-1-naphthylethylenediamine dihydrochloride (NED) were added and mixed. The reaction was incubated at RT for 10 mins. The OD_{540} was read on a nanodrop spectrophotometer and values averaged to construct the standard curve.

The OD_{600} of the anaerobic cultures was taken and the number of cells/ml medium was deduced (given that an OD_{600} of 1.000 is equal to 8 x 10^8 cells/ml). 1 x 10^9 cells were washed in 0.5 ml 50 mM MOPS buffer to remove any residual nitrate; heat killed cells placed on a heat block at 72 ^{o}C for 10 mins. Washed cells were resuspended in 5 ml assay buffer (50 mM MOPS buffer pH 8.5, 10 mM glucose), and 100 µM sodium nitrite added at T_0. The cultures were placed in a 37 ^{o}C static incubator and samples taken over a period of 2 h. At T_0, and at half hour intervals up to 2 h, 40 µl of culture was

placed, in triplicate, in a 96 well plate. To each of the samples in the 96 well plate, 40 µl of 1 % sulphanilamide and 40 µl of 0.2 % NED were added, and the reaction incubated for 5 mins. The OD_{540} was recorded on a nanodrop spectrophotometer, and an average taken. The standard curve was used to infer the concentration in µM of nitrite in the medium.

2.4.3 PhoA assay

The PhoA assay was performed as described by Manoil (1991), and the cell growth as described by Rapp *et al* (2004) although minor alterations have been made. Cells from an overnight growth culture were diluted into fresh medium, grown until an OD_{600} between 0.4 – 0.6, and induced with 0.1 % arabinose. After induction cells were incubated for a further 4 h. A 1 ml sample was harvested by centrifugation at 16000 x *g* for 1 min and washed once with cold 10 mM Tris pH 8.0, 10 mM magnesium sulphate. The cell pellet was resuspended in 1 ml cold 10 mM Tris pH 8.0 and the OD_{600} recorded. Cell suspension (0.1 ml) was added to 0.9 ml assay buffer (1 M Tris, pH 8.0, 0.1 mM zinc chloride) and 1 ml assay buffer was retained as a blank. To each sample including the blank, 50 µl 0.1 % SDS and 50 µl chloroform was added, mixed by vortexing, incubated at 37 ^{o}C for 5 mins, and placed on ice for 5 mins. 0.1 ml of 0.4 % p-nitrophenyl phosphate (pNPP) (in 1 M Tris) was added to the reaction and incubated at 30 ^{o}C for 90 mins (standardised for all reactions), until the mix had turned from colourless to a pale yellow. To stop the reaction 120 µl 0.5 M EDTA, 1 M potassium dihydrogen phosphate in a 1:5 ratio was added. The values for were recorded. The activity of the phosphatase was calculated using the following equation:

$$PhoA\ activity\ (units) = \frac{OD_{420} \times 1000}{time(\min) \times OD_{600} \times vol.cells(ml)}$$

2.4.4 GFP assay

The GFP assay was described by Rapp *et al* (2004), however minor changes have been made. The experiments were carried out in triplicate and repeated at least 5 times. Cells containing plasmid with a C-terminal GFP tag on the specific insert were grown overnight in LB containing the appropriate antibiotic. They were subcultured 1:50 dilution into 5 ml fresh LB medium containing the appropriate antibiotic, and grown at 37 ^{o}C for 1.5 – 2 h until cells were in mid-exponential phase. At this point, expression

of the GFP fused insert was initiated by the addition of 0.05 mM IPTG and cells allowed to grow for a further 3 h. The OD_{600} was measured and recorded. 1 ml of the cell culture was harvested by centrifugation at 2800 x *g* for 1 – 2 mins and resuspended in 1 ml GFP assay buffer (50 mM Tris pH 8.0, 200 mM sodium chloride, 15 mM EDTA). The cell suspension was incubated for 30 mins at room temperature. Approximately 300 µl of the cell suspension was transferred to a 3 x 3 mm microcuvette and read at excitation 490 nm and emission 520 nm, on a VersaFluor™ Fluorometer System (Bio-Rad, Hemel Hempstead, UK).

2.5 *Plasmodium falciparum* Culture

The work detailed below was carried out in the laboratory of Professor S. Müller at the University of Glasgow.

2.5.1 Maintenance of parasite cultures

The *in vitro* culture of *Plasmodium falciparum* intraerythrocytic stages was first described by Trager and Jensen (1976), and the method was modified as described below. *P. falciparum,* strains 3D7 (The Netherlands) and D10 (Papua New Guinea) were cultivated in complete RPMI medium (15.9 g/L RPMI 1640 (containing L-glutamine and 25 mM HEPES) (Invitrogen), 0.5 % Albumax II (Invitrogen), 0.1 % (w/v) sodium bicarbonate, 20 µg/ml gentamycin, 200 µM hypoxanthine, pH 7.4). The medium was filter sterilised using a 0.22 µm stericup filter unit (Millipore) and stored at 4 °C. The medium was prewarmed to 37 °C before use. Cultures (10 ml) were propagated with a 5 % haematocrit (HCT) in a 50 ml TC (non-vented) culture flasks (Greiner, Stonehouse, UK) at 37 °C. Parasites were maintained under reduced oxygen conditions (1 % oxygen, 3 % carbon dioxide and 96 % nitrogen).

Parasitaemia was maintained between 0.5 – 2 %, cultures were diluted and supplied with fresh medium and red blood cells (RBC) accordingly. Parasitaemia was determined by a thin blood smear stained with Giemsa stain solution (BDH). Air dried slides were fixed with 100 % methanol for about 1 min, followed by staining in surplus Giemsa (diluted 1:10) for 5 – 10 mins. The slide was rinsed with water to remove excess Giemsa, and air dried before analysis by light microscopy with an objective magnification of 100 fold. The parasitaemia was recorded by counting the number of

infected and uninfected erythrocytes. At least 1000 RBCs were counted. Parasitaemia is expressed as a percentage of infected erythrocytes (IE).

2.5.2 Blood

Human blood was received weekly from the Blood Transfusion Service. Briefly, packed erythrocytes were washed 3 – 4 times in blood wash medium (RPMI medium described above without the addition of Albumax II) and the white blood cells or 'buffy coat' and serum removed by centrifugation at 600 x *g*, 4 °C for 10 mins. After washing, a small aliquot was taken, mixed with growth medium and incubated at 37 °C for two days to check for growth of any contaminants. Blood was stored at 4 °C is v/v blood wash medium, and was stored for no longer than four weeks after the bleed date.

2.5.3 Drug selectable markers

Drug selectable markers (DSM's) were used to select for parasites containing transfected plasmids. 5-fluorocytosine (5-FC) was made at 4 mM and was stored at room temperature. While in use, WR99210 and blasticidin were stored at 4 °C, blasticidin is stable for 1 week at 4 °C and was refreshed accordingly. WR99210 was dissolved in dimethyl sulphoxide (DMSO) to a concentration of 20 mM and stored at -80 °C. Every four weeks a working solution was made in blood wash medium at a concentration of 20 µM. Blasticidin powder was dissolved in HEPES buffer (pH 7.8) to a final concentration of 10 mg/ml and stored in 50 µl aliquots at -20 °C. Table 2.4 presents the working concentrations of each drug.

Drug	Working concentration
WR99210	0.0125 nM
Blasticidin	0.025 µg/ml
5-FC	1 µM

Table 2.4: Concentrations of DSM used in the culture of *P. falciparum*.

2.5.4 Preparation of *P. falciparum* stabilates

Stabilates were prepared using cultures of mixed stage parasites containing ≥ 2 % ring stages. A culture was centrifuged at 380 x *g* for 5 mins at 4 °C with a low deceleration, and the supernatant aspirated. The RBC pellet volume of fresh cold medium was added to the RBCs. Twice the RBC pellet volume of freezing solution (PBS (phosphate buffered saline) + 30 % v/v glycerol) was added and mixed gently.

Aliquots of 600 µl were transferred to 1 ml cryovials and cooled to 4 °C before the vials were frozen in liquid nitrogen.

2.5.5 Thawing of *P. falciparum* stabilates

Stabilates were removed from liquid nitrogen and placed in a 37 °C incubator until thawed. One tenth the original volume of cells of thawing solution 1 (12 % w/v sodium chloride) was added slowly to the falcon tube and mixed, followed by incubation for 5 mins at room temperature. Ten times original volume of thawing solution 2 (1.8 % w/v sodium chloride) was added slowly and mixed, after which, the cell suspension was incubated at room temperature for 5 mins. Finally, ten times original volume thawing solution 3 (0.9 % w/v sodium chloride, 0.2 % w/v glucose) was added slowly and mixed, followed by incubation at room temperature for 5 mins. The samples were centrifuged at 1500 x *g* for 5 mins. The supernatant containing the freezing and thawing solutions was discarded and the parasites transferred to a new flask containing 9.3 ml medium and 200 µl fresh blood.

2.5.6 Synchronisation of *P. falciparum* RBC stages

To synchronise the parasites to the same phase of growth, cultures were centrifuged at 380 x *g* at room temperature for 5 mins with a low deceleration. The medium was aspirated and the parasitised RBCs resuspended in 5 volumes of 5 % sorbitol (w/v) in phosphate buffer (10 mM potassium phosphate, pH 7.2). The parasite mixture was placed in a 37 °C incubator for 5 mins and centrifuged as previously described. The supernatant was aspirated and the pellet resuspended in 10 ml RPMI complete medium, gassed and returned to the 37 °C incubator (Lambros and Vanderberg, 1979).

2.5.7 *P. falciparum* transfection

Prior to carrying out the transfection 50 – 100 µg of plasmid DNA was ethanol precipitated and subsequently resuspended in 30 µl of filter sterilised MQ water (see section 2.2.6). The DNA was stored at -20 °C until use in the transfection. A synchronised culture of early ring stage parasites was required for transfections, with a parasitaemia of no more than 5 %. As the parasites required 4 – 6 weeks continuous growth to recover from the transfection, fresh human RBCs were used in the transfection. Transfections were carried out in duplicate on separate days.

Transfections were carried out using the method of Wu *et al* (1995) and Crabb *et al* (2004). A 10 ml parasite culture was centrifuged at 4 °C, 380 x *g* for 5 mins, the supernatant aspirated and the infected RBC (iRBC) pellet washed in 5 ml ice cold Cytomix (120 mM potassium chloride, 0.15 mM calcium chloride, 2 mM EGTA, 5 mM magnesium chloride, 10 mM potassium dihydrogen phosphate/dipotassium hydrogen phosphate, 25 mM HEPES, pH 7.6). The washed iRBCs were stored on ice. Three hundred and seventy microlitres ice cold Cytomix was added to 30 µl of plasmid DNA, and 200 µl infected RBCs added with gentle but thorough mixing. The mixture was placed in a 2 mm electroporation cuvette (BioRad). Electroporation carried out at 310 V and 950 µF using a Gene Pulser Xcell (BioRad). After electroporation, the mixture was immediately added to a culture flask, containing 300 µl fresh RBCs and 8 ml pre-warmed RPMI 1640 complete medium. Parasites were gassed and returned to the 37 °C incubator. After 6 h, the medium was aspirated and replaced with fresh medium. The following day, the medium was again removed and replaced and the selection drug either WR or Blasticidin (section 2.5.5.2) was added to the flask at the appropriate concentration, this was repeated daily for 10 days. After 10 days medium was replaced every second day for up to 8 weeks. One hundred microlitres fresh RBCs were added to cultures once a week. To assess whether transfected parasites started to appear, Giemsa stained thin smears were prepared twice a week. When parasites began to appear in thin blood smears, stabilates were made (section 2.5.3). When parasitaemia was greater than 5 % in a 50 ml culture genomic DNA was extracted to analyse the genotype.

2.5.7.1 Drug cycling

To facilitate the selection of parasites where the transfected plasmid has integrated into the desired gene locus, a 'drug cycling procedure' was employed. Transfected parasite lines were selected with the positive DSM until the parasitaemia reached ≥ 5 %. Following the production of stabilates and isolation of gDNA, the selection drug was removed to encourage loss of the transfected plasmid. After 3 weeks the selection drug was re-applied until the parasite culture reached ≥ 5 % parasitaemia, at which point more stabilates and gDNA were made. The drug cycling was repeated multiple times and is depicted in Figure 2.7. In each of the sections where drug cycling has taken place, this diagram is used again.

The negative selectable marker 5-fluorocytosine (5-FC) was applied to those parasites lines that were previously transfected with either pCC1 or pCC4 knock-out plasmids. The presence of the *cd* chimera gene on the transfection plasmid leads to death of cells containing the plasmid when 5-FC is added to the medium. The chimeric CD protein converts 5-FC to the toxic 5-fluorouracil and 5-fluorouridine monophosphate. which causes an inhibition of RNA synthesis and the enzyme thymidylate synthase (Duraisingh *et al*, 2002 and Maier *et al*, 2006).

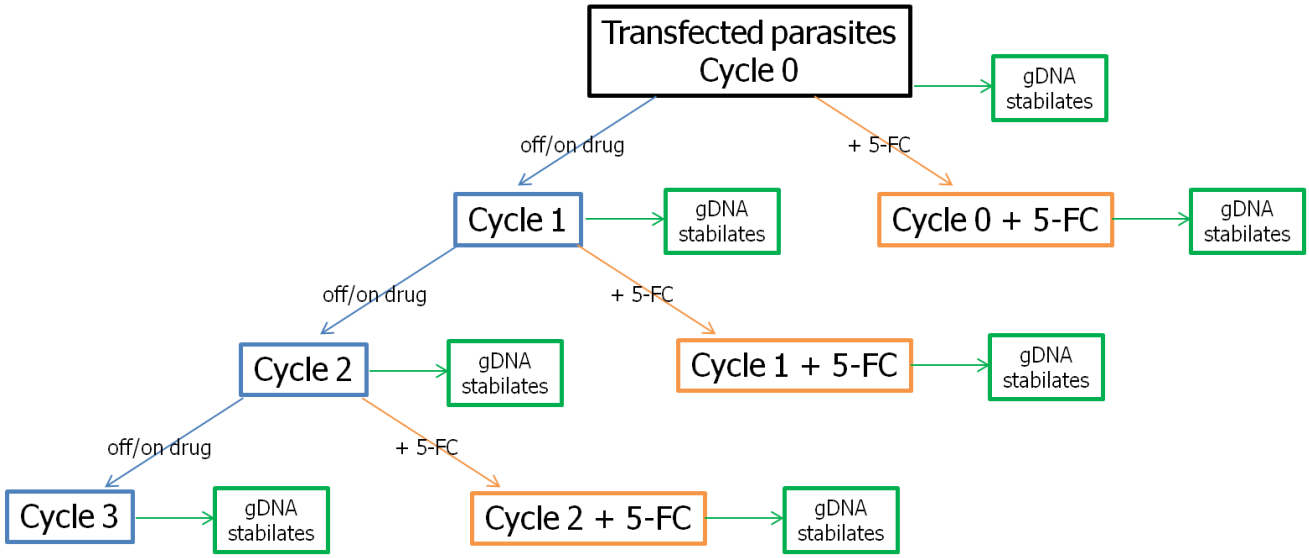


Figure 2.7: Overview of the drug cycling process. Parasites after transfection, growing on selectable marker (either WR 99210 or Bla) are termed cycle 0 (black box). Drug cycling of transfected parasites is carried out to select for integration of plasmids into the parasites' genome. Off/on drug cycling (either WR 99210 or Bla) is indicated by blue boxes/arrows and treatment with 5-FC is indicated in orange. At each stage gDNA is taken for analysis using PCR and/or Southern blot, and stabilates are made for the preservation of the parasite line. This is indicated by a green box/arrow combination

2.5.8 Saponin lysis of *P. falciparum* infected RBCs

Parasites from a 25 or 50 ml culture were resuspended in complete medium. For generation of gDNA 0.2 % saponin (v/v) was added to the culture and for isolation of parasite proteins 0.1 % saponin was added, followed by incubation at 4 °C for 10 mins. Subsequently the cells were centrifuged at 2000 x *g* for 5 mins at 4 °C with a low deceleration, followed by washes with PBS and transferred to a microfuge tube. The centrifugation was repeated at 2300 x *g* for 3 mins at 4 °C. The wash was repeated until the PBS was clear of any haemoglobin. For protein extraction 0.1 % saponin

(w/v) was used. The final pellets were stored at -80 °C until use (Umlas and Fallon, 1971).

2.5.9 gDNA extraction

2.5.9.1 Kit

Parasite gDNA was extracted using a Qiagen QIAamp DNA mini kit. A 50 ml culture was lysed using saponin (section 2.5.6) and resuspended in 200 µl sterile PBS. DNA extraction was carried out using the Blood or Body Fluid Spin Protocol with two minor changes. Parasite extracts were incubated at 56 °C for 30 mins to fully lyse the parasites and elution of gDNA was carried out twice using 100 µl of MQ water, and stored at 4 °C.

2.5.9.2 Phenol extraction

Saponin lysis was carried out as described (Section 2.5.6) and the parasite pellet resuspended in resuspension buffer (1200 µl SDW, 200 µl 10xTN9 (500 mM Tris, 2 M sodium chloride, pH 9) 400 µl 0.5 M EDTA, 40 µl 50 mg/ml proteinase K), after which 200 µl 10 % SDS (w/v) was added to prevent precipitation of solutions. The cell suspension was mixed in a 50 ml falcon tube and incubated at 50 °C with agitation overnight. The following day, 1 ml phenol was added and the sample mixed on a spinning wheel for 1 h at room temperature, after which the sample was centrifuged at 3000 x *g* for 10 mins. The upper phase was collected and transferred to a new 15 ml falcon tube, the phenol extraction was repeated. The upper phase was again transferred to a new 15 ml falcon tube, and 66.7 µl (1/30 volume) 3 M sodium acetate and 1200 µl (0.6 volume) isopropanol added. The tube was inverted vigorously until the DNA formed a white precipitate. The precipitate was removed to a microfuge tube containing 200 µl TE buffer and incubated at 4 °C overnight.

2.5.10 Protein extraction

To separate the soluble and insoluble (containing membranes) fractions of the parasites, a culture was lysed with saponin (as above) and the pellet frozen at -80 °C. To the frozen pellet and equal volume of lysis buffer (1% Triton X-100, 5 µg/ml RNaseA, 1 mM PMSF, 1 mM benzamidine, 20 µM leupeptin, 10 µM E64, 4 µM pepstatin A, 2 µM 1, 10-phenanthroline, 250 µg/ml DNase in PBS) was added. The pellet was

thawed at room temperature and the buffer mixed with cells. The parasite cell suspension was frozen on dry ice and thawed at 37 °C. The freeze/thaw process was repeated three times. After the final thawing the resuspension was incubated on ice for 20 mins and centrifuged at full speed on a benchtop centrifuge for 20 mins. The supernatant was transferred to a fresh microfuge tube and kept on ice. The pellet containing the insoluble fraction was resuspended in lysis buffer and the pellet homogenised. The homogenised pellet incubated on ice with gentle agitation for 30 mins, and centrifuged as before. This was repeated twice. Proteins were quantified using the BCA assay.

2.6.11 Microscopy

2.6.11.1 Live cell imaging

Parasite culture (500 µl) at a high parasitaemia (> 4 %) was placed in a microfuge tube, and 5 µl of 0.5 µg/ml Hoechst 33258 added. Cells were incubated for 5 mins at 37 °C. Cells were centrifuged at 200 x *g* for 30 s and washed in 500 µl medium to remove any excess DNA stain and centrifugation repeated. Five hundred microlitres fresh RPMI complete medium was added to the microfuge tube without disturbing the pellet. A small volume of packed RBCs was mixed with a little of the medium and placed on a slide and a cover slip placed on top.

Cells were visualised using an Axioskop (Zeiss) or Delta Vision (Applied Precision) microscope, at a 100x magnification. To visualise the Hoechst 33258 nuclear stain a DAPI filter was used. To visualise proteins tagged with GFP a GFP or FIT-C filter was used.

2.6.11.2 Immunofluorescence assay

Immunofluoresence assays were carried out using the method of Tonkin *et al*, (2004). Parasite culture (5 – 10 µl of ≥ 5% parasitaemia, 2.5 % HCT) was placed into a chamber slide pre-treated with poly-L-lysine, and allowed to air dry in a laminar flow hood. One hundred microlitres of a 4 % para-formaldehyde + 0.0075 % gluteraldehyde premade solution was added to each chamber and left for 10 – 30 mins to fix the cells. After washing three times with 250 µl PBS the cells were premeabilised with 100 µl 0.1% Triton X-100 for 10 mins. Following another wash step (three times

with 250 µl PBS) 100 µl 0.1 mg/ml sodium borohydride in PBS was added and incubated for 10 mins. Again the slide was washed three times with 250 µl PBS. The samples were blocked overnight with 3% BSA in PBS at 4 ^{o}C with humidity. The following day the slide was washed with 250 µl PBST (PBS + 0.05% Tween 20). The primary antibody was diluted in 3% BSA in PBS and 200 µl added to each chamber for 1 h at room temperature with agitation. The slide was washed in PBST three times. The secondary antibody was diluted in 3 % (w/v) BSA in PBS, 200 µl added to each chamber and incubated for 1 hour at room temperature with agitation. Prior to the final washes, 3 µl of a 1:10 dilution of Hoechst 33258 (0.5 µg/ml in 3 % (w/v) BSA) was added to each chamber and left for 1 min. The samples were washed three times with PBS + 0.05 % Tween 20. The plastic well divider was removed and the slide mounted with 500 µl 2.5% DABCO (1,4-diazabicyclo[2.2.2]octane), 50 % (v/v) glycerol in PBS. A cover slip was added, ensuring no bubbles, and sealed with black nail polish.

Cells were visualised using an Axioskop (Zeiss) or Delta Vision (Applied Precision) microscope, at a 100x magnification. To visualise the Hoechst 33258 nuclear stain a DAPI filter was used. To visualise proteins tagged with GFP or antibodies conjugated with AlexaFluor488 a GFP or FIT-C filter was used, and a TRIT-C or Alexa594 filter for antibodies conjugated with AlexaFluor594.

2.6 Biocomputational Analyses

2.6.1 Alignments and homology searches

There is a single annotated formate-nitrite transporter putative gene predicted in PlasmoDB. In PlasmoDB, the genomic structures of protein coding genes were predicted using computer based programs and manual curation (Gardner *et al,* 2002a). Protein coding genes were annotated using electronic means, many have a putative function and some no function at all (Kissinger *et al*, 2002). More information on the sequencing and annotation of genes in *P. falciparum* can be found in Hall *et al* (2002), Hyman *et al* (2002) and Gardner *et al* (2002b). Protein sequences from *P. falciparum* and *E. coli* were used as queries in order to identify genes in other organisms (bacteria, parasites, etc) that encode potential orthologues to the FNT proteins. Protein sequences retrieved from the PlasmoDB database (*Plasmodium*) and NBCI database (*E. coli*) were submitted to BLASTP searches of the online PlasmoDB

database for *Plasmodium* orthologues and through the NCBI database non-redundant protein sequences (nr) for other organisms. The top-scoring sequence for a particular organism was selected as the putative FNT orthologue, unless the protein had been annotated as a FNT protein or putative FNT protein (including FocA, NirC, FDHC and transporter annotations). Information on the various BLASTP searches performed in this thesis can be found in Appendix 5.

Alignments of multiple amino acid sequences were created using MAFFT (Katoh *et al*, 2002) version 6, the L-INS-i mode was used unless stated otherwise, for the phylogenetic analysis. For the alignments in the thesis an online MAFFT[19] program was used and the ClustalW output mode was selected. The MAFFT program, as found on the EBI website, implements a selection of multiple alignment methods as described on the MAFFT website (http://mafft.cbrc.jp/alignment/software/).

2.6.2 Membrane topology analyses

Hydropathy plots were used to estimate the number of transmembrane domains (TMDs) in a given amino acid sequence, and to identify the likely locations for the TMDs. The computer programs used for membrane topology analysis are listed below and their URLs are given in Table 2.5.

Program	URL
TMHMM Server v 2.0	http://www.cbs.dtu.dk/services/TMHMM-2.0/
HMMTOP version. 2.0	http://www.enzim.hu/hmmtop/
SOSUI engine ver. 1.11	http://bp.nuap.nagoya-u.ac.jp/sosui/
TMpred	http://www.ch.embnet.org/software/TMPRED_form.html
TopPred 0.01	http://mobyle.pasteur.fr/cgi-bin/portal.py?form=toppred

Table 2.5: Membrane protein topology prediction programs.

TMHMM predicts the number of transmembrane helices in a protein sequence given the amino acid properties, and gives the likelihood of these being true TMDs using other properties such as length of predicted TMD. The overall structure is used to determine the probability of the N-terminus being on the cytoplasmic side (Möller *et al*, 2001). HMMTOP (hidden markov model for topology prediction) is software used for the prediction of transmembrane helices and the topology of proteins given the protein sequence (Tusnády and Simon, 1998). TMpred predicts membrane-spanning regions

[19] http://www.ebi.ac.uk/Tools/msa/mafft/. Accessed on 10 August 2012.

and the overall orientation of a given amino acid sequence using statistical analyses of known transmembrane proteins present in the TMbase database (Hofmann and Stoffel, 1993). SOSUI (Mitaku *et al*, 2002) and TopPred (von Heijne, 1992) are also membrane topology and transmembrane domain prediction programs.

2.6.3 Protein modelling

Data gained from hydropathy plots and visual analysis were used to generate the data that was combined to produce a 2D model of the proteins.

The 3D protein model in this thesis was created using I-TASSER v1.1[20] from the Zhang Lab (University of Michigan, MI, USA). The program is based on multiple-threading alignments by LOMETS and iterative TASSER assembly situations (Zhang, 2008). Models retrieved with the highest confidence score (indicating the significance of alignment) were selected as the best models.

Conserved residues were identified through multiple alignments and those deemed to be important from two of the FocA structure papers (Wang *et al*, 2009 and Waight *et al*, 2010) were mapped onto the corresponding 3D structure for visualisation of possible binding interactions using PyMol v. 1.3 (Schrodinger, Camberley, UK).

2.6.4 Phylogenetic tree reconstruction

Sequences to be used in the phylogenetic tree were downloaded from NCBI and PlasmoDB, and aligned using MAFFT as detailed in section 2.6.1. The alignment was submitted to ModelGenerator (Keane *et al*, 2006) and the Bayesian information criterion (BIC) selected to choose a phylogenetic substitution model for the data. The number or Gamma categories selected in ModelGenerator was set as four. Each gene ID entry was edited give a less than 9 letter sequence identifier in order to make them suitable for input into the PhyML application. The maximum likelihood phylogeny was created by inputting the alignment file to PhyML 3.0[21] (Guindon *et al*, 2010), the substitution model was set to LG as chosen by ModelGenerator and type of tree improvement was set to SPR & NNI. The data were bootstrapped and set to 250 replicates. Output from PhyML was visualised using TreeView v 1.6.6 (Page, 1996).

[20] http://zhanglab.ccmb.med.umich.edu/I-TASSER/. Accessed on 25 November 2011.
[21] http://atgc.lirmm.fr/phyml/. Accessed on 10 August 2012.

2.6.5 Statistical analyses

Statistical analyses were carried out using SPSS (PASW Statistics, now IBM, Portsmouth, UK). To compare the means of data sets the one-way ANOVA (analysis of variance) was used. A one-way ANOVA makes the assumption that the variances of the groups being tested is similar. Therefore a homogeneity of variance' test was chosen. This tests for similar variences in the groups being tested, if the test returns a significant value, there is variation in the variences of the groups and a one-way ANOVA is not the correct test to use. If the result is not significant, the homogeneity of variences has been met, and the one-way ANOVA can be carried out. *Post hoc* tests were selected to compare each of the data sets by ANOVA, this was set to Bonferroni comparisons, the most stringent method in SPSS to correct for problems that could arise when comparing many data sets. If data did not conform to parametric assumptions a Kruskal-Wallis test was used instead of an ANOVA.

2.6.6 Subcellular localisation prediction

To investigate the possible cellular location for a specific protein, a selection of computer programs were used. The subcellular prediction computer programs used and their URLs are given in Table 2.6.

The PlasmoAP prediction program uses an algorithm to identify protein sequences containing an N-terminal amino acid sequence predicted to contain a signal and transit peptide consistent with apicoplast targeted sequences from *P. falciparum*. The transit peptides contain more basic (histidine, lysine and arginine) than acidic (aspartate and glutamate) residues (Foth *et al*, 2003). The PATS predication software uses neural network analysis to predict the transit sequence from an amino acid sequence (Zuegge *et al*, 2001), it is suggested by the authors that a secondary analysis is carried out on sequences predicted to be targeted to the apicoplast. PlasMit is used to predict mitochondrial transit peptides in *P. falciparum*. PlasMit consists of four neural networks, and output is based on all four. The networks consider amino acid composition and the physicochemical properties of the residue sequence (Bender *et al*, 2003). iPSORT is used to predict signal peptides, mitochondrial targeting peptides and chloroplast transit peptides in the N-termini of protein sequences. The iPSORT program uses an 'amino acid index rule' analysing the biochemical properties of each of the first 30 amino acids and 'alphabet indexing' which detects patterns in the

residue sequence (Bannai *et al*, 2002). Signal P predicts the presence and location of any cleavage peptides and signal peptides in eukaryote organisms, and is based in a combination of artificial neural networks (Petersen *et al*, 2011).

Program	URL
PlasmoAP	http://v4-4.plasmodb.org/restricted/PlasmoAPcgi.shtml
PATS v 1.2.1N	http://gecco.org.chemie.uni-frankfurt.de/pats/pats-index.php
PlasMit	http://gecco.org.chemie.uni-frankfurt.de/plasmit/
iPSORT	http://ipsort.hgc.jp/
SignalP 4.0 Server	http://www.cbs.dtu.dk/services/SignalP/

Table 2.6: Subcellular localisation programs for *Plasmodium* spp.

3 Substrate Specificity of *Plasmodium falciparum* NitA

3.1 Introduction

The *pfnita* gene was annotated as a formate-nitrite transporter. In this chapter, the investigation of the possible substrate/s of the *Plasmodium* protein was pursued. Protein orthologues in other organisms are known to transport formate, nitrite or the hydrosulphide ion. However, there are organisms with orthologues that have unknown functions and no other possible substrates have been proposed. For example, the substrates of the *E. gracilis* (Deloménie *et al*, 2007) and the *S. cerevisiae* (Macuk *et al*, 2001) orthologues are unknown. The green algae *C. reinhardtii* contains six homologous proteins annotated as FNTs, one of which has been shown to transport bicarbonate (Mariscal *et al*, 2006). In *Clostridia* spp. and other bacteria the orthologues are known to transport the hydrosulphide ion (Czyzewski and Wang, 2012). Orthologues are found in protozoan parasites such as *Trichomonas vaginalis*, and the Apicomplexans *Plasmodium*, *Babesia* and *Theileria*. No orthologue has been found in *Cryptosporidium*, the only Apicomplexan not to contain an apicoplast (Zhu *et al*, 2000).

In this chapter the results of functional complementation assays are described. The assays were designed to aid in the identification of the substrate of PfNitA. Given the annotation of PfNitA as a formate-nitrite transporter, two possible substrates provided the basis for investigation: formate and nitrite. The first model organism chosen was *Aspergillus nidulans*. *A. nidulans* is a useful organism as it can be genetically manipulated with ease when compared to other organisms. This is due to a deletion in the *nkua* gene which allows an integration efficiency at the target locus of greater than 90 % (Dr. J.R. Kinghorn, personal communication). This organism has also been extensively researched with regard to nitrate and nitrite uptake (Wang *et al,* 2008 and Symington, 2009 and references therein) in this laboratory making it an ideal place to initiate this line of enquiry with regard to *Plasmodium*. A second model organism used for functional complementation was *Escherichia coli*. *E. coli* is easily propagated in the laboratory, and work has been carried out on the PfNitA orthologues EcFocA a formate

transporter (Suppmann and Sawers, 1994) and EcNirC a nitrite transporter (Clegg *et al*, 2002).

3.1.1 Possible roles for formate in *Plasmodium*

Formate is a by-product of fermentation in anaerobic bacteria such as *E. coli*, *Salmonella* and *Clostridium* species. Formate is formed in the production of acetyl CoA from pyruvate by pyruvate:formate lyase, an enzyme only expressed in anaerobic conditions (Sawers, 2005). The formate can be toxic to the bacteria, and needs to be exported from the cell to prevent acidification of the cytosol (Suppmann and Sawers, 1994). Experimental results characterising carbon metabolism in *P. knowlesi* have demonstrated formate and acetate to be end-products of pyruvate metabolism (Scheibel and Pflaum, 1970) and in *P. berghei* (Fulton and Spooner, 1956). This formate produced as a by product of cellular processes could be toxic to *Plasmodium* parasites. If so, the formate would require removal from the parasite and RBC, and the PfNitA transporter could efflux unwanted the formate.

Formate is produced as a by-product in the conversion of guanosine triphosphate to dihydroneopterin by GTP cyclohydrolase (GCH) I, this is the first step of the folate biosynthetic pathway (Krungkrai *et al*, 1985). In *P. falciparum* the GCH I enzyme shows variation in the copy number between parasites (Nair *et al*, 2008) most likely due to antifolate drug pressure which affects enzymes involved in folate biosynthesis. The formate produced is likely to be detrimental to the parasites and, as before, a transporter to eliminate the toxic product would be of benefit.

3.1.2 Possible roles for nitrite in *Plasmodium*

It is known that there are nanomolar concentrations of nitrite present in human red blood cells (RBCs), and that nitrate is found in the RPMI used in malaria culture medium (Ostera *et al*, 2011 and Invitrogen[22]). The removal of nitrate from RPMI leads to a decrease in parasite growth suggesting nitrate and/or nitrite are used for growth, and that parasites are able to utilise intraerythrocytic nitrite (Ostera *et al*, 2011). Nitric oxide (NO) and reactive nitrogen species (RNS) derived from NO are found in the parasites' digestive vacuole, and play a role in the formation of soluble and crystallised

[22] http://www.invitrogen.com/site/us/en/home/support/Product-Technical-Resources/media_formulation.117.html. Accessed on 17 September 2012.

haem. However, there is no obvious nitric oxide synthase gene in the malaria genome. There are however three possible functionally similar molecules, namely a putative flavodoxin-like protein (PFI1140w), a putative cytochrome b_5 (PFL1555w) and putative NADH-cytochrome b_5 reductase (PF13_0353) that have partial similarity to inducible nitric oxide synthase (iNOS) and plant nitrate reductases, respectively. It has been shown that two of these proteins, PF13_0353 and PFL1555w, also localise to the parasites DV in the trophozoite stage (Ostera *et al*, 2008 and Ostera *et al,* 2011). These data imply that (i) nitrite and nitrate from the erythrocyte and *in vitro* growth media, respectively, are used in the production of nitric oxide for use in the haemoglobin detoxification process (Ostera *et al,* 2011) and (ii) that nitrite could be produced as a by product of NO generation which is exported from the parasite and incorporated into the existing RBC nitrite pool (Ostera *et al,* 2008). The transport of nitrite and/or nitrate into or out of the parasite cell, or into the parasites' DV may require a permease. This engenders a role for the *pfnita* gene as a transporter for nitrite.

The role of nitric oxide in the pathogenesis and immune response to malaria is not yet fully understood. NO has been proposed, however, to have a protective effect against severe disease (Gramaglia *et al*, 2006 and Weinberg *et al*, 2008). For instance, inhibition of NO has also been demonstrated to reduce the parasitaemia of *P. falciparum in vitro* (Ostera *et al*, 2008) and in the *P. berghei* rodent model (Koka *et al*, 2008). A protective role for NO could be related to decreased cytokine production that may help avert the related immunopathologies, and reduce endothelial adhesion molecules. Such events would prevent rolling adhesion and sequestration of parasitised RBCs (Serirom *et al*, 2003). The supplementation of exogenous NO donors has been demonstrated to protect mice infected with *P. berghei* from the development of cerebral malaria thus limiting the disease symptoms but not parasitaemia (Gramaglia *et al*, 2006). During the infection, the levels of NO and arginine (a precursor of NO) are seen to be diminished, the reason for this observation is unclear. It has been considered that malarial arginase uses arginine from the hosts blood to produce ornithine thus diverting the arginine from being converted to NO or that haem released by the RBC rupture binds to NO rendering it inactive (Weinberg *et al*, 2008). Another possible explanation could be, that as in *S. typhimurium* infection (Section 1.5.3.1) (Das *et al*, 2009), the *Plasmodium* parasites use the formate-nitrite transporter

to prevent nitrosative stress. The exact role that the *Salmonella* transporter plays is, however, yet to be defined.

mRNA expression data from PlasmoDB[23] indicates the PfNitA transporter is active throughout intraerythrocytic and gametocyte stages. There is no current data on the gene expression in the mosquito stages available (at time of writing). The mosquito elicits an immune response which includes NO production leading to the expression of anti-microbial peptides which have anti-parasitic effects (Herrera-Ortiz *et al*, 2010). A mechanism to limit the amount of NO produced, such as the *pfnita* gene function, could be beneficial to the parasites' development and transmission. The justification is similar for the exo-erythrocytic stage of infection as NO is produced by the hepatocytes as a response to infection (Klotz *el al*, 1995).

Glutamine synthetase uses ammonia as a substrate in the synthesis of glutamine from glutamate. Glutamine is thought to be a carbon source for the parasites unusual TCA cycle and therefore an essential nutrient (Olszewski *et al*, 2011). A source of ammonia may be from the reduction of nitrite, a molecule, as noted above, to be freely available in the host erythrocyte.

3.2 Sequence Considerations

The *pfnita* gene from *P. falciparum* has been so given this designation due to its similarity to formate-nitrite sequences and has been so designated in accord to the FNT gene of *A. nidulans* known as *annita*, which has been much researched in the Unkles laboratory. The gene is not found in higher organisms making it an ideal choice for research studies as a selective drug target. Orthologues of this protein have been crystallised from various bacteria (Wang *et al*, 2009a, Waight *et al*, 2010, Lü *et al*, 2011 and Czyzewski and Wang, 2012 and Chapter 1). The crystal data has led to the identification of key residues involved in substrate binding and maintenance of the protein structure.

[23] www.plasmodb.org. Accessed on 01 January 2012.

3.2.1 Gene structure

The *pfnita* genomic sequence is predicted to contain four exons and three introns (PlasmoDB) (Figure 3.1), the UTR's are unknown. However, caution must be exerted in these regards as, to my knowledge, no cDNA copy has been sequenced and compared to the genomic copy, indeed the A+T rich genome makes cDNA sequence analysis difficult in *Plasmodium* species (López-Barragán *et al*, 2011). To obtain the 5' and 3' untranslated regions, 5' and 3' rapid amplification of cDNA ends (RACE) could be carried out. Each of the putative introns contains the canonical GU AG dinucleotide sites at its periphery for 5' and 3' splicing. The A+T content of the three introns averages at 86.67%, typical of *P. falciparum* (Zhang *et al*, 2011). The two nucleotides adjacent to the 5' splice site in *P. falciparum* are most often AG, this is observed in the first and third putative exons. In exon two however, the nucleotides are AC. The AC in this position is not well represented in *P. falciparum* with approximately 3 % of introns having an ACgt boundary (Zhang *et al*, 2011) (uppercase representing exons and lower case introns). The final exon is only 25 nucleotides long, with the final 3 making up the stop codon.

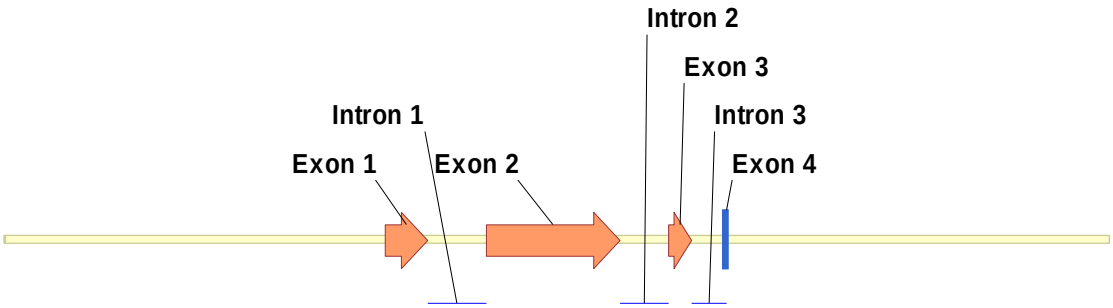


Figure 3.1: Putative genomic structure of *pfnita*. A schematic to illustrate the putative intron and exon structure of the *pfnita* gene (PFC0725c). Exon 4 is distorted due to its small size. The UTR's are unknown and have not been included. The diagram was created using Vector NTI.

3.2.2 Protein secondary structure predictions

Membrane topology analysis using online computer software was used to investigate the number of predicted transmembrane domains (TMDs) in the protein (Section 2.6.2 and Table 2.5). FNTs are thought to have 6 – 7 transmembrane domains. The PfNitA amino acid sequence was subjected to five common topological prediction programs (Table 3.1). These hydropathy plots consistently demonstrated the likely presence of

six TMDs in the amino acid sequence, this is in agreement with other FNTs. The plots only estimate the likely position of TMDs in amino acid sequences using a hydrophobicity index. To confirm the output data the sequence was analysed for hydrophobic residues known to be found in transmembrane domains and charged residues not regularly found in TMDs. The putative transmembrane domain positions from each program are given in Table 3.1. A representative hydropathy plot has been reproduced in Figure 3.2.

TMD	TMHMM	HMMTOP	SOSUI	TMpred	TopPred
1	48-70	41-64	43-65	49-68	47-67
2	83-105	77-101	84-105	76-98	80-100
3	129-151	132-155	130-152	132-156	138-158
4	184-206	186-203	182-204	186-206	186-206
5	216-238	210-227	219-241	209-228	209-229
6	258-280	258-281	266-282	262-281	261-281

Table 3.1: Hydropathy plot output data for PfNitA. The amino acid positions of each of the transmembrane domains of PfNitA, as predicted by each TM predictor program.

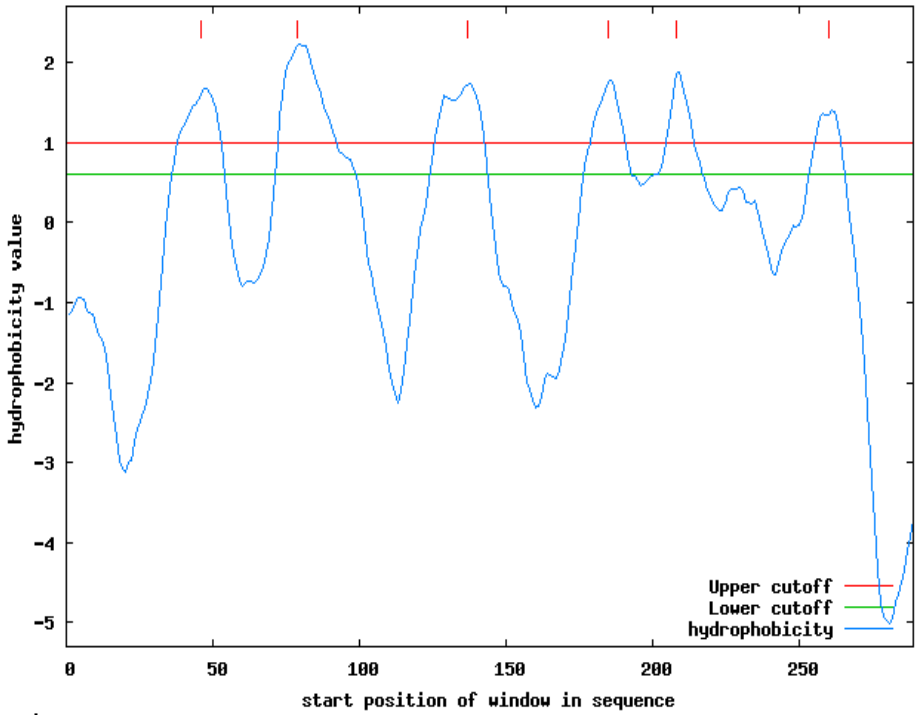


Figure 3.2: Hydropathy plot to indicate the predicted transmembrane domains of PfNitA. An upper and lower hydrophobicity value was set automatically and can be seen as a red and green line running across the images, respectively. Peaks above the red line are taken as transmembrane domains, and those above the green line are putative transmembrane domains. 'start position of window in

sequence' denotes the sliding window of 21 amino acids, which can be found in the 309 amino acid PfNitA sequence. Image taken from Mobyl@Pasteur v1.0 [24].

A 2D structure model of the PfNitA protein was created using data generated from the combination of hydropathy plots (Table 3.1 and Figure 3.2) and visual analysis of the sequence as well as known important residues (Figure 3.3). The model does not however indicate the precise locations of the transmembrane domains. The TMDs of PfNitA are predicted to form α-helices and therefore, hydrophilic faces of such TMDs may form the lining of an aqueous accessible region through which the substrate is transported.

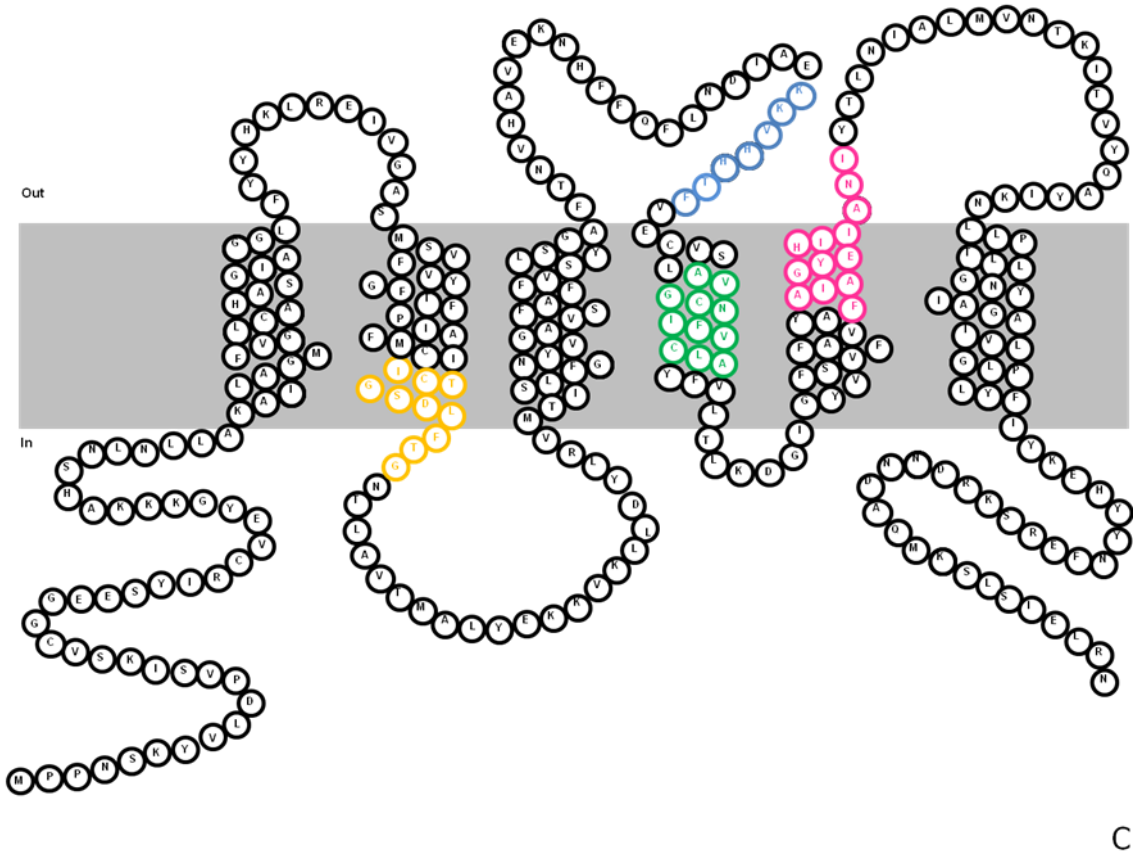


Figure 3.3: 2D structure model of the PfNitA protein. Structural model for PfNitA based on the outputs from the hydropathy plot programs (Table 3.1 and Figure 3.2) and sequence data. The known FNT motifs 1, 2 and 3 are shown in orange, green and pink, respectively. The novel formate motif is indicated in blue. N and C denote the termini. Diagram is not to scale.

There are three conserved motifs conserved in FNT proteins (Saier *et al*, 1999). The first motif [LIVMA]-[LIVMY]-X-G-[GSTA]-[DES]-L-[FI]-[TN]-[GS], is present in the

[24] http://mobyle.pasteur.fr. Accessed on 28 January 2012.

PfNitA amino acid sequence as ICTGSDLFTG. This motif is found in the second TMD and covers amino acids 98 – 107.

Residues D103, L104 and T106 are proposed to play a role in the transporters activity and structural fold. D103 and T106 correspond to *E. coli* and *V. cholerae* residues that are involved in hydrogen bonding, while L104 corresponds to a channel lining constriction residue in *E. coli* (Wang *et al*, 2009a) or in *V. cholera* a residue forming part of the selectivity filter (Waight *et al*, 2010). In Figures 3.3 and 3.4 this motif is illustrated in orange.

The second motif [GA]-X(2)-[CA]-N-[LIVMFYW](2)-V-C-[LV]-A, is present as AVGCNIFVCLA. This second motif is found in the fourth TMD and spans residues 189 – 199. Residues N193 and V196 are proposed to be in the selectivity filter of *V. cholerae* (Waight *et al*, 2010) and V196 a channel lining residue of *E. coli*, implying they might play a role in the translocation of the substrate. Residue N193 corresponds to a residue involved in bonding to T91 in *E. coli*, which is T106 in the *P. falciparum* orthologue and found in the first FNT motif (Wang *et al*, 2009a). This data points to a structural role for these latter residues. In Figures 3.3 and 3.4 this motif is presented in green.

The third motif F-[IVFA]-X-[LIS]-G-[LEYT]-[EQ]-H-[SVCY]-[VI]-[AG]-[ND]-[MLQ] is present in The *P. falciparum* protein as FAIAGYEHIIANI. This motif is found in TMD 5 and spans residues 223 – 234. F223, H230 and A233 are found as channel lining residues and as part of the selectivity filter. E229 and N234 correspond to residues that form hydrogen bonds with each other in *E. coli* (E208 and N213) (Wang *et al*, 2009a), and residue H230 was shown to be in the selectivity filter and form hydrogen bonds in *V. cholerae* (Waight *et al*, 2010). In Figures 3.3 and 3.4 this motif is indicated the by pink colouring.

```
PfNitA   MPPNNSKYVLDPVSIKSVCGGEESYIRCV----EYGKKKAHYSNLNLLAKAILAGMFVGL
PkFNT    MSKGKSKYVIDPISVKTACTSEESYIRCV----EYGKGKAHYPNLSLLAKAILAGVFVGV
PvFNT    MTKG-SKYTIDPISVKTACTSEESYIRCV----EYGKGKAHYPNLSLLAKAILAGVFVGV
PbFNT    MGKSGKQYIIDPASVKTTCSSEESYIRCV----EYGKAKATYSNFNLFLKAILAGIFVGL
PcFNT    MGKIGKQYILDPTSVKTTCSSEESYIRCV----EYGKAKAAYSNFNLFLKAILAGIFVGL
PyFNT    MGKTGKQYILDPPCVKTTCSSEESYIRCV----EYGKAKATYSNFNLFLKAIMAGIFVGL
BbFNT    M---------EVGTVSQIRKAGDAYKVAI----DTGIYKKSCPIYILLVKAMMGGYFAAL
TaFNT    M----------FEYTRQLASAKDNYERIA----KEAGDKVNGNILTLFVKSLLGGWFVAI
TgFNT    M---------------VVTASPDTYLHVI----DYGLKKVRLRFDRLLLQAFMAGVYIGM
EcFocA   M---------KADNPFDLLLPAAMAKVAE----EAGVYKATKHPLKTFYLAITAGVFISI
StFocA   M---------KADNPFDLLLPAAMAKVAE----EAGVYKATKHPLKTFYLAITAGVFISI
VcFocA   M---------E-HNQFDSLLPPQMAERAA----ITGEGKAKKAAYKSFLLAISAGIQIGI
StNirC   M--------------FT-----DSINKCAAKLRASAPVSANNPL-GFWVSSAMAGAYVGL
EcNirC   M--------------FT-----DTINKCAANAARIARLSANNPL-GFWVSSAMAGAYVGL
CdHSC    M---------G--RAHK-----ETLDKLTNAAINSINLLNTSKV-KYLVSSAFAGLYVGI
MtFDHC   M---------G--SSFK--SPADTAKACS----AIAELKEKAPLGKVIVLSFLAGAYIAF
MfFDHC   M---------A--SSFK--SPADTAKACV----GVAALKEKAPLSNLIVLSFVAGAYIAF
AnNitA   M---------P--PSIDAYTPLEVVEFVS----RAGSVKGHMRLDKIFLSSLSAGCLLAL
EgFTH    M--------------PNHVKAVDVLEGAI----NAGVMKAHMPFMDMAIRGLYAGFFLGV
CaYHA8   M------------DDNLYLTTYEAALAVV----ATAMKKARLRIDVLVINSLMGGMLFTT
ScYHL    M-----------VDDSNYLTPHETALAVV----ATAMKKARLQLDTLLINSILGGVLFSS

PfNitA   CAHAS-GIAGGLFYYHKLREIVGASMSVFVYGFTFPIAFMCIICTGSDLFTGNTLAVTMA
PkFNT    CAHAS-GIAGGHFYYHKLREYVGISMSAFVYGFTFPIAFLCIIATGSDLFTGNTLAVTTA
PvFNT    CAHAS-GIAGGHFYYHKLREHVGISMSAFVYGFTFPIAFLCIIATGSDLFTGNTLAVTTA
PbFNT    CAHAS-GIAGGLFYYHKLRAYVGPSFSSFVYGFTFPIAFLCIICTGSDLFTGNTLAVTVA
PcFNT    CAHAS-GIAGGLFYYHKLRAHVGASFSSFVYGFTFPIAFLCIICTGSDLFTGNTLAVTIA
PyFNT    CAHAS-GIAGGLFYYHKLRAYVGASFSSFVYGFTFPIAFLCIICTGSDLFTGNTLAVTVA
BbFNT    GGHAA-MVLASYYY-MD--GHHGGAKLAF--GVIFSGALLCIVFTGTDLVTSNCMNFAFL
TaFNT    GGYAA-SVIASLFYEQD--ASNGAARAAF--SLIFPGALCAILFTGSDLYTGNTMSFTFA
TgFNT    AGNACISLAGGFSTDPADPKAITAGVQKFIYASIFPVAFIAIIMTGAELFTGNTMTMLIC
EcFocA   AFVFYITATTGTGT-----M--PFGMAKLVGGICFSLGLILCVVCGADLFTSTVLIV-VA
StFocA   AFVFYITATTGTGA-----M--PYGMAKLIGGICFSLGLILCVICGADLFTSTVLIV-VA
VcFocA   AFVFYTVVTTGAHD-----M--PYGVTKLLGGLAFSLGLILVVITGGELFTSSVLIL-VA
StNirC   GIILIFTLGNLLDP-----SVRP--L---VMGATFGIALTLVIIAGSELFTGHTMFLTLG
EcNirC   GIILIFTLGNLLDP-----SVRP--L---VMGATFGIALTLVIIAGSELFTGHTMFLTFG
CdHSC    GILLIFTIGGLLTD-----AGSP--MTKIVMGLSFAIALSLVIMTGTELFTGNNMVMSAG
MtFDHC   GGLLAEVVTGGMAK-----AGYPAGLVKLVFGAVFPVGLMLVVIAGSELFTGNCMYMPLG
MfFDHC   GGLLAEVATGGMAA-----AGWPTGLVKLVFGGVFPVGLMLVVIAGSELFTGNCMYMPMG
AnNitA   ACGTTLSTNASPWF-----TENAPGLIRTISALVFPYGLVLIILTGADLCTGSFMFTTVA
EgFTH    TASFNFQLTALYGN-------------AMIGGLLFFAGFMMIIFYGFELVTGNMLVLPLA
CaYHA8   GGMLYDLIRAGFSG----INETNPGVISLLQGICYPIGLFYVVILGVDLFNSNILFFSTA
ScYHL    GSFLLVAVYSEDPD----IVARNPGIVNLITGVNFAMGLFYVVMMGADLFNSNILFFSVG

PfNitA   LYEKKVKLLDYLRVMTISLFGNYVGAVSFAFFVSYLSGAFT-NVHAVEKN-HFFQFLNDI
PkFNT    LLQRKVSLLQYLRVMSISLFGNYLGAVSFAFFVSHLSGAYE-KHTDVTKN-HIFQFLNDI
PvFNT    LLQRKVTLLEYLRVMSISLFGNYVGAVSFAFFVSHLSGAFK-KHEEIGKN-HIFQFLNDI
PbFNT    LLQKKVKLLSYIRVMSISLLGNYIGAVSFAFIVSYGSGAFH-QHSDIEKN-HIYEFLNAI
PcFNT    LLQKKIKLLSYIRVMSISLLGNYIGAVSFAFIVSYGSGAFH-QQASIEKN-HIYEFLNAI
PyFNT    LLQKKVKLLSYIRVMLISLLGNYIGAVSFAFIVSYGSGAFH-QHSDIEKN-HIYEFLNAI
BbFNT    AYSRQVTLFQYLTRMGTSLLGNYCGAILGAALLTAGTGYFS-L--D-KRP-E--AYLKAI
TaFNT    LCKKHITFLNYFLKLGISIMGNYIGAVIGALILSGGTTYFL-K--DVGKG-A--PYLLDM
TgFNT    WFERRITIWQLLQNWAGSFLGNWLGTMFSAYFLTYLCCPFD-H------D-PYLSYLNYT
EcFocA   KASGRITWGQLAKNWLNVYFGNLVGALLFVLLMWLSGEYMT-A------NGQWGLNVLQT
StFocA   KASGRITWGQLAKNWLNVYFGNLIGALLFVLLMWLSGEYMT-A------NGQWGLNVLQT
VcFocA   KASGKISWKELVRNWTVVYFGNLCGSIILVFIMLATRQFME-D------GGQLGLNAMAI
StNirC   VKAGTISHGQMWAILPQTWLGNLVGSVFVALLYSWGGGS-L-L------P-VDTSIVHSV
EcNirC   VKAGSISHGQMWAILPQTWLGNLVGSVFVAMLYSWGGGS-L-L------P-VDTSIVHSV
CdHSC    MLNKGVSIKDTSKIWAYSWVGNLIGALVLGIIFVGTGLV-D-K------G-PVAEFFANT
MtFDHC   ILDKRASIMGLIRNWVTSWVFNLVGAVFVAYFLAVATGILT-A------D-PWQAGALTV
MfFDHC   ILQGEASVMGTIKNWVGSWVFNLVGALFVAYVLAYLTGILT-A------E-PWAATAVTI
AnNitA   ALHRRLSWPKMLLHWFITFWGNLCGSLFVVAIIFGYGEVFS-A------D-PFRSAVISF
EgFTH    LMARRAPWLGCPKNLLTVYTGNFLGCALYLAFFYGSSLILDCGWTAIKVAQLYCSWKTKT
CaYHA8   LCRGAVSFLDLFISWFVSWWFNLVGNIFVCYIFCYYSDVVR--------TQLMVVGSVEV
ScYHL    VLRKAVTIYDLMISWVVSWLGNIAGSLFVSYLFGHLSGISS--------QKLWIIGSRQI
```

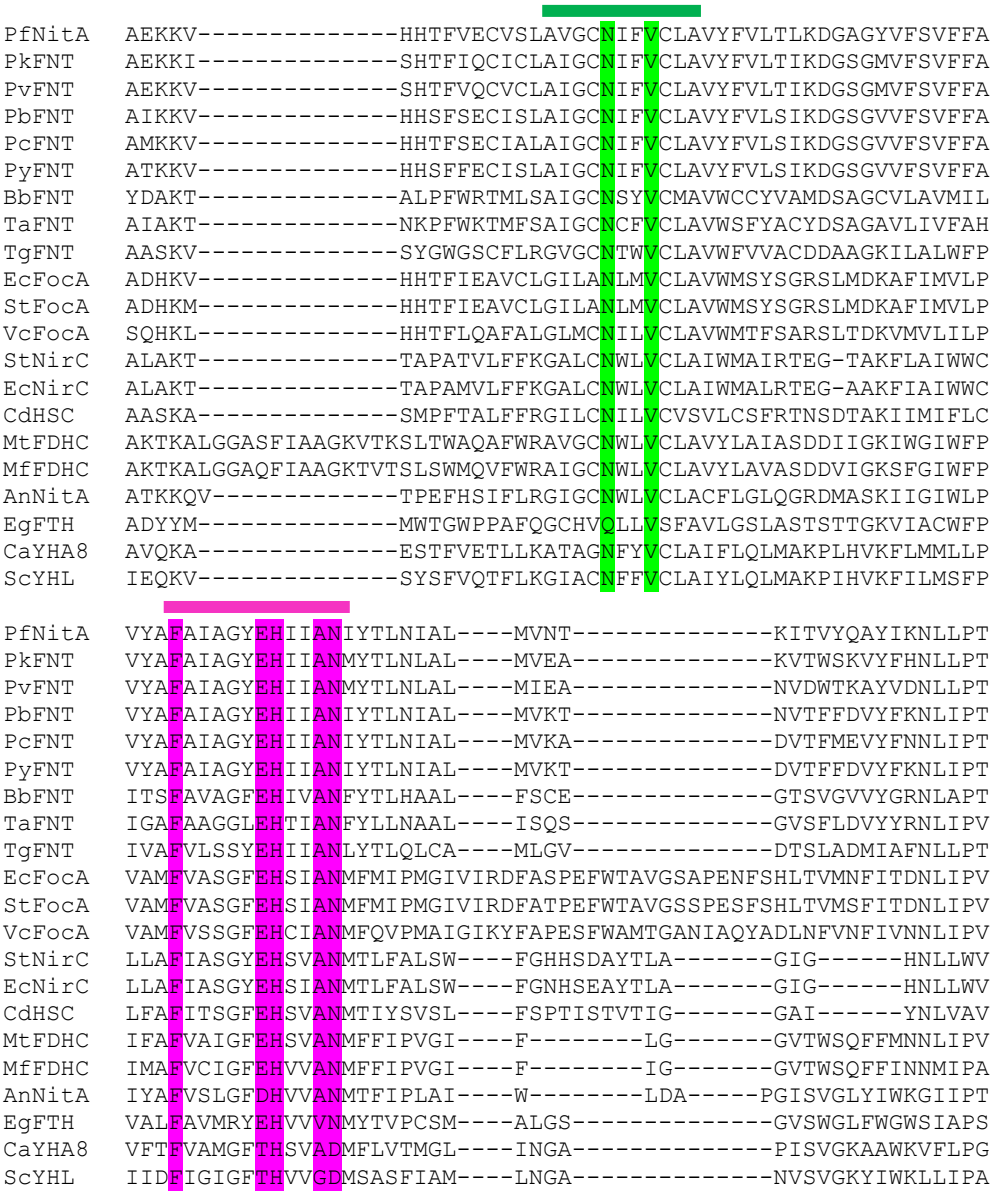


Figure 3.4: Partial multiple alignment of FNT orthologue amino acid sequences. The three highly conserved FNT sequence motifs 1 2 and 3 (as described above) are indicated by orange, green and pink bars, respectively. Residues within these sequences are postulated to play a role in the selectivity and activity of the group of permeases and are coloured appropriately. StNirC, EcNirC, and AnNitA are known to transport nitrite. EcFocA, VcFocA and StFocA are known to transport formate. CdHSC is known to transport the hydrosulphide ion. MtFDHC and MfFDHC are known to transport formate. Ca YHA8 and ScYHL are yeast orthologues. PfNitA, PkFNT, PvFNT, PbFNT, PcFNT, PyFNT TaFNT, BbFNT, TgFNT, and EgFTH are protist orthologues with unknown substrates. Pf – *P. falciparum*, Pk – *P. knowlesi* Pv – *P. vivax,* Pb – *P. berghei*, Pc – *P. chabaudi*, Py – *P. yoelii* , Bb – *Babesia bovis,* Ta – *Theileria annulata*, Tg – *T. gondii,* Ec – *Escherichia coli*, St – *Salmonella typhimurium*, Vc – *Vibrio cholerae*, Cd – *Clostridium difficile*, Mt – *Methanothermobacter thermautotrophicus*, Mf – *Methanobacterium formicicum*, Eg – *Euglena gracilis*, Ca – *Candida albicans* an Sc – *Saccharomyces cerevisiae.* The alignment was made using MAFFT and sequences were truncated at the C-terminals. Accession numbers can be found in Appendix 3.

A fourth putative motif HKXHHTF, found to be conserved between putative formate transporters (Dr. J.L. Cecile, personal communication) is also present in *Plasmodium* orthologues, and is found in an extracellular loop. Interestingly this novel motif is not found in other Apicomplexan organisms such as *Theileria* and *Babesia*, or formate transporting archaea. An alignment of select *Plasmodium* and bacterial orthologues highlighting the presence of this novel motif can be found in Figure 3.5. From the alignment, the HKXHHTF motif (KKVHHTF in *P. falciparum*) is observed only in putative bacterial formate transport proteins and *Plasmodium* proteins. In Figures 3.3 and 3.5 this motif is shown in blue.

```
PfNitA 103 SLFGNYVGAVSFAFFVSYLSGAFTNVHAVEKNHFFQFLNDIAEKKVHHTFVECVSLAVGCNIF 166
PkFNT  103 SLFGNYLGAVSFAFFVSHLSGAYEKHTDVTKNHIFQFLNDIAEKKISHTFIQCICLAIGCNIF 166
PcFNT  103 SLLGNYIGAVSFAFIVSYGSGAFHQQASIEKNHIYEFLNAIAMKKVHHTFSECIALAIGCNIF 166
EcFocA 87  VYFGNLVGALLFVLLM-WLSGEYMTANGQWGLNVLQ----TADHKVHHTFIEAVCLGILANLM 145
StFocA 87  VYFGNLIGALLFVLLM-WLSGEYMTANGQWGLNVLQ----TADHKMHHTFIEAVCLGILALMV 145
VcFocA 86  VYFGNLCGSIILVFIM-LATRQFMEDGGQLGLNAMA----ISQHKLHHTFLQAFALGLMCNIL 144
```

Figure 3.5: Partial multiple alignment of *Plasmodium* orthologues and known formate transporters. The novel formate transporter sequence motif (HKXHHTF) is indicated by the blue bar. EcFocA, VcFocA and StFocA are known to transport formate. PfNitA, PkFNT, and PcFNT are example *Plasmodium* orthologues with unknown substrates. Pf – *P. falciparum*, Pk – *P. knowlesi,* Pc – *P. chabaudi*, Ec – *Escherichia coli*, St – *Salmonella typhimurium*, Vc – *Vibrio cholerae*. Accession numbers can be found in Appendix 3.

3.2.3 Protein tertiary structure predictions

Four crystal structures have been produced for formate-nitrite transporters, namely FocA, of *E. coli*, *V. cholerae*, and *Salmonella typhimurium* and FNT3/CdHSC of *Clostridium difficile*. These four structures make it relatively simple to create a prediction of the PfNitA tertiary structure. The PfNitA amino acid sequence was aligned to the sequences with known crustal structures. The FocA proteins from *E. coli* and *S. typhimurium* demonstrate the greatest sequence similarity to PfNitA (Table 3.2).

The PfNitA amino acid sequence was uploaded onto the I-TASSER[25] (Zhang, 2008) sever, and this approach yielded five predicted 3D models. This program computes the likely structure and returns a model based on the template. These 5 models were studied with regard the predicted TMDs, structural motifs crystal data, and from the

[25]http://zhanglab.ccmb.med.umich.edu/I-TASSER/. Accessed on 4 September 2012.

outcomes the most suitable chosen. The model, however, is a prediction and does not show the true structure of the PfNitA protein. So caution must be exerted since to obtain the precise information the crystal structure of the protein must be solved.

PfNitA amino acid sequence comparison with crystallised orthologues		
Organism	Identity (%)	Similarity (%)
V. cholerae	23.4	38.7
E. coli	24.7	41.5
C. difficile	22.3	36.2
S. typhimurium	24.6	41.4

Table 3.2: Comparison of the amino acid sequence from PfNitA and homologous transporters with known crystal structures. Identity is an identical match between two amino acids, similarity is where two amino acids share similar properties. Alignments were carried out using EMBOSS Needle – Alignment[26].

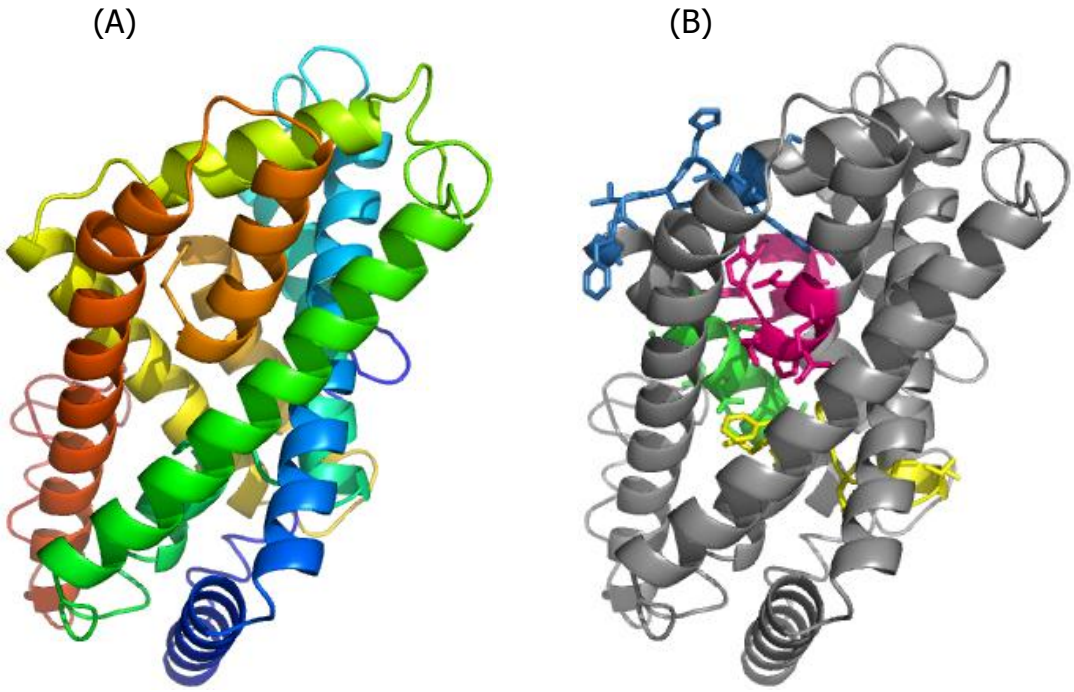


Figure 3.6: Predicted tertiary structure for the PfNitA monomer. (A) The protein as viewed from the planar angle, both termini are at the bottom of the diagram as they are predicted to be intracellular. From this angle, the split helices of domains 2 (yellow/green) and 5 (pale orange) can be seen. (B) The four motifs discussed in the text are displayed with respect to the crystal structure while the other residues are shaded in grey.

[26] http://www.ebi.ac.uk/Tools/psa/emboss_needle/. Accessed on 15 August 2012.

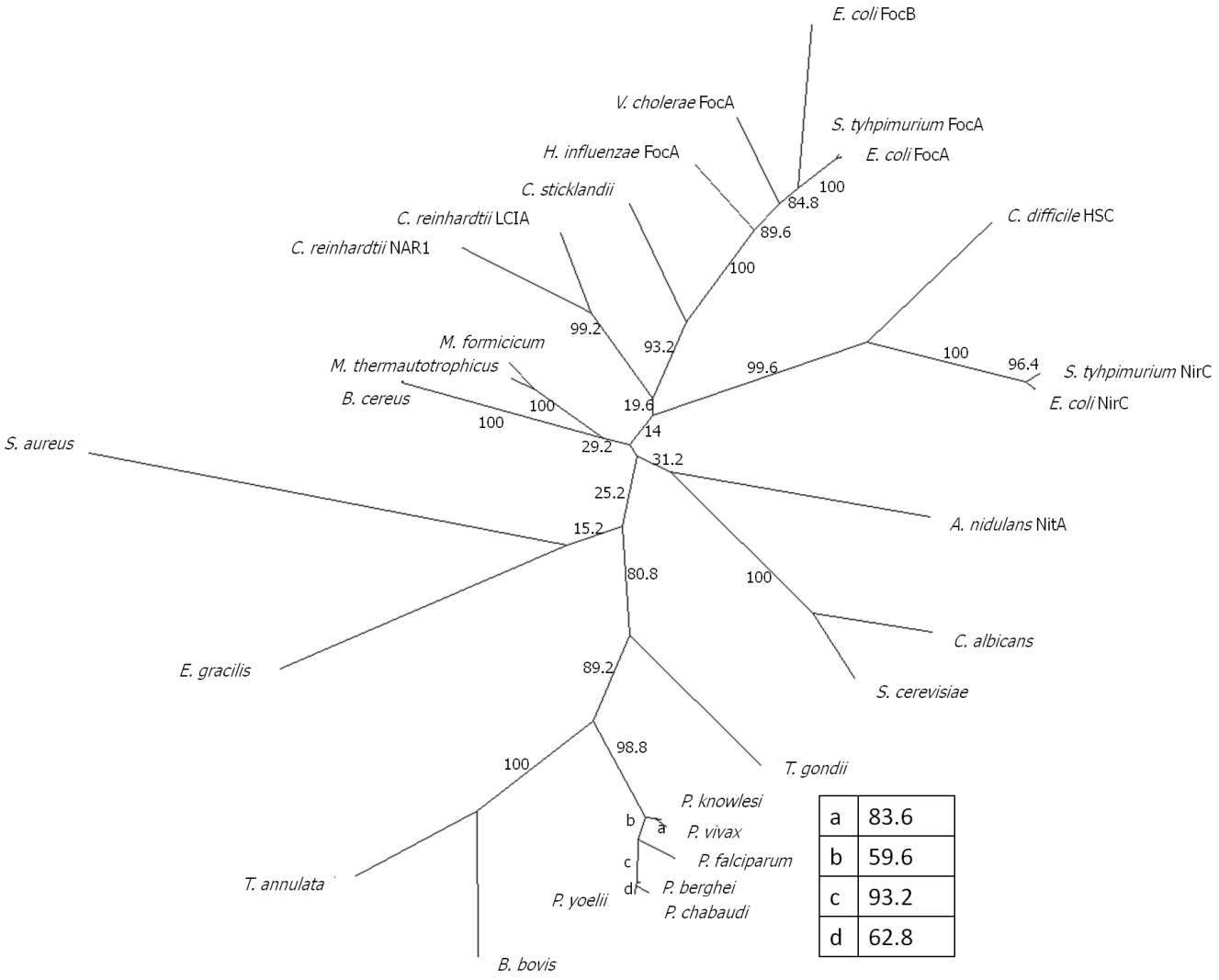
Secondary structure programs predict the N and C termini of the PfNitA protein to be intracellular. The image in Figure 3.6, A illustrates a 3D structural model of the PfNitA protein with the termini coloured red and blue. With respect to B, the important residues from each motif have been highlighted, the colours corresponding to previous figures. The structures illustrate possible interactions of the important residues with each other and where they are likely to be found within the 3D structure.

3.2.4 Construction of a phylogenetic tree

A phylogenetic tree was constructed to explore which of the 4 known clusters, I – IV, (Saier *et al*, 1999) of FNT proteins the PfNitA protein most closely related to. It was anticipated that a phylogenetic study might give further clues to the possible substrate/s of the transporter. Initially a large tree was constructed using 28 of the FNT protein sequences from prokaryotes, archaea and eukaryotes. The longer sequences (*T. gondii*, *E. gracilis* and the yeast orthologues) created too many gaps in the alignment to be processed by PhyML. To eliminate these gaps, the putative FNT domains from the four sequences were chosen and the N- and C-termini truncated. The truncation left an approximately 250 amino acid fragment for each sequence. This inspection was carried out manually though an alignment made using MAFFT. The *T. gondii* FNT domain was created between amino acids 81 – 332, *E. gracilis* domain 9 – 257, *S. cerevisiae* 1 – 261 and *C. albicans* 1 – 260. The data presented in Figure 3.7 shows a bootstrapped (250 replications) unrooted phylogenetic tree. The phylogenetic tree remained unrooted because no outlier could be chosen from the known orthologues and another factor taken into consideration is that the evolutionary history of the FNT family is as yet unknown. Bootstrap support numbers are indicated on branches.

The large unrooted tree clusters the *Plasmodium* orthologues into their own group which have the closest evolutionary history the apicomplexan orthologues, and further with the eukaryote orthologues, as might have been expected (Figure 3.7). The protozoan parasite orthologues show the closest history with the other eukaryotic organisms, however, given the low bootstrap values extending from this part of the tree, it is difficult to draw any meaningful conclusions about function. The very low bootstrap support values (i.e. those less than 50 %) implied that there either was not enough, or was too much variation between the sequences to find a useable

phylogenetic 'signal' (Dr. D. Barker, personal communication). It is possible that repeating the analysis using the nucleotide sequences from each of the orthologues could assist the creation of a more meaningful tree.



a	83.6
b	59.6
c	93.2
d	62.8

Figure 3.7: Unrooted large phylogenetic tree. A phylogenetic tree made from a MAFFT alignment of 28 FNT orthologues from each of the four known FNT clusters and protozoan organisms. Gaps found in the sequences were eliminated (see text) for submission to PhyML. Numbers indicate the percentage support for each branch from a total of 250 replicates. The branch lengths are arbritary as they are generated by the software.

A second unrooted tree was created (Figure 3.8) using two organisms from each of the reported clusters, the *Clostridium* orthologue and the single *P. falciparum* sequence. In this tree the *Plasmodium* orthologue shows the closest evolutionary history with the formate transporters of *E. coli* and *V. cholerae*.

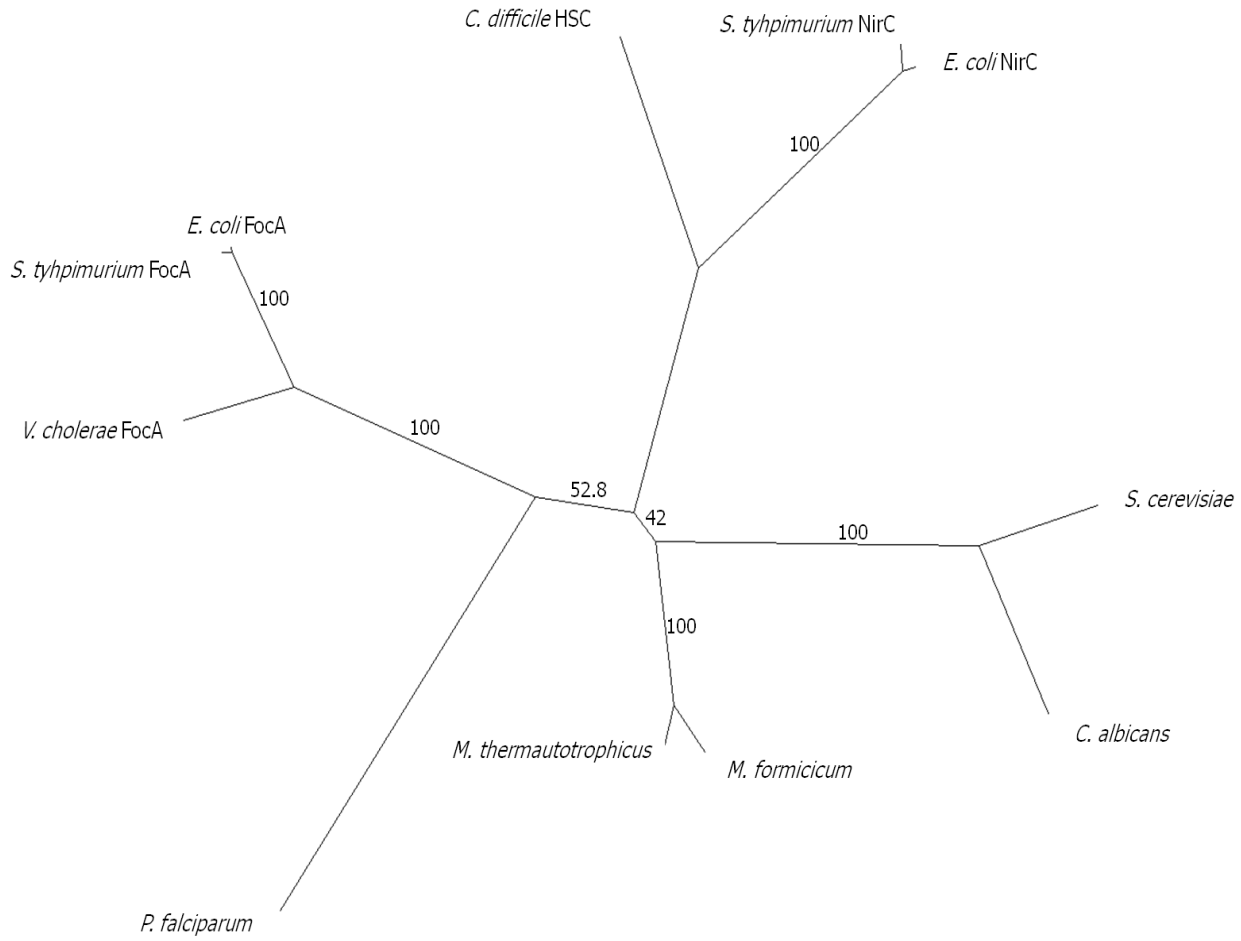


Figure 3.8: Smaller unrooted phylogenetic tree. A phylogenetic tree made from a MAFFT alignment of two orthologues from each of the four reported FNT clusters (yeast transporters, bacterial formate transporters, bacterial nitrite transporters and archeae formate transporters), the *Clostridium* HSC, and the *P. falciparum* FNT orthologue. Numbers indicate the support for each branch in percentages from a total of 250 replicates. The branch lengths are arbritary as they are generated by the software.

The bootstrap support numbers on each branch are fairly high from the smaller tree indicating that the data is significant. Given that the bootstrap values are much higher on the second smaller tree than on the initial larger tree, indicating that the data and analysis strongly support most of the groupings. Thus the data from the phylogenetic analyses points towards a role in the transport of formate for the *P. falciparum* transporter. However, care must be taken in the assessment of the smaller tree as it contains a smaller number (eleven) of protein sequences. Long branch attraction (LBA) is a phenomenon where sequences are grouped together in a cluster even though they have disparate sequences (Bergsten, 2005). Due to the limited number of sequences in this analysis, it is possible that some of the sequences have been placed

together due to LBA. In this example, the *P. falciparum* protein sequence may have been wrongly placed in the tree due to its potentially divergent sequence and the long branch length observed in Figure 3.8.

3.3 Complementation in *A. nidulans*

To facilitate expression in *A. nidulans*, the *pfnita* gene was redesigned to reduce the A+T richness from 70 % to 45 %. The codon-optimised synthetic *pfnita_opt* gene was cloned into two *Aspergillus* expression vectors upstream of a V5 tag (Figure 3.9, A and B) by Dr. S.E. Unkles.

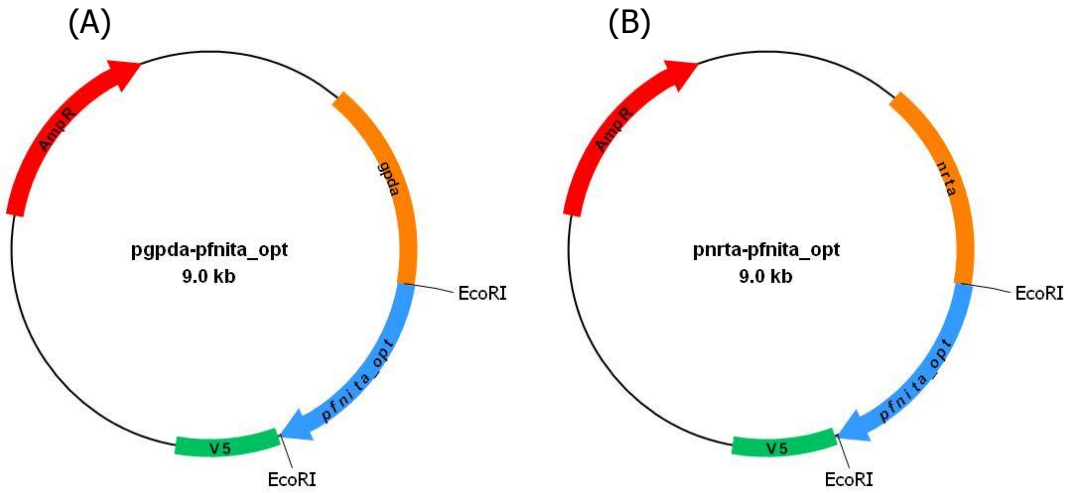


Figure 3.9: Cloning *pfnita_opt* for expression in *A. nidulans*. Plasmid constructs showing the *pfnita_opt* gene downstream from a V5 tag, and the two different promoters (A) *gpda* a strong promoter and (B) *nrta* a nitrite/nitrate inducible promoter. The restriction sites used to clone in the *pfnita_opt* gene, *Psf*I and *Cla*I are shown as well as the ampicillin resistance (AmpR) gene which acted as a selectable marker during cloning. Images are not to scale.

As this was the first work carried out on this protein, coupled with the difficulty in expressing and purifying membrane proteins, the V5 tag was fused to the C-terminus of the protein to allow detection using a V5-antibody (Table 2.3). The *pfnita_opt* optimised gene was cloned into the *Aspergillus*-V5 vectors using the restriction endonuclease site *Eco*RI. The orientation of the insert was checked using two single restriction enzymes digests using the restriction enzymes *Pvu*II and *Sfu*I, inserts in the correct orientation were sequenced. The *Aspergillus*-V5 vectors contain either a strong *gpda* promoter or a nitrite/nitrate inducible *nrta* promoter, and a V5 tag for detection.

Both constructs p*gpda-pfnita_opt* and p*nrta-pfnita_opt* (Figure 3.9, A and B) were transformed into a mutant strain of *A. nidulans* (deleted for the two nitrate transporter genes, *annrta* and *annrtb*, and the single nitrite transporter *annita* gene, as well as the *anargb*2 gene, a system which allowed for selection for arginine prototrophs) by Dr. S.E. Unkles. Integration of the plasmid into the *A. nidulans* genome was confirmed by Southern blot analyses of DNA isolated from transformants.

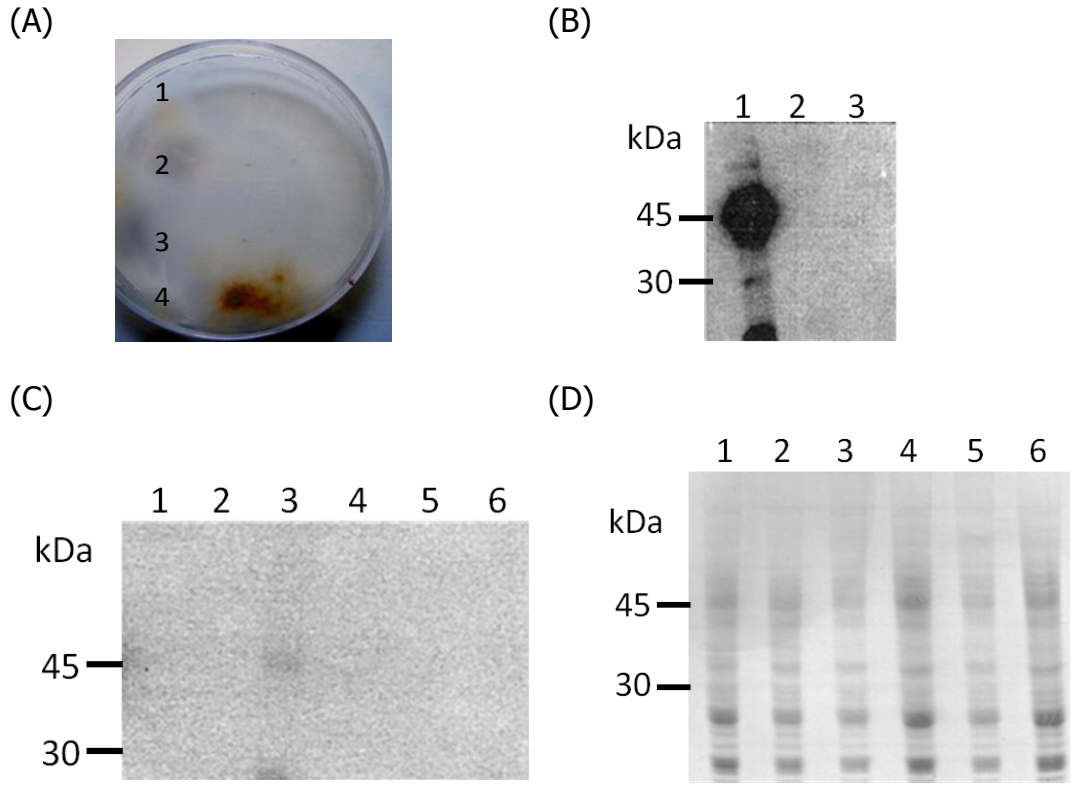


Figure 3.10: Expression of PfNitA_opt in *A. nidulans*. (A) *Aspergillus* minimal medium with 2 mM nitrite as sole nitrogen source was point inoculated with a small number of fungal spores and the growth phenotype assessed after three days incubation at 37 ^{o}C. 1. p*gpda-pfnita_opt* 2. p*gpda-pfnita_opt -intron-Y*212*K* 3. Triple mutant lacking *annrta*, *annrtb* and *annita* 4. WT *A. nidulans*. (photograph courtesy of Dr J.R. Kinghorn) (B) Western blot directed against the V5 tag on PfNitA_opt with PF11 a single copy (lane 2) and PF12 a multicopy (lane 3), a control sample of another protein containing the V5 tag is shown in lane 1. (C) A western blot using an anti-V5 antibody directed against the 3' V5 tag. Lane 1 contains membranes from PF104, lane 2 PF105, and lane 3 PF106. All samples are 'intron' mutated and expressed with the *nrta* promoter. Lane 4 contains membrane from PF124, lane 5 PF125, and lane 6 PF128, again all are 'intron' mutated however under the control of the *gpda* promoter. (D) CBB stained gel of the same samples shown in the western blot in (C) to confirm protein is present in each of the samples.

From the first transformation experiment using p*gpda-pfnita_opt*, two transformed strains were chosen for further analysis, PF11 a single copy and PF12 a multi copy of the transformed plasmid containing *pfnita_opt*. Fungal cells were inoculated onto minimal agar medium supplemented with nitrite as the sole source of nitrogen, to assess any change in growth phenotype. The strains containing the codon-optimised synthetic *pfnita_opt* gene did not grow on nitrite as sole nitrogen source and therefore did not complement the mutant. The growth responses presented in Figure 3.10, A shows that there is no growth of the transformed strain containing the *Plasmodium* based gene p*gpda-pfnita_opt* (lane 1) which is similar to the triple mutant defective control (lane 3). A WT with respect to all three transport proteins is also present as a comparison where typical growth is observed (lane 4). To determine if the p*gpda-pfnita_opt* was expressed, western analyses were carried out using a HRP conjugated anti-V5 antibody (Figure 3.10, B and C).

Membranes were prepared from *A. nidulans* transformants as the expressed PfNitA_opt protein required to be present in the plasma membrane to complement the mutant. After an exposure time of 30 mins the blot showed the PfNitA_opt protein had not been expressed in the membranes of PF11 and PF12, (Figure 3.10, B). A strain known to express a fungal protein with a V5 fusion in the membrane was included as a positive control, A, lane 1) is approximately 45 kDa in size and was present as expected. It was noted that the expected size of the PfNitA_opt protein is 36 kDa (ExPASy[27]).

Although there could be several reasons for the lack of expression, two more plausible possibilities were chosen for further consideration. The first was the presence a potential *Aspergillus* 5' intron boundary that was formed during the cloning procedure at the codon-optimised gene stop codon and plasmid junction. (Dr. S.E. Unkles, personal communication). The second possible problem was the presence of a highly conserved amino acid residue (a lysine) among PfNitA orthologues that was absent from the *Plasmodium* residue sequence. The potential intron boundary was altered by PCR, using the oligonucleotide primers NitAoptF and NitAoptR producing *pfnita_opt-intron*. Secondly, using mutagenesis primers Y212KF and Y212KR (Appendix 2) a mutation was created by means of spliced overlap extension (Figure 2.2).

[27] http://web.expasy.org/compute_pi/. Accessed on 28 January 2012.

Between the TAG stop codon of the cloned *pfnita_opt* gene sequence and the *Eco*RI site used for cloning there is an additional *Sna*BI for cloning and/or mutagenesis. After the initial problems with expression of PfNitA_opt in *A. nidulans* the sequence was re-visited by Dr. S. E. Unkles. This revealed another potential drawback. The sequence read TAG**TACGTA**GAATTC, the underlined nucleotides could make a 5' intron boundary in *Aspergillus*, and the *Sna*BI site is in bold. If the adjacent 'intron' were to be spliced out, the stop codon would be lost and the following sequence nonsensical until the next stop codon was reached. To circumvent this problem, the *Sna*BI restriction site was mutated. The resulting insert was termed producing *pfnita_opt-intron*. This gene/plasmid is referred to as 'intron' mutated in Figure 3.10. This section of the experimental work was carried out by Dr. S. E. Unkles.

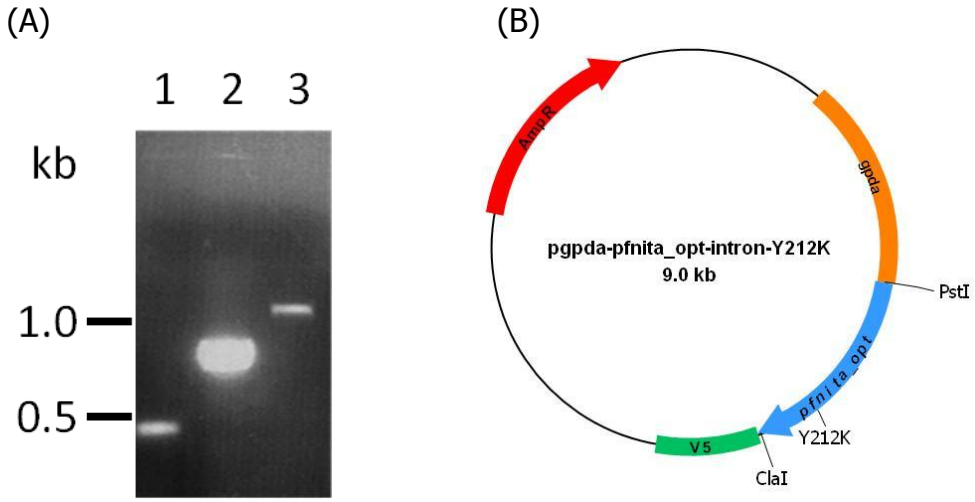


Figure 3.11: Mutagenesis and cloning of *pfnita_opt* for expression in *A. nidulans*. (A) Mutagenesis PCR products lane 1 contains the 3' PCR (400 bp) product and lane 2 the 5' PCR product (800 bp). Lane 3 contains the final 1200 bp product of the splice overlap mutagenesis PCR. The final product was cloned into the pV5M and pV5GPD vectors. (B) Plasmid construct showing the *pfnita_opt* gene downstream from a V5 tag, and the *gpda* promoter. The site of the Y212K mutation has been indicated. The restriction sites used to clone in *pfnita_opt* are shown as well as the ampicillin resistance (AmpR) gene which acted as a selectable marker during cloning. Diagram is not to scale.

The *P. falciparum* amino acid, tyrosine was mutated using spice overlap extension mutagenesis (for description see Section 2.2.2 and Figure 2.2) to the less bulky and hydrophilic lysine. The PCR products generated in this process are given in Figure 3.11, A. This produced a second mutant *pfnita_opt-intron-Y*212K. These two PCR products were cloned into the previously mentioned expression vectors creating three

new constructs: p*gpda-pfnita_opt-intron*, p*nrta-pfnita_opt-intron* and p*gpda-pfnita_opt-intron-Y*212K (Figure 3.11, B). The expression results from the new constructs were no more encouraging as none grew on nitrite as the sole nitrogen source. In these respects, the results in Figure 3.10, A part 3 show that no growth of the doubly mutated p*gpda-pfnita_opt-intron-Y*212*K* transformants was obtained.

Given the large number of arginine prototrophic strains generated from transformations with the three new constructs that lacked growth on nitrite, it was decided not to perform Southern blots (Dr. S.E. Unkles, personal communication). The transformants from p*nrta-pfnita_opt-intron-Y*212*K* (designated PFT 100 – 108) and p*gpda-pfnita_opt-intron* (designated PFT 120 – 128) were analysed by western blot and a selection are shown in Figure 3.10, C. Each of these shows no expression of the protein, except certain spurious bands (lane 3) around 25 kDa and 45 kDa, most likely revealed by the lengthy exposure time. The data in D shows a CBB stained gel to confirm that there is protein in the membrane preparations. It was decided that as the Y212K/intron mutants did not grow on nitrite, it was not worthwhile carrying out Southern or western analyses. The continuation of this general approach, i.e. the characterisation of the *P. falciparum* transporter in *A. nidulans,* would most likely not be of any benefit and was consequently abandoned.

3.4 Complementation in *E. coli*

As *E. coli* contains at least two individual transporters for formate or nitrite, it was decided this bacterium should be used as a model for functional complementation of both the formate and nitrite transporter genes. A second codon-optimised gene was purchased (Invitrogen) and is designated *pfnita_optEc*. Both of the optimised genes, *pfnita_opt* and *pfnita_optEc*, were used in the functional complementation assays, as fungal genes without introns are usually expressed in *E. coli*. The nitrite assay was performed in two *E. coli* strains, (i) the *ecnirc* gene, and the nitrate/nitrite antiporters (i.e. *ecnaru* and *ecnark*) have been knocked out and (ii) contains the *ecnirc* gene but lacks the nitrate/nitrite antiporters. The formate assay was performed using three methods with each method requiring strains in which a mutation had been generated in the *ecfoca* gene resulting in the mutants being unable to transport hypophosphite and by extension formate (Suppmann and Sawers, 1994), and a WT *foca* strain.

The *pfnita_opt* gene was amplified from a plasmid using the primers NitAopt TTQ F and NitAopt V5 TTQ R (Appendix 2) which amplified the *pfnita* gene with V5 tag, with the addition of the restriction sites for *Eco*RI and *Hind*III. The *pfnita_optEc* gene was excised from its plasmid using *Eco*RI and *Pst*I cleavage. One or other of the two products were cloned into the TTQ18 plasmid (Figure 3.12). The *pfnita_opt* gene was not cloned in frame with the hexahistidine tag as it contained a C-terminal V5 tag, while *pfnita_optEc* was cloned with a C-terminal hexahistidine tag. Proteins were expressed from the TTQ18 plasmid under *tac* promoter gene regulation, resulting in plasmid constructs pTTQ18_*pfnita_opt* and pTTQ18_*pfnita_optEc*.

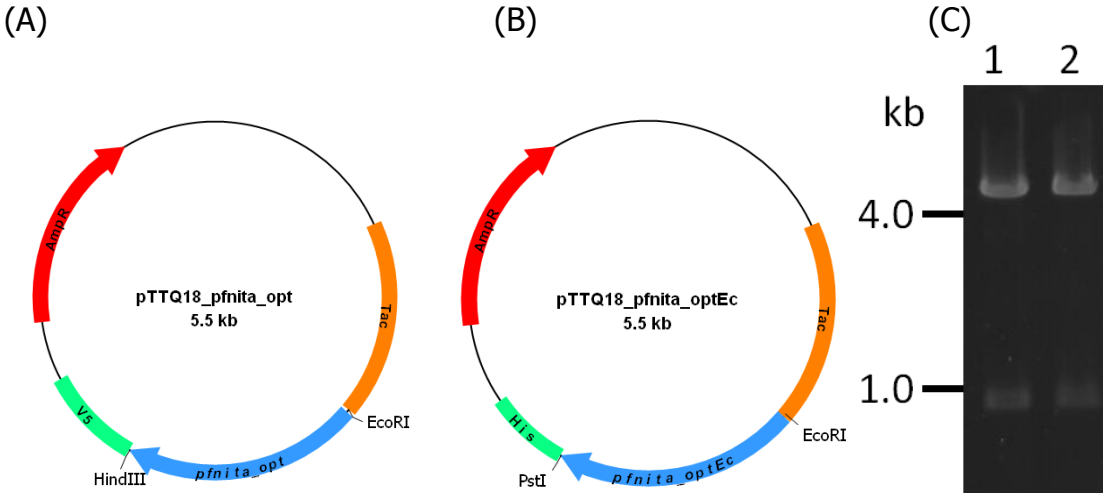


Figure 3.12: Cloning vectors for expression of the synthetic codon-optimised *pfnita* genes in *E. coli*. (A) Plasmid construct showing *pfnita_opt* upstream from a V5 tag. (B) Plasmid construct showing *pfnita_optEc* upstream from a His-tag. The restriction sites used to clone in *pfnita* are shown as well as the ampicillin resistance (AmpR) gene which acted as a selectable marker during cloning. Diagrams are not to scale. (C) Restriction digest of pTTQ18_*pfnita_opt* (lane 1) with *Hind*III and *Eco*RI. Digest of pTTQ18_*pfnita_optEc* (lane 2) with *Pst*I and *Eco*RI.

To act as a second positive control, the deleted/mutated genes, *ecnirc* and *ecfoca* were also expressed from the TTQ18 plasmid. The *ecfoca* gene was amplified from *E. coli* genomic DNA using the primers FocA F and FocA R, and *ecnirc* amplified from a plasmid using the primers NirC F and NirC R (Appendix 2). These two products were inserted into the TTQ18 vectors upstream from the hexahistidine tag, to produce re-expressor constructs pTTQ18*ecfoca* and pTTQ18*ecnirc*.

3.4.1 Nitrite as a possible substrate

Six strains were used in the nitrite complementation assays, these are given in Table 3.3.

E. coli Strain	Strain Designation
Wild Type	4018
Knock out	$4520^{-/-}$
ecnirc complementing	$4520^{-/+}$
ecfoca complementing	$4520^{-/ecfoca}$
pfnita_opt complementing (*A. nidulans*)	$4520^{-/pfnita\_opt}$
pfnita_optEc complementing (*E. coli*)	$4520^{-/pfnita\_optEc}$

Table 3.3: *E. coli* strains used for the nitrite assays. Each of the strains used in the nitrite assay are given in this table, the 'Strain Designation' column gives the names of each strain as referred to in the text.

3.4.1.1 Expression of complementing proteins in cells grown under anaerobic conditions

Cells lacking EcNirC, $4520^{-/-}$ were transformed with plasmids pTTQ18*ecnirc*, pTTQ18*ecfoca*, pTTQ18*pfnita_optEc*, and pTTQ18*pfnita_opt* (Figure 3.12) to give strains $4520^{-/+}$, $4520^{-/ecfoca}$, $4520^{-/pfnita\_opt}$ and $4520^{-/pfnita\_optEc}$, respectively (Table 3.3). Initially $4520^{-/pfnita\_opt}$ cells were grown in JC minimal salts media under anaerobic conditions to investigate the expression levels of the PfNitA_opt protein. Trials were established and bacterial cells containing the pTTQ18*pfnita_opt* plasmid were grown initially for 4 h at 37 °C, induced with one of two different IPTG concentrations and were grown for 16 or 20 h after induction. All experiments were carried out in JC media at 37 °C again under anaerobic conditions. *E. coli* crude membrane preparations were analysed for PfNitA_opt expression using NuPAGE® protein electrophoresis and western blotting against the V5 protein fused to PfNitA_opt (Figure 3.12, A). Two bacterial colonies were examined to investigate possible variation in expression between colonies. A total of 15 µg protein from a crude membrane preparation of induced bacterial cultures was applied to each gel lane. It was observed that a combination of (i) induction with 1 mM IPTG after an initial growth period of 4 h, and (ii) cells grown for a further 22 h showed best expression of PfNitA_opt (Figure 3.13, A lane 4).

The experimental blot for the 16 h time point is not presented as there was little to no protein expression in each of the samples. It is noteworthy that there was observable variation in expression levels between the colonies (data not shown). In addition the

V5 antibody has detected a much lower band on the gel, most likely this is due to proteolysis of the expressed protein.

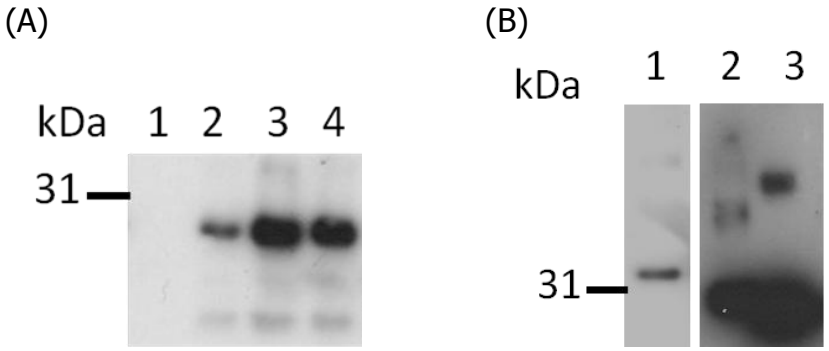


Figure 3.13: Expression of complementing proteins in *E. coli* in JC minimal salts medium. (A) A western blot to show expression of PfNitA_opt, after detection with an anti-V5 antibody, directed against the C-terminal tag on the expressed protein. Lane 1 contains a negative control with no IPTG. IPTG concentrations used were 0.1 mM (lane 2), 0.5 mM (lane 3), and, 1 mM (lane 4). The protein is running slightly lower than its expected molecular weight. Certain lower molecular weight products can be observed in each of the induced samples, probably degradation products. (B) Western blot of PfNitA_optEc (lane 1), EcFocA (lane 2) and EcNirC (lane 3) membrane proteins, using an anti-His antibody directed against the C-terminal tag. The blot confirms expression of all constructs.

The expression trials were repeated for the re-expresser strains containing the *ecnirc* and *ecfoca* sequences in cells grown for 22 h with 0.1 mM IPTG concentration of inducer. As before, all trials were carried out in JC media at 37 ^{o}C. The *ecnirc*, *ecfoca* and *pfnita_optEc* constructs did not contain a V5 tag, and therefore was replaced using an anti-His antibody against the C-terminal His-tag. The *pfnita_optEc* gene was cloned into the TTQ18 expression vector upstream of a His-tag. Expression trials for this protein were carried out using the 22 h regime with 0.1 mM, 0.5 mM and 1 mM IPTG inducer concentration. All trials were carried out in JC media at 37 ^{o}C. The data presented in lane 1 of Figure 3.13, B) shows the expression of *pfnita_optEc* using 1 mM IPTG, the lower concentrations did not induce protein expression.

The *ecnirc* and *ecfoca* constructs were expressed using different conditions to the ones for the *pfnita_opt* gene. These changes were 0.1 mM IPTG and 22 h growth after induction at 37 ^{o}C. The results presented in Figure 3.13, B includes a western blot for FocA (lane 2) and NirC (lane 3) using optimised conditions. It was found that EcNirC and EcFocA were expressed at relatively high levels.

3.4.1.2 Nitrite transport assays

A nitrite transport assay was used to determine if the *pfnita* optimised genes would complement an *E. coli* strain lacking the endogenous nitrite transporter, *ecnirc*. Initially two control strains were selected: (i) JCB 4018 containing the WT *ecnirc* gene and lacking the nitrite/nitrate transporters *ecnaru* and *ecnark* as the positive control, and (ii) JCB 4520, mutant in all three genes to act as the negative control. MOPS buffer at pH 8.5 was used as the assay buffer, it was essential to have a high pH. This ensures that no nitrite could diffuse into the cells to give false positive readings. It has been shown that in *A. nidulans*, nitrite diffusion does not occur readily until pHs lower than pH 6.5 (Unkles *et al*, 2011).

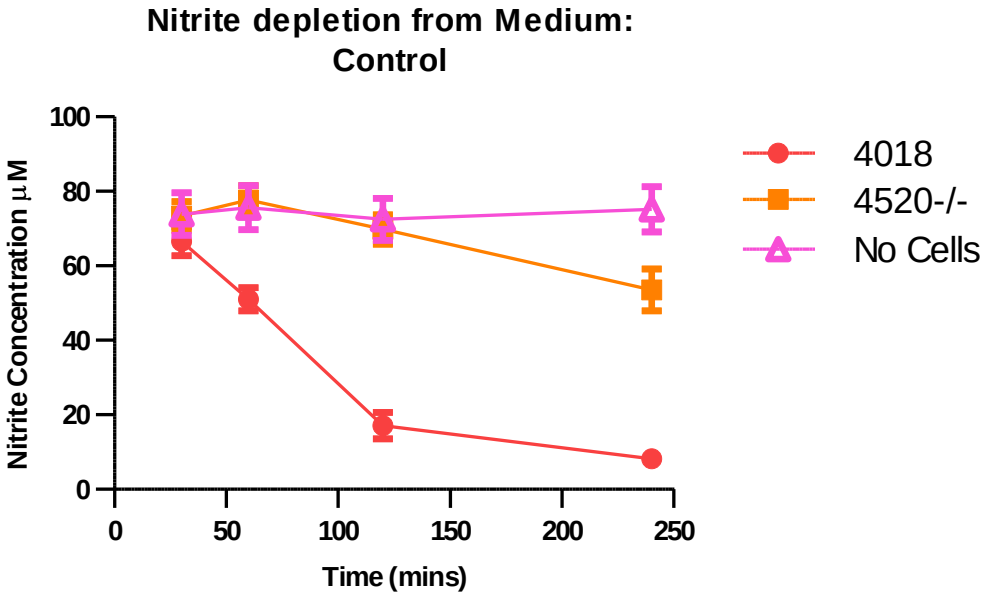
To detect the levels of nitrite present in the medium, the Griess Assay was used. NED and sulphanilamide were combined with an aliquot of each of the samples in triplicate and the results read on a spectrophotometer. The production of a pink/purple azo dye (from colourless) confirmed the presence of nitrite in the media. This meant that it was known, early on, if the controls were working as expected and we had proof of correct experimental design. Bacterial cells were resuspended in MOPS buffer and 100 µM sodium nitrite added to the assay medium to induce and therefore initiate the nitrite transport. The nitrite uptake assay was carried out over 4 h, and samples were taken at 4 time points to check for the disappearance of nitrite from the medium (and therefore uptake into cells). The assay was repeated 4 times, and samples from the two strains were taken in duplicate. An average value for each sample at each time point was taken and used to construct a graph (Figure 3.14). There was some variation observed in the levels of uptake between colonies and between experiments, but, the trend was similar and is reflected in the chart (Figure 3.14, A). For a second negative control, assay buffer with 100 µM sodium nitrite and no cells was included (designated No Cells). The chart in Figure 3.14, A shows the depletion of nitrite from the assay medium over the 240 mins of the assay. Strain 4018 showed a reduction in the concentration of nitrite from the medium to approximately 10 µM which is indicative of nitrite uptake into the WT cells expressing EcNirC. The negative controls $4520^{-/-}$ and No Cells did not show the same reduction in assay medium nitrite concentration. After approximately 120 mins, strain $4520^{-/-}$ shows a reduction in nitrite concentration to approximately 60 µM thus exhibiting intermediate levels of nitrite uptake. The pH of the assay medium was sampled after 240 mins to determine if the

pH had dropped which might explain the apparent uptake of nitrite in $4520^{-/-}$. The pH was still above pH 8.0 after 240 mins confirming nitrite entry to the cell through means other than passive diffusion. It was decided that in future assays that the reaction should be halted after 120 mins. It is of interest, that even though 100 µM of nitrite (final concentration) was added to the assay medium none of the samples show this concentration at T_0.

The results presented in Figure 3.14, B shows a box and whisker plot for the 120 mins time points. The box represents the high quartile (top line), median (middle line) and low quartile (bottom line). The mean is represented by a +, and the upper and lower whiskers represent the range. The data revealed in B shows the levels of nitrite remaining in the medium after 120 mins. At this time the levels of nitrite transport by the two negative controls, $4520^{-/-}$ and No Cells, are similar with a mean value of approximately 70 µM. The range is greater for the $4520^{-/-}$ sample than the No Cells control which is not surprising as variation is more likely to be observed in samples with cells than without cells.

The nitrite assay results showed clear differences between the positive and negative controls and therefore the two test strains, $4520^{-/pfnita\_opt}$ and $4520^{-/pfnita\_optEc}$, were next investigated. As a further positive control, the endogenous *ecnirc* gene was expressed on a plasmid (see above) to ensure that complementation could take place (sample $4520^{-/+}$). Two additional negative controls were used, (i) heat killed JCB 4018 and therefore non-viable cells (the WT positive control strain was chosen to ensure the heat treatment had fully killed the cells) (ii) No Cells. An *ecfoca* complementing strain was also included in the experiments.

(A)



(B)

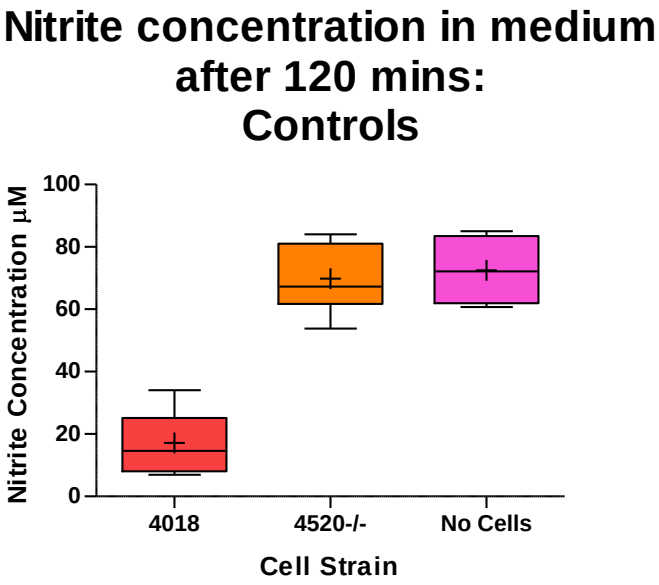


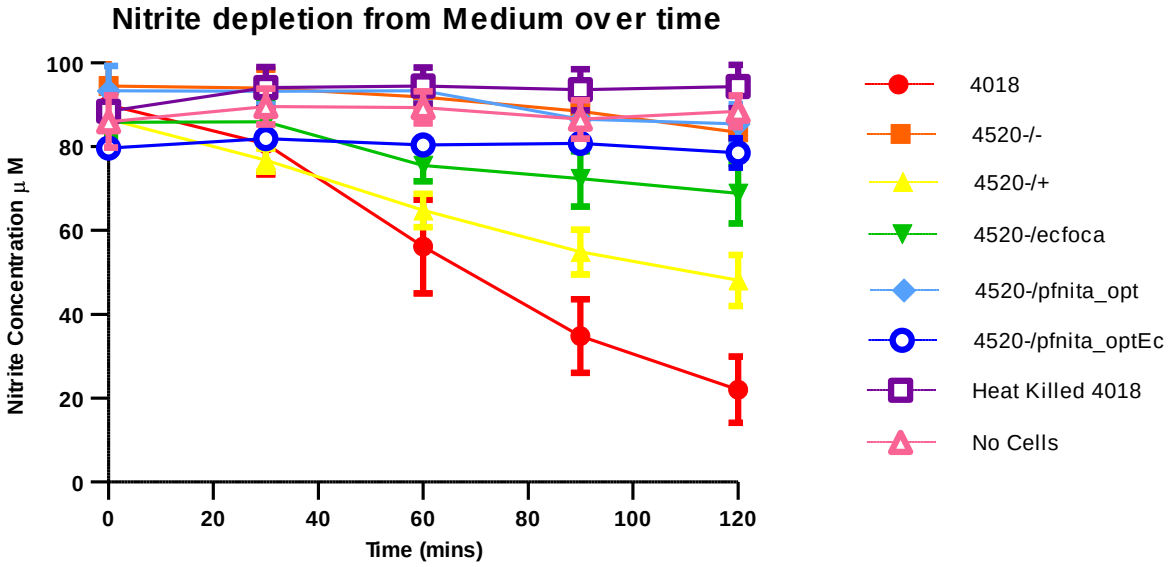
Figure 3.14: Graph of nitrite depletion from the medium in control samples. (A) The control strains were investigated for the uptake of nitrite from the medium by sampling the concentration of nitrite that remained in the medium at various time points over 240 mins. Depletion of nitrite from the medium equates to transport of nitrite into the cell. Error bars indicate the standard error of the mean, 4018 and $4520^{-/-}$ are from 8 replictates and No Cells is from 3. (B) The results at 120 mins were made into a box and whisker plot to demonstrate the difference in the uptake patterns with regard to the positive and negative controls. The box represents the high quartile (top line), median (middle line) and low quartile (bottom line). The mean is represented by + and is from 8 (4018 and $4520^{-/-}$) or 4 (No Cells) replicates, the upper and lower whiskers represent the range. Raw data can be found in Appendix 7.1.1.

The information presented in Table 3.4 highlights the results of nitrite concentration as evaluated from the Griess assay, with the average shown and the range in brackets for each time point as well as for each of the strains. Charts were created for the sample results in which the depletion of nitrite from the assay medium was recorded over the 120 mins assay incubation time. The charts are displayed in Figure 3.15. As before, the results shown in one chart (A) shows the depletion of nitrite from the medium and a second box and whisker plot data (B) for 120 min time point displays the variation in nitrite transport levels between the strains.

	µM nitrite				
E. coli strain	0 mins	30 mins	60 mins	90 mins	120 mins
4018	89.90 (75.24-100.95)	80.57 (59.05-87.62)	56.19 (14.29-76.19)	34.8572 (8.57-59.05)	22.03 (9.52-50.48)
$4520^{-/-}$	94.48 (77.14-116.19)	93.90 (81.91-106.67)	91.81 (81.91-101.91)	88.38 (80.95-100.11)	83.43 (76.19-95.24)
$4520^{-/+}$	86.48 (72.38-109.52)	76.76 (69.52-83.81)	64.76 (54.29-75.24)	54.86 (40.95-64.76)	48.00 (31.43-60.95)
$4520^{-/ecfoca}$	85.72 (72.38-92.38)	85.91 (70.48-100.00)	75.43 (62.86-85.71)	72.29 (52.38-90.95)	68.76 (44.73-86.67)
$4520^{-/pfnita\_opt}$	93.33 (78.10-107.62)	93.14 (86.67-104.76)	93.33 (86.67-104.76)	86.48 (81.91-90.47)	85.33 (77.14-102.86)
$4520^{-/ pfnita\_optEc}$ *	79.55 (75.20-83.16)	81.87 (77.89-86.84)	80.39 (78.71-84.21)	80.75 (77.08-85.96)	78.46 (69.44-87.90)
Heat Killed 4018	88.38 (80.00-100.95)	94.10 (86.67-113.33)	94.48 (80.95-107.62)	93.52 (83.81-110.48)	94.29 (82.86-112.38)
No Cells	85.90 (69.52-105.71)	89.52 (80.95-105.71)	89.33(81.91-103.81)	86.48 (76.19-100.95)	88.38(81.91-102.86)

Table 3.4: The rate of nitrite depletion from the medium. This table contains the average concentration of nitrite that remains in the media at each time point, followed by the range in brackets. The average is from 5 replicates. * The data for strain $4520^{-/ pfnita\_optEc}$ was collected by Dr. S. E. Unkles. Raw data can be found in Appendix 7.1.1.

(A)



(B)

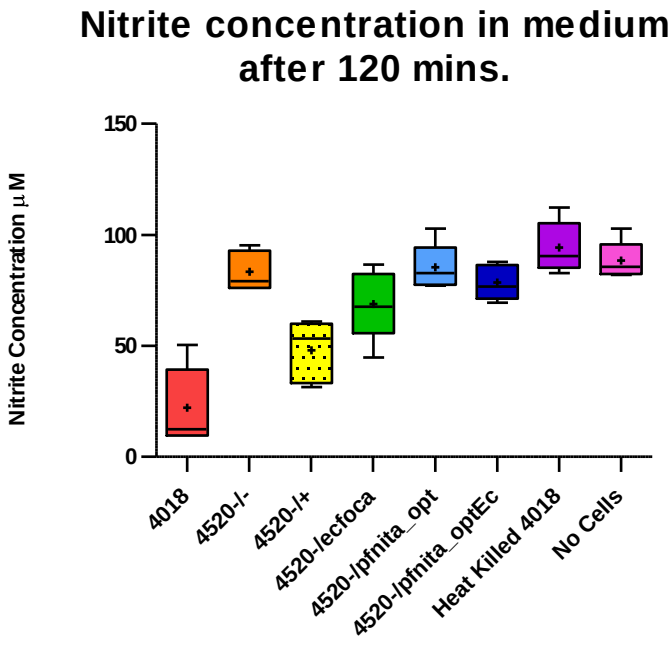


Figure 3.15: Graph of nitrite depletion from the medium. (A) All strains were examined for the disappearance of nitrite from the medium by sampling the concentration of nitrite that remained in the medium over 120 mins. Depletion of nitrite from the medium equates to transport of nitrite into the cell. Values are the mean of 5 replicates, error bars show the standard error of the mean (SEM). (B) The results at 120 mins were presented as a box and whisker plot to demonstrate the difference in the uptake between the positive and negative controls. The box represents the high quartile (top line), median (middle line) and low quartile (bottom line). The mean is represented by + and is from 5 replicates, the upper and lower whiskers represent the range. Raw data can be found in Appendix 7.1.1.

It can be seen from the data in Figure 3.15, strain $4520^{-/+}$ (48.00 µM at T_{120}) does not restore the average uptake value of nitrite to that of WT value where 22.03 µM remained in the medium after 120 mins. The *pfnita_opt* test strains $4520^{-/pfnita\_opt}$ and $4520^{-/pfnita\_optEc}$ show very little alteration in the levels of nitrite in the media illustrating that little to no nitrite uptake took place over the time period. The average nitrite concentration remaining in the medium for strain $4520^{-/pfnita\_opt}$ was found to be 85.33 µM (from 93.33 µM at T_0) and for strain $4520^{-/pfnita\_optEc}$ is 78.46 µM (from 79.55 µM at T_0). These results are consistent with the data obtained for the three negative controls, $4520^{-/-}$ (at T_{120} 83.43 µM), heat killed strain 4018 (at T_{120} 94.29 µM) and sample No Cells (at T_{120} 88.38 µM). The $4520^{-/ecfoca}$ strain shows a slight increase in nitrite transport compared to the negative controls (at T_{120} 68.67 µM compared to 83.43 µM for strain $4520^{-/-}$), but does not reach the levels of nitrite transport attained by the *ecnirc* complementing strain ($4520^{-/+}$).

3.4.1.3 Statistical analysis

A one-way ANOVA (analysis of variance) was carried out to examine if there is any significant differences between the levels of nitrite in the media from each of the samples at the 120 min time point. Bonferroni post-hoc tests were used to compare each of the means from the sample results.

The positive control strain 4018 shows significant differences ($p < 0.001$) between all samples except strain $4520^{-/+}$ where the KO has been complemented with a plasmid containing *ecnirc* where $p = 0.059$, and thus loses significance (Figure 3.15 and Table 3.5). Strain $4520^{-/+}$ the re-expresser positive control, shows significant differences when compared with the negative control strains, and no significant differences to 4018 and the EcFocA complementing strain $4520^{-/ecfoca}$. The negative control strains, Heat Killed 4018 and No Cells controls show significant differences with strains 4018 and $4520^{-/+}$, $p < 0.001$ and $p = 0.002$, respectively. These data confirm that the positive and negative controls are significantly different from each other.

The two test strains $4520^{-/pfnita\_opt}$ and $4520^{-/pfnita\_optEc}$ show a p value of 1.000 when compared to each other, the three negative controls and $4520^{-/ecfoca}$. These values are $p < 0.001$ when compared to 4018 the WT positive control, $p = 0.001$ ($4520^{-/pfnita\_opt}$) and $p = 0.012$ ($4520^{-/pfnita\_optEc}$) when compared to the re-expresser positive control

$4520^{-/+}$. Overall, these data imply that the *pfnita* synthetic codon-optimised genes are unable to complement the loss of function nitrite transport defective mutant of *E. coli* and therefore is incapable of transporting nitrite at least in this bacterium. A table containing the relevant statistical data is provided below (Table 3.5).

Strain	Compared to	Significance
4018	$4520^{-/-}$	0.000*
	$4520^{-/+}$	0.059
	4520^{-}/*ecfoca*	0.000*
	4520^{-}/*pfnita_opt*	0.000*
	4520^{-}/ *pfnita_optEc*	0.000*
	Heat Killed 4018	0.000*
	No Cells	0.000*
$4520^{-/-}$	4018	0.000*
	$4520^{-/+}$	0.002*
	4520^{-}/*ecfoca*	1.000
	4520^{-}/*pfnita_opt*	1.000
	4520^{-}/ *pfnita_optEc*	1.000
	Heat Killed 4018	1.000
	No Cells	1.000
$4520^{-/+}$	4018	0.059
	$4520^{-/-}$	0.002*
	4520^{-}/*ecfoca*	0.326
	4520^{-}/*pfnita_opt*	0.001*
	4520^{-}/ *pfnita_optEc*	0.012*
	Heat Killed 4018	0.000*
	No Cells	0.000*
4520^{-}/*ecfoca*	4018	0.000*
	$4520^{-/-}$	1.000
	$4520^{-/+}$	0.326
	4520^{-}/*pfnita_opt*	1.000
	4520^{-}/ *pfnita_optEc*	1.000
	Heat Killed 4018	0.068
	No Cells	0.462
4520^{-}/*pfnita_opt*	4018	0.000*
	$4520^{-/-}$	1.000
	$4520^{-/+}$	0.001*
	4520^{-}/*ecfoca*	1.000
	4520^{-}/ *pfnita_optEc*	1.000
	Heat Killed 4018	1.000
	No Cells	1.000
4520^{-}/*pfnita_optEc*	4018	0.000*
	$4520^{-/-}$	1.000
	$4520^{-/+}$	0.012*
	4520^{-}/*ecfoca*	1.000
	4520^{-}/*pfnita_opt*	1.000
	Heat Killed 4018	1.000
	No Cells	1.000
Heat Killed 4018	4018	0.000*
	$4520^{-/-}$	1.000
	$4520^{-/+}$	0.000*
	4520^{-}/*ecfoca*	0.068
	4520^{-}/*pfnita_opt*	1.000
	4520^{-}/ *pfnita_optEc*	1.000
	No Cells	1.000
No Cells	4018	0.000*
	$4520^{-/-}$	1.000
	$4520^{-/+}$	0.000*
	4520^{-}/*ecfoca*	0.462
	4520^{-}/*pfnita_opt*	1.000
	4520^{-}/ *pfnita_optEc*	1.000
	Heat Killed 4018	1.000

Table 3.5: Statistical analyses for nitrite uptake by *E. coli* cells. This table displays the significance as determined by a one-way ANOVA and Bonferroni comparisons. Where the significance is 0.000, $p < 0.001$. An * designates significance. Raw data can be found in Appendix 7.1.2.

3.4.2 Formate as a possible substrate

Three strategies were adopted to investigate the prospect of formate being a substrate for the PfNitA permease. *E. coli* cells containing a loss-of-function mutation in the *ecfoca* gene were transformed with plasmids pTTQ18*ecnirc*, pTTQ18*ecfoca*, pTTQ18*pfnita_optEc*, pTTQ18*pfnita_opt* and pTTQ18*pccir*2 (refer to Chapter 6). A list of strains used is provided in Table 3.6.

E. coli Strain	Strain Designation
Wild Type	MC4100
Mutant 1	REK701$^{-/-}$
ecfoca complementing	REK701 $^{-/+}$
ecnirc complementing	REK701$^{-/ecnirc}$
pfnita_opt complementing	REK701$^{-/pfnita_opt}$
pfnita_optEc complementing	REK701$^{-/pfnita_optEc}$
Mutant 2	RM 201$^{-/-}$
ecfoca complementing	RM 201$^{-/+}$
ecnirc complementing	RM 201$^{-/ecnirc}$
pfnita_opt complementing	RM 201$^{-/pfnita_opt}$
pfnita_optEc complementing	RM 201$^{-/pfnita_optEc}$
*pccir*2 plasmid	RM 201$^{-/pccir2}$

Table 3.6: *E. coli* strains used for formate assays. Each strain in which the formate assays were performed are included in this table, the 'Strain Designation' column gives the names of each strain as referred to in the text.

3.4.2.1 Hypophosphite inhibition assay

Initially a hypophosphite inhibition assay, as demonstrated by Suppmann and Sawers (1994) was used to assay formate transport capability. Hypophosphite (HP) has been reported as a toxic analogue of formate, and thus cells capable of transporting formate and *ergo* HP, would be killed. The structures of formate and hypophosphite are provided in Figure 3.16 for reference. In the 1994 article (Suppmann and Sawers), the OD_{420} of the cells were measured and the WT strain MC4100 showed a 53 % reduction in growth compared to the FocA mutant REK701$^{-/-}$ in the presence of HP concentrations of 100 mM. This approach was repeated using the WT and null mutant strains. However, as cell death was being investigated additionally an optical density of OD_{600} was measured. The test strains, MC4100, REK701$^{-/-}$, REK701$^{-/+}$ and REK701$^{-/pfnita_opt}$ were initially analysed for their ability to grow in WM minimal salts media (Appendix 1) under aerobic conditions. It was found that after an overnight growth period (without IPTG induction) cells containing plasmids, REK701$^{-/+}$ and REK701$^{-/pfnita_opt}$ did not grow to sufficient densities to permit inoculation the test liquid culture

medium. The removal of ampicillin from the growth media allowed the strains to increase in cell density but unfortunately the cells had lost their plasmid. This infers that the complementing plasmids were unable to replicate in WM media. A second minimal medium, M9 minimal salts, was investigated for growth of plasmid containing strains. As before, the levels of growth were poor for all strains when glycerol or glucose was the sole carbon source. In contrast a third medium, namely TGYEP allowed sufficient growth and consequently the bacteria were able to be propagated anaerobically as required for the assay conditions.

The two control strains MC4100 and REK701$^{-/-}$ were studied for inhibition of growth by hypophosphite using various media. WM minimal medium experiments resulted in growth inhibition values of 41 % for strain REK701$^{-/-}$ and 87.5 % for MC4100. The results were obtained from two assays only, but show the WT growth is severely reduced by the addition of hypophosphite compared to the null mutant. Using M9 medium supplemented with glucose the growth inhibition values are as follows: 30.5 % for REK701$^{-/-}$, 68.0 % for MC4100, 37 % for REK701$^{-/+}$ and, 35.5 % for REK701^{-} /*pfnita_opt*. Overall, the growth inhibition responses suggest that the two complementing strains are similar to the negative control and do not appear to transport HP and ergo formate. The inhibitory effect of hypophosphite could not be repeated using TGYEP in our laboratory as inhibition values of only 84 to 93 % compared to the null mutant strain (if growth of null mutant with hypophosphite is 100 %) were observed.

A second null mutant strain RM201 was obtained, and the experiment repeated using TGYEP media. Using the WT strain, the reductions in growth observed were between 77 % and 85 % compared to the RM201 null mutant. Given the very low growth inhibition between the WT and two FocA null mutants, REK701$^{-/-}$ and RM201$^{-/-}$ it was decided not to pursue this assay.

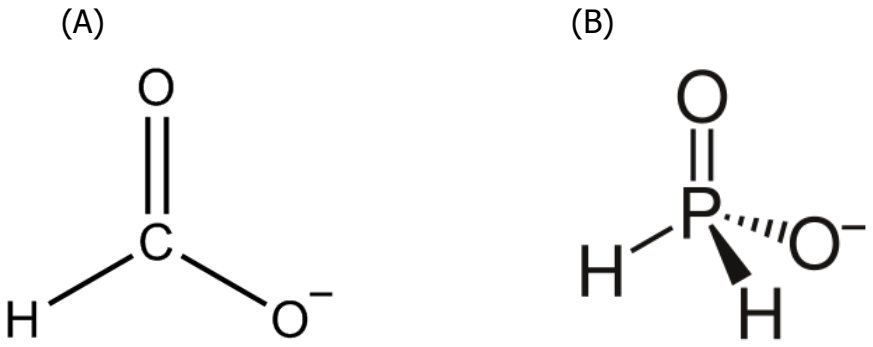


Figure 3.16: Chemical structure comparison. (A) The chemical structure of formate and (B) hypophosphite demonstrating similarities.

3.4.2.2 Radioactive tracer assay

A more direct way of characterising the uptake of formate, using tracer C^{14} formic acid, was used. From this it was believed that data on the transporters kinetics could also be acquired.

The two control strains MC4100 and REK701$^{-/-}$ were tested for differences in the uptake of C^{14} formic acid. Initially MOPS buffer (containing 10 mM glucose) was used at pH 8.5 in an attempt to prevent possible diffusion of formic acid across the cell membranes. Bacterial cells representing each strain were grown anaerobically in TGYEP liquid media for 22 h at 37 ^{o}C in static culture. Cells were pelleted and resuspended in a sodium formate/C^{14} formic acid mixture, and incubated in a 37 ^{o}C water bath for the duration of the assay. A 50 mM sodium formate concentration was chosen as standard. At different time points between 1 and 40 mins, a 1 ml sample was taken. Samples were rapidly filtered and washed three times in 50 mM sodium formate. The filter membranes were tested for radioactivity using a scintillation counter. An aliquot of the sodium formate/C^{14} formic acid mixture was also taken at the end of each experiment and used as a control. Five microlitres of the control sample gave scintillation readings between 9014.50 and 11972.80 units whilst none of the control samples gave readings as high as this (Table 3.7) although slight differences were observed between the MC4100 and REK701$^{-/-}$ strains with the WT having slightly higher values, indicating higher levels uptake into these bacterial cells. As noted previously, the diffusion of nitrite in the fungus *A. nidulans* only occurs at pHs below 6.5, the assay was repeated in MES buffer at pH 6.5 (Table 3.8). A minor difference is observed in the uptake of formate between the two cell lines, unlike the

experiment carried out using MOPS buffer, the uptake is greater into the *ecfoca* mutant. Each experiment was carried out in duplicate. The values for uptake of C^{14} formic acid into the control strains imply the uptake was very low and a different approach was undertaken.

Mins	1	5	10	20	40
MC4100	66.8	49.5	88.0	153.9	265.1
REK701	35.1	47.8	102.0	103.1	169.8

Table 3.7: Formate uptake over time using MOPS buffer pH 8.5.

Mins	1	5	10	20	40
MC4100	33.8	40.0	122.9	84.6	203.3
REK701	37.1	55.3	97.1	124.3	176.8

Table 3.8: Formate uptake over time using MES buffer pH 6.5.

The experiment was repeated using a single time point of 20 mins, and the concentration of sodium formate was varied between 50 and 400 mM to circumvent problems due to formate concentration being insufficient to produce any measurable uptake. As before, the two buffers MOPS pH 8.5 and MES pH 6.5 were tested. Only one experiment was carried out for each buffer as the levels of uptake were again poor. The data is presented in tables 3.9 and 3.10.

[Formate]	50	100	200	400
MC4100	18.8	23.3	20.3	27.0
REK701	25.8	24.3	19.8	26.0

Table 3.9: Formate uptake using MOPS buffer pH 8.5.

[Formate]	50	100	200	400
MC4100	24.0	25.0	27.0	64.8
REK701	45.5	42.3	113.3	75.0

Table 3.10: Formate uptake using MES buffer pH 6.5.

3.4.2.3 β-galactosidase assay

An indirect assay using β-galactosidase fused to the *fdhF* promoter was employed by Falke *et al* (2009) to examine the uptake of formate into bacterial cells. The *fdhF* promoter is induced by formate. The bacterial strain $RM201^{-/-}$ contains a fusion and is a *ecfoca* null mutant. Plasmid constructs of interest were transformed into the null mutant strain, $RM201^{-/-}$, to generate strains $RM201^{-/+}$, $RM201^{-/ecnirc}$, $RM201^{-/pfnita\_opt}$ and $RM201^{-/pfnita\_optEC}$. Strain $RM201^{-/-}$ was used as a negative control. As strain $RM201^{-/-}$ has been shown to have residual β-galactosidase activity (Falke *et al*, 2009) when grown with and without additional formate, a second protein expressing negative control was created. Since residual β-galactosidase activity was found in cells grown without formate it was deemed unnecessary to include a control of cells grown without formate. Heat Killed cells were considered, but as these should lack β-galactosidase activity, they were not included in this series of experiments as the negative control

strain does have residual β-galactosidase activity. The second negative control is the loss-of-function mutant strain expressing the Cir2 protein from the pTTQ18 plasmid (Chapter 6), RM201$^{-/PcCir2}$. Expression of the PcCir2 protein should not facilitate the uptake of formate from the media. There was no positive control of a strain containing both the β-galactosidase fusion and a chromosomal *foca* gene available.

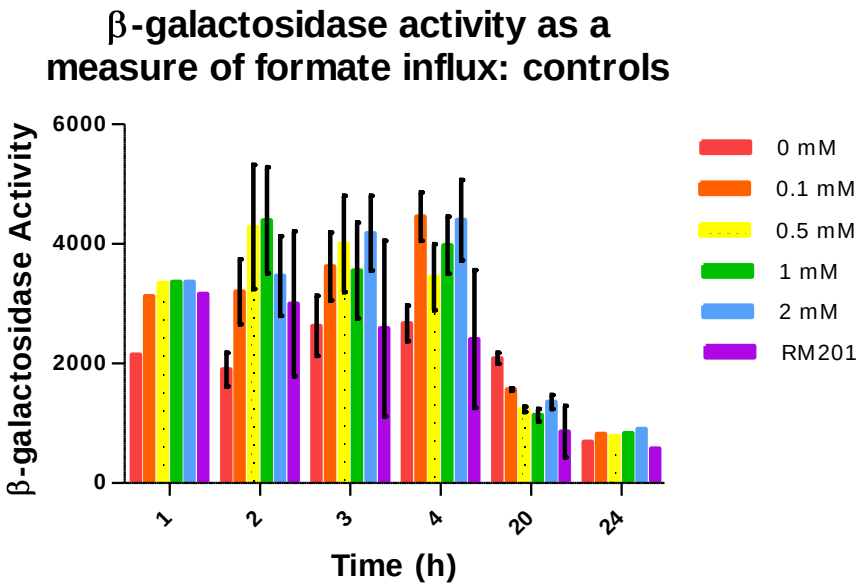
3.4.2.3.1 Study of the control strains

Initial trials were set up using the RM201$^{-/-}$ strain as the negative control and RM201$^{-/+}$ as the positive control re-expressing the EcFocA protein. The objective was to determine if expression of formate carriers in the RM201$^{-/-}$ cell line could restore formate uptake into the cells. These tests were carried out using IPTG concentrations between 0.1 and 2 mM with 1 – 4, 20 and 24 h growth time after induction. The 2, 3 and 4 h assays were repeated in triplicate.

From the results presented in Figure 3.17 it can be observed that for cells grown for 20 and 24 h, the levels of β-galactosidase activity had substantially dropped compared to the 1 – 4 h samples. Moreover the activities at 20 and 24 h were not greatly different to the null mutant. β-galactosidase activity at 1 h after induction was again not particularly different between the induced and mutant strains, and the activity was quite high. Therefore only times of 2, 3 and 4 h were investigated in future assays. The concentration of IPTG which gave the best β-galactosidase activity was found to be 0.1 mM. The uninduced sample yielded a β-galactosidase activity somewhat higher than the RM201$^{-/-}$ strain indicating the expression of *ecfoca* from the TTQ18 plasmid may be leaky allowing expression of the formate permease.

Statistical analyses using one-way ANOVA with Bonferroni *post-hoc* comparisons show no significant differences between any of the treatments or between the time points. The lowest p value was $p = 0.696$ for the 4 h time point for strain RM201 and 0.1 mM IPTG.

(A)



(B)

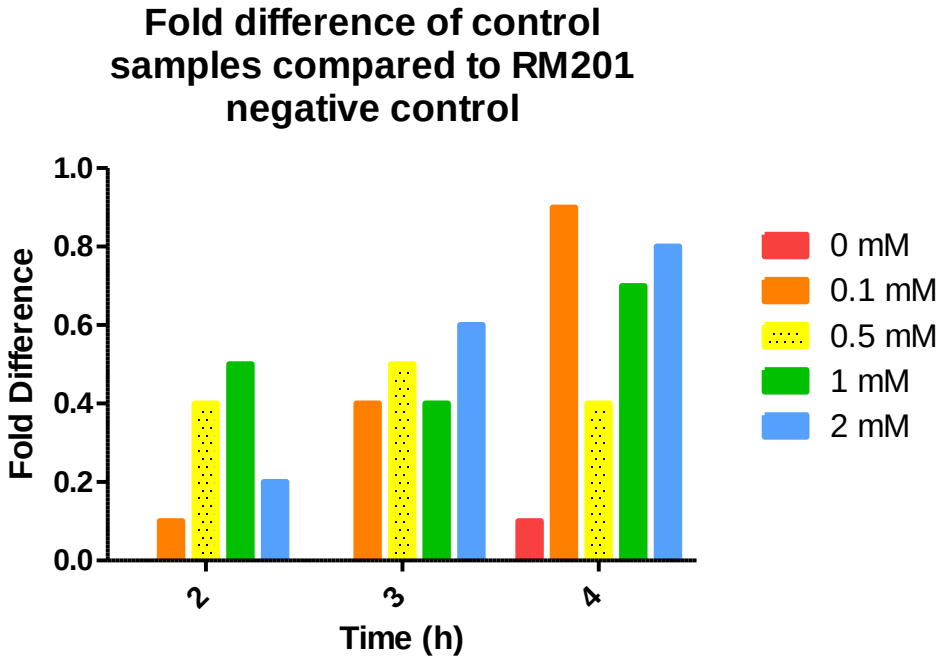


Figure 3.17: β-galactosidase activity of controls. (A) Bar chart showing the β-galactosidase activity in Miller Units on the y-axis against incubation time (h) on the x-axis. The expression of FocA was induced using different IPTG concentrations and the cells incubated for varying times to determine the optimal conditions for this study. Mean values are plotted where possible, the error bars indicate the standard error of the mean. The 1 and 24 h experiments were carried out once only, the 20 h experiment in duplicate and the 2, 3, and 4 h experiment in triplicate. (B) This graph shows the fold difference in the average ß-galactosidase activity of the induced samples when compared to the RM201$^{-/-}$ control (= 0) over each of the time points. β-galactosidase activity is an indicator of formate import into the bacterial cell cytoplasm. Raw data can be found in Appendix 7.2.1.

Therefore 0.1 mM IPTG inducer concentration was chosen for further experiments as it gave the most striking differences compared to RM201$^{-/-}$ and the no IPTG control when presented in the chart in Figure 3.17, A. The lack of significance is most likely due to the large standard deviations observed for each of the samples (Figure 3.17). The results exhibited in Figure 3.17, B shows the fold difference of each of the samples compared to the RM201$^{-/-}$ control strain. The RM201$^{-/-}$ control strain therefore equalled 0. From this chart, the marked difference of the induced samples is clearly observed.

3.4.2.3.2 Expression of complementing proteins in cells grown under anaerobic conditions

Null mutant *ecfoca* cells, RM201$^{-/-}$ were transformed with plasmid constructs pTTQ18*ecnirc*, pTTQ18*ecfoca*, pTTQ18*pfnita_opt*, pTTQ18*pfnita_optEc* (Figure 3.12) pTTQ18*pccir2* to give strains RM201$^{-/+}$, RM201$^{-/ecnirc}$, RM201$^{-/pfnita_opt}$, RM201$^{-/pfnita_optEc}$, and RM201$^{-/pccir2}$ respectively (Table 3.6).

Initially strain RM201$^{-/pfnita_optEc}$ cells were grown in TGYEP liquid medium under anaerobic conditions to test the expression levels of the PfNitA_optEc protein. Trials were established and bacterial cells containing construct pTTQ18*pfnita_optEc* were grown initially for 4 hours at 37 ^{o}C, induced with one of three different IPTG concentrations and were grown for a further 22 h after induction with IPTG. All trials were carried out in TGYEP liquid medium at 37 ^{o}C in a static incubator. *E. coli* crude membrane preparations were analysed for PfNitA_optEc expression using NuPAGE® protein electrophoresis and western blotted against the His-tag fused to the C-terminus (Figure 3.18). A total of 15 µg protein from a crude membrane preparation of induced bacterial cultures was applied to each gel. It was found that induction with 1 mM IPTG after an initial growth period of 4 h with cells grown for a further 22 h showed best expression of PfNitA_optEc (Figure 3.18, A). The band at the expected molecular weight from 0.5 mM IPTG induction (lane 3) is darker than the others, however the gel contains many other products in the lane implying more protein had been applied to the lane. This observation made the results difficult to interpret. Experiments were repeated using 0.1 mM IPTG inducer concentration for strains RM201$^{-/+}$ and RM201$^{-/ecnirc}$ as this concentration had given best results for the functional complementation of the null mutant harbouring the *ecfoca* plasmid (Figure 3.18). The expression trials for strain RM201$^{-/pfnita_opt}$ were probed with an anti-V5 antibody against the C-terminal V5

tag contained on the expressed protein. These western blots failed to produce bands on a gel, even when control strains were included. Therefore strain RM201$^{-/pfnita_opt}$ was not used for the formate assays.

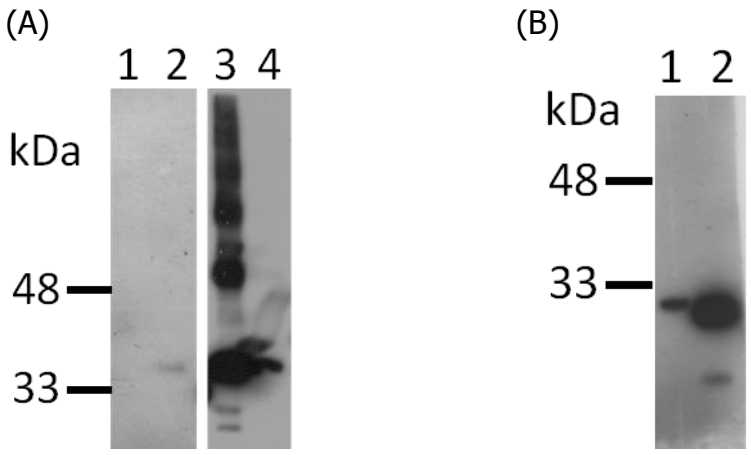


Figure 3.18: Expression of complementing proteins in *E. coli* in TGYEP medium. (A) Western blot to show expression of PfNitA_optEc, after detection with an anti-His antibody, directed against the C-terminal tag on the expressed protein. Lane 1 contains a negative control with no IPTG. IPTG concentrations used were 0.1 mM (lane 2), 0.5 mM (lane 3), and, 1 mM (lane 4). (B) Western blot of EcFocA (lane 1) and EcNirC (lane 2) membrane proteins, using an anti-His antibody directed against the C-terminal tag. The blots confirm expression of all constructs under the assay conditions.

3.4.2.3.3 Formate uptake assay

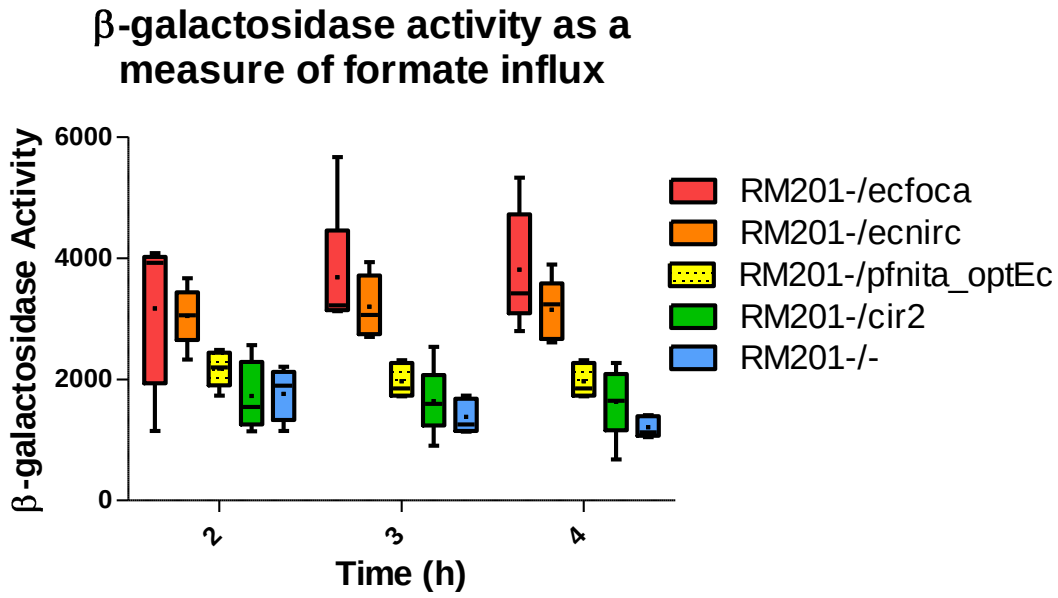
A formate uptake assay was used to determine whether the synthetic *pfnita* genes, optimised in accord with *E. coli* or *A. nidulans* codon usage, would permit uptake in an *E. coli* strain mutated in *ecfoca* and thus lacking its own formate transport ability. There is no positive control strain for this assay, therefore the null mutant strain, RM201$^{-/-}$ transformed with a plasmid containing the *foca* gene acted as the only positive control. The negative control was the untransformed bacterial strain. As a further negative control, strain RM201$^{-/-}$ was transformed with a plasmid containing *pccir*2 which is not a transporter (See Chapter 5). The assay was carried out in cells grown in liquid TGYEP media with a pH value of 6.7. This pH value switches the activity of the transporter from the efflux of formate to importing formate (Sawers, 2005). The *fdhF* promoter possesses a downstream *lacZ* fusion (fdhF::lacZ (Falke *et al*, 2009)), thus if formate enters the cells there should be a concomitant increase in β-galactosidase activity. β-galactosidase activity is recorded using the Miller assay and

thus is presented in Miller Units. Cells were grown as described in the methods section 2.4.1.3. After 2, 3 and 4 h, 1 ml aliquots were taken and the β-galactosidase assay carried out. The assay was repeated at least 5 times on each sample.

β-galactosidase activity (Miller units)			
E.coli strain	2 h	3 h	4 h
RM201$^{-/-}$	1759.169 (1146.667-2206.061)	1380.612 (117.391-1731.915)	1209.104 (1043.165-1407.273)
RM201$^{-/+}$	3167.754 (1146.667-4082.353)	3684.167 (3126.829-5670.588)	3810.153 (2793.750-5326.316)
RM201$^{-/ecnirc}$	3045.330 (2328.571-3666.667)	3196.273 (2700.000-3936.842)	3147.887 (2608.333-3894.737)
RM201$^{-/pfnita_optEc}$	2175.971 (1731.034-2482.759)	1969.726 (1720.000-2311.765)	2034.421 (1626.667-3166.667)
RM201$^{-/pccir2}$	1724.224 (1140.909-2561.905)	1641.829 (904.000-2535.714)	1629.220 (677.143-266.667)

Table 3.11: β-galactosidase activity as an indirect measure of formate uptake. This table contains the average β-galactosidase activity values for the different strains at each time point, followed by the range in brackets. The average of from 5 samples. Raw data can be found in Appendix 7.2.1.

(A)



(B)

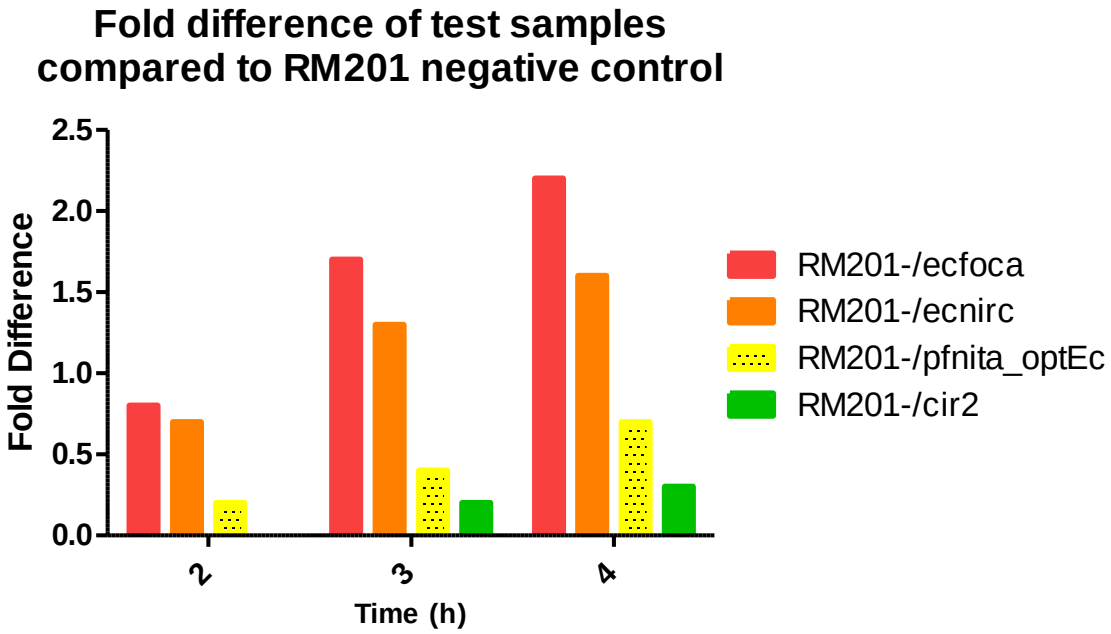


Figure 3.19: β-galactosidase activity. (A) Box and whisker plot showing β-galactosidase activity in each of the strains. The box represents the high quartile (top line), median (middle line) and low quartile (bottom line). The mean is represented by a + and is from 5 replicates, the upper and lower whiskers represent the range. (B) Graph showing the fold difference in average β-galactosidase activity of the test samples (from the five replicates) when compared to the RM201$^{-/-}$ control (= 0) over each of the time points. β-galactosidase activity is an indicator of formate import to the cell cytoplasm. Raw data can be found in Appendix 7.2.1.

An average for each of the samples at the given time points have been used to create the graph presented in Figure 3.19. There was variation observed between the activity levels of the colonies tested. In this regard, Table 3.11 provides the average activity value for each sample with the range in brackets.

3.4.2.3.4 Statistical analysis

A one-way ANOVA (analysis of variance) was carried out to examine if there is any significant differences between the levels of β-galactosidase activity from each of the samples at the 2, 3 and 4 h. Bonferroni post-hoc tests were used to compare each of the means from the samples. As an ANOVA could not be used for the 2 h samples, a Kruskal-Wallis test was. This test showed there were no significant differences ($p = 0.18$) between β-galactosidase activity values in the strains tested. A one-way ANOVA was used for the 3 and 4 h samples and the data is presented in Table 3.12 below.

3 h Time point			4 h Time point		
Strain	Compared to	Significance	Strain	Compared to	Significance
RM201$^{-/-}$	RM 201$^{-/+}$	0.000*	RM201$^{-/-}$	RM 201$^{-/+}$	0.000*
	RM 201$^{-/pfnita_optEc}$	1.000		RM 201$^{-/pfnita_optEc}$	0.520
	RM 201$^{-/pccir2}$	1.000		RM 201$^{-/pccir2}$	1.000
	RM 201$^{-/ecnirc}$	0.002*		RM 201$^{-/ecnirc}$	0.001*
RM201$^{-/+}$	RM 201$^{-/-}$	0.000*	RM201$^{-/+}$	RM 201$^{-/-}$	0.000*
	RM 201$^{-/pfnita_optEc}$	0.004*		RM 201$^{-/pfnita_optEc}$	0.002*
	RM 201$^{-/pccir2}$	0.001*		RM 201$^{-/pccir2}$	0.000*
	RM 201$^{-/ecnirc}$	1.000		RM 201$^{-/ecnirc}$	1.000
RM201^{-} /pfnita_optEc	RM 201$^{-/-}$	1.000	RM201^{-} /pfnita_optEc	RM 201$^{-/-}$	0.520
	RM201$^{-/+}$	0.004*		RM201$^{-/+}$	0.002*
	RM 201$^{-/pccir2}$	1.000		RM 201$^{-/pccir2}$	1.000
	RM 201$^{-/ecnirc}$	0.060		RM 201$^{-/ecnirc}$	0.114
RM201^{-} /pccir2	RM 201$^{-/-}$	1.000	RM201^{-} /pccir2	RM 201$^{-/-}$	1.000
	RM201$^{-/+}$	0.001*		RM201$^{-/+}$	0.000*
	RM 201$^{-/pfnita_optEc}$	1.000		RM 201$^{-/pfnita_optEc}$	1.000
	RM 201$^{-/ecnirc}$	0.009*		RM 201$^{-/ecnirc}$	0.011*
RM201^{-} /ecnirc	RM201$^{-/-}$	0.002*	RM201^{-} /ecnirc	RM201$^{-/-}$	0.001*
	RM201$^{-/+}$	1.000		RM201$^{-/+}$	1.000
	RM 201$^{-/pfnita_optEc}$	0.060		RM 201$^{-/pfnita_optEc}$	0.114
	RM 201$^{-/pccir2}$	0.009*		RM 201$^{-/pccir2}$	0.011*

Table 3.12: Statistical analyses of β-galactosidase activity in strains. This table displays the significance as determined by a one-way ANOVA and Bonferroni comparisons. Where the significance is 0.000, $p < 0.001$. An * designates significance. β-galactosidase activity is an indicator of formate import to the cell cytoplasm. Raw data can be found in Appendix 7.2.2.

At the 3 and 4 h time points, the negative control RM201$^{-/-}$ shows significant differences to strain RM201$^{-/+}$ ($p = 0.000$ at both times) and strain RM201$^{-/ecnirc}$ ($p = 0.002$ at 3 h, and, $p = 0.001$ at 4 h). This result shows that the *ecfoca* and *ecnirc* genes encoded on the complementing plasmids produce proteins that import formate from the extracellular medium and/or bacterial cell periplasm. There are no significant differences to the strains containing the PfNitA gene construct ($p = 1.000$ at 3 h, and, $p = 0.520$ at 4 h) or PcCir2 complementary plasmids ($p = 1.000$). The graphs presented in Figure 3.19 show that the RM201$^{-/pfnita_optEc}$ strain has appreciably more ß-galactosidase activity than either of the negative control strains. The additional negative control strain, RM201$^{-/pccir2}$, shows significant differences to strain RM201$^{-/+}$ ($p = 0.009$ at 3 h, and, $p = 0.011$ at 4 h) and RM201$^{-/ecnirc}$ ($p = 0.002$ at 3 h, and, $p = 0.001$ at 4 h), the same as the RM201$^{-/-}$ negative control. Interestingly, there is no significant differences between strains RM201$^{-/ecnirc}$ and RM201$^{-/pfnita_optEc}$ ($p = 0.060$ at 3 h, and, $p = 0.114$ at 4 h). This suggests the uptake levels between the two strains are fairly similar.

From the results shown in Figure 3.18 it is clear that EcFocA and EcNirC proteins are expressed to higher levels than PfNita_optEc. It is possible that this comparably low level of protein expression is the reason the ß-galactosidase activity observed is so low in cells expressing PfNita_optEc. Densitometry analysis was unable to be carried out as the PfNita_optEc western blot was separate to the blot for EcFocA and EcNirC as further optimisation of expression from the initial studies was required. Taken together, these data imply that formate may be transported by the PfNitA permease. However further study is required before interpretation of this can be confirmed.

3.5 Summary

- The several possible functions for a formate-nitrite transporter in *P. falciparum* were discussed. Nitrite and nitrate have a putative role in the digestive vacuole, possibly in the degradation of haemoglobin to produce amino acids and haem. Formate may be produced as a by product of cellular processes such as glucose catabolism or folate biosynthesis.

- Structural predictions for the *P. falciparum* formate-nitrite transporter suggest an overall topology that is consistent with published crystal structures. The protein contains six transmembrane domains and the C- and N-termini are intracellular. The *P. falciparum* NitA protein sequence contains three amino acid motifs found in all FNT orthologues. The conserved residues in these motifs are proposed to play a role in the selectivity of the transport protein. In addition to this, PfNitA contains a novel sequence motif shared by bacterial formate transporters.

- Phylogenetics places the *Plasmodium* FNT transporters in their own phylogenetic group. This group displays the most recent evolutionary history with other apicomplexan organisms (*Toxoplasma*, *Theileria* and *Babesia*). A streamlined tree showed PfNitA to have the closest evolutionary history with the bacterial formate transporters.

- The codon-optimised synthetic *pfnita_opt* gene was unable to be expressed in terms of activity and protein production in the lower eukaryote fungus *Aspergillus nidulans*. Attempts were made to alter the *pfnita*/plasmid sequences to facilitate expression. Western blotting indicated the mutated PfNitA was not expressed in the fungi. Therefore it remains unknown if the transporter would have been able to complement an *A. nidulans* mutant strain.

- Attempts to express the *Plasmodium* FNT transporter in the bacterium *Escherichia coli* were carried out. Two codon-optimised synthetic *pfnita* genes were chosen for expression in *E. coli* based on bacterial or fungal codon usage rules. Western blots confirmed expression of both the PfNitA_opt and PfNitA_optEc in *E. coli* membranes. The uptake of nitrite by a null mutant (Δ*nirc*) was not significantly different to the

uptake levels by the null mutant overexpressing the synthetic codon-optimised *pfnita* genes. The lack of nitrite uptake from the assay medium could be due to one of two things (i) the expressed protein was not active in the uptake of nitrite from medium or (ii) the protein was not functionally expressed.

- Three methods were pursued to study formate transport in bacterial cells. From these an indirect assay proved to be the most useful. This assay procedure measures β-galactosidase activity as a marker for formate uptake. β-galactosidase activity observed in a null mutant overexpressing the PfNitA_optEc protein was markedly higher than the activity in the null mutant (Δ*foca*). However statistical analysis using a one-way ANOVA confirmed such differences not to be significant. The PfNitA_optEc was poorly expressed under assay conditions and the standard deviations of the samples were quite large. Potentially, if PfNitA_optEc was expressed to higher levels a significant difference may have been observed. The lack of a meaningful outcome strongly indicates that further research into the possibility that PfNitA functions as a formate transporter needs to be carried out (e.g. characterisation in *Xenopus* oocytes).

- The addition of a His or V5 tag to the C-terminal of the protein was deemed not to have a detrimrntal effect on the oloigomerisation and activity of the recombinant PfNitA_opt and PfNitA_optEc proteins as the addition of a C-terminal His-tag to the endogenous EcFocA and EcNirC proteins did not prevent the proteins from functioning in the uptake of formate and/or nitrite. A V5 tag added to the C-terminal of the endogenous AnNitA protein does not affect the function and activity of the protein (Dr. S.E. Unkles, personal communication).

4 Genetic Manipulation of *Plasmodium falciparum nita*

4.1 Introduction

The overall objective of the line of study into the characterisation of the PfNitA protein was to assess whether the PfNitA transporter is suitable as a novel chemotherapeutic target. Drug targets are required to be genetically and chemically validated. In the research described in this chapter the subcellular localisation of PfNitA was established using episomal expression of a tagged synthetic codon-optimised (for *A. nidulans*) PfNitA protein in *Plasmodium falciparum*. Research efforts were made to (i) generate a genetically modified cell line deficient in the *pfnita* gene using double cross-over recombination, and (ii) to investigate its essentiality using reverse genetics. In addition, we wanted to demonstrate that the gene locus was targetable using a knock-in approach to tag the native gene.

4.2 Subcellular Localisation of PfNitA

4.2.1 Subcellular localisation predictions

Orthologues of the PfNitA permease are found in all apicomplexan parasites excluding *Cryptosporidium* which does not contain an apicoplast (Zhu *et al*, 2000). Therefore it is possible that the PfNitA protein might be present in the apicoplast membrane. The apicoplast is bound by four membranes, and therefore has a requirement for permeases to transport molecules into and out of the organelle. To investigate a possible apicoplast location two apicoplast targeting motif predictors, PlasmoAP (Foth *et al*, 2003) and PATS (prediction of apicoplast targeted sequences) (Zuegge *et al*, 2001), were used. Further localisation predictors were employed, that determined (i) a possible mitochondrial location using PlasMit (Bender *et al*, 2003) and iPSORT (Bannai *et al*, 2002), and (ii) the presence of signal peptides using iPSORT and SignalP (Petersen *et al*, 2011). A list and overview of the prediction programs utilised are provided in Section 2.6.6. The full-length amino acid sequences of all known PfNitA orthologues from *Plasmodium* species were studied. Each program employed returned a negative result for a signal peptide and for localisation to the mitochondrion or apicoplast (Table 4.1). The amino acid sequence does not encode a known signal

sequence to direct it outside the parasite and to the RBC such as the PEXEL/VTS motif (Hiller *et al*, 2004, and Marti *et al*, 2004). Proteins are known to be transported to the RBC using other routes (Girolamo *et al*, 2008). It has been predicted that the cellular location for PfNitA is the parasite plasma membrane (PPM), due to the hydrophobic nature of the protein and annotation as a transporter. A PPM location has already been inferred in Figure 1 of Martin *et al*, (2009a).

	PATS	PlasmoAP	MitoProt	PlasMit	iPSORT	SignalP
P. falciparum	No	No	No	No	No	No
P. chabaudi	No	No	No	No	No	No
P. berghei	No	No	No	No	No	No
P. yoelii	No	No	No	No	No	No
P. knowlesi	No	No	No	No	No	No
P. vivax	No	No	No	No	No	No

Table 4.1: Subcellular localisation predictions. PATS and PlasmoAP predict targeting to the apicoplast, MitoProt, PlasMit to the mitochondrion and SignalP, iPSORT predict targeting sequences. The full-length PfNitA amino acid sequence and orthologue from other *Plasmodium* spp. were analysed using prediction software programs. An output from the program that suggests positive location is indicated by Yes and a negative result by No.

4.2.2 Biochemical subcellular localisation

To determine the subcellular localisation of the PfNitA protein, the synthetic codon-optimised *pfnita* gene (*pfnita_opt*) was episomally expressed in *P. falciparum*. The 927 bp *pfnita_opt* gene was amplified using the primers NitAopt GW F and NitAopt GW R (Appendix 2), and cloned upstream of a 3xHaemagglutinin ($(HA)_3$) epitope tag or green fluorescent protein (GFP) tag using the MultiSite Gateway® technology (Invitrogen) (Tonkin *et al*, 2004, and van Dooren *et al*, 2005). Expression of the cloned gene was placed under the control of the *Hsp86* promoter in both plasmids. The final transfection plasmids were termed pCHD-Hsp86-*pfnita_opt*-GFP and pCHD-Hsp86-*pfnita_opt*-$(HA)_3$, and are displayed in Figure 4.1, A and B, respectively. An overview of the MultiSite Gateway® cloning procedure is provided in Figure 2.4.

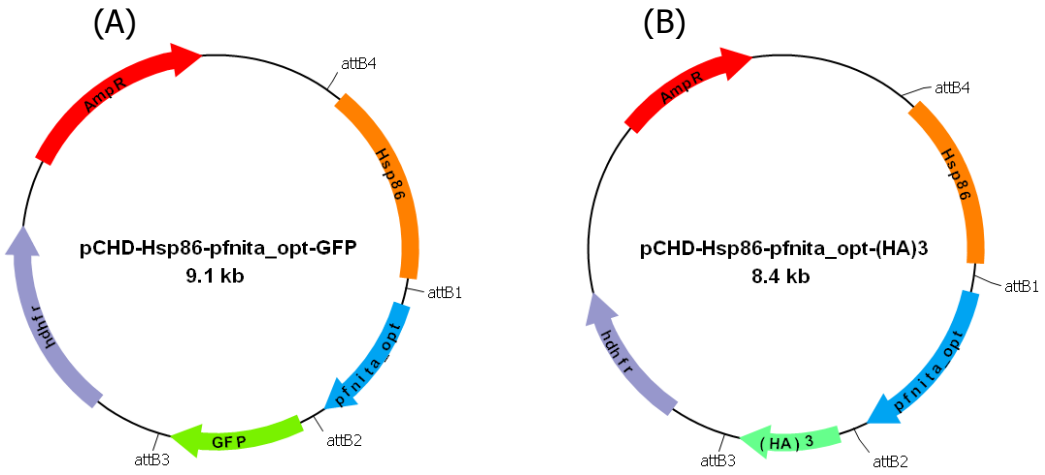


Figure 4.1: Subcellular localisation transfection plasmids. The final plasmids (A) pCHD-Hsp86-*pfnita_opt*-GFP and (B) pCHD-Hsp86-*pfnita_opt*-$(HA)_3$ encode the codon-optimised *pfnita* gene and a GFP or $(HA)_3$ tag, respectively. These plasmids were transfected into 3D7 and D10 parasites. The *att* sites that bring the constructs together during the gateway cloning LR reaction are indicated. The ampicillin and hDHFR selectable markers used in bacterial cloning and parasite transfection/selection procedures, respectively, are also given.

Given that the codon-optimised *pfnita* gene was under transcriptional control of the *Hsp86* promoter, its mRNA expression profile was analysed for an Hsp86 family protein (PF07_0030) and for *pfnita* using the PlasmoDB website[28] and information therein (Le Roch *et al*, 2003). The only obvious differences in the mRNA expression profiles of the two genes are during the merozoite and gametocyte stages where the *pfhsp86* family gene is expressed at slightly elevated levels than *pfnita* (Figure 4.2). During the sporozoite stage the *pfnita* transcript is expressed at higher levels than that of *pfhsp86*. Therefore, the substitution of the native promoter with the Hsp86 promoter would not be expected to cause marked overexpression of the PfNitA protein. Interestingly, analysis of the *P. falciparum* proteome found PfNitA protein to be present in the gametocytes but not in gametes (Supplementary Table B, Lasonder *et al*, 2002).

Two strains of *P. falciparum* were transfected, namely 3D7 and D10. The main differences between these strains are that D10 cells can grow to a higher parasitaemia *in vitro* than 3D7 and do not produce gametocytes. Parasites were thus transfected with the plasmids pCHD-Hsp86-*pfnita_opt*-GFP or pCHD-Hsp86-*pfnita_opt*-$(HA)_3$ on two separate occasions (1 and 2).

[28] www.plasmodb.org. Accessed on 01 January 2012.

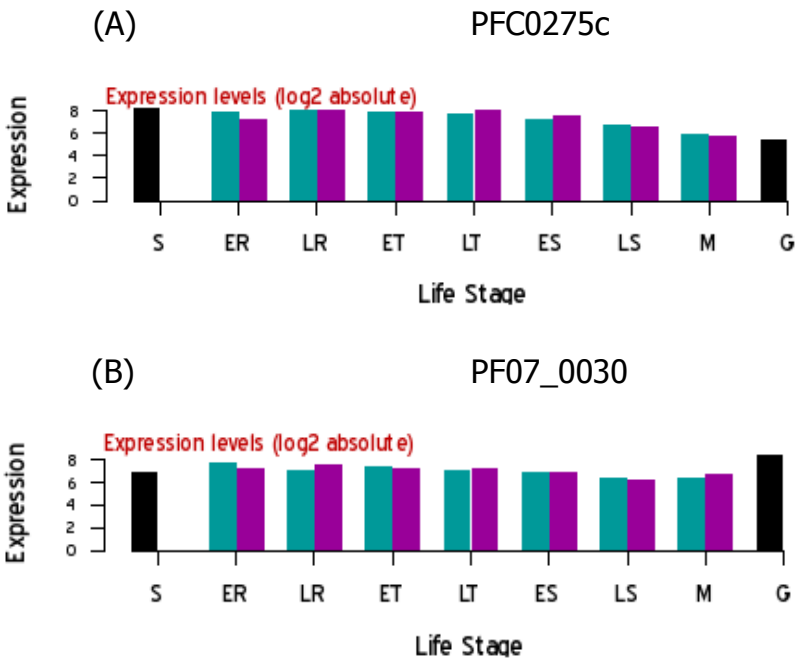


Figure 4.2: Expression profiles of PFC0725c (*pfnita*) and PF07_0030. Relative mRNA expression level of the genes intensity is compared relative to all other genes for a given experiment. Green bars represent stage synchronization using sorbitol, purple bars represent stage synchronization by temperature. (A) PFC0725c, (B) PF07_0030. S = sporozoites, ER = early rings, LR =late rings, ET = early trophozoites, LT = late trophozoites, ES = early schizonts, LS = late schizonts, M = merozoites, G = gametocytes. Images taken from PlasmoDB[29].

These plasmids encode the human dihydrofolate reductase (*hdhfr*) gene conferring resistance to Walter Reed Institute anti-DHFR antifolate 99210 (WR 99210). Transfected parasites resistant to WR 99210 were observed in thin blood smears approximately four weeks after transfection, giving rise to strains D10$^{pfnita_opt\text{-}GFP}$-1 and -2, 3D7$^{pfnita_opt\text{-}GFP}$, and, D10$^{pfnita_opt\text{-}HA}$-1 and -2, 3D7$^{pfnita_opt\text{-}HA}$-1 and -2.

4.2.2.1 Expression and localisation of GFP tagged PfNitA

To confirm that the correct sized protein had been expressed, protein was extracted from transfected parasites for western blot analysis. The extracted protein was separated into insoluble and soluble fractions, with the insoluble fraction being washed well to remove the majority of remaining cytoplasmic proteins (Section 2.5.10). These fractions were separated by a 7.5 % SDS-PAGE gel, blotted onto nitrocellulose and the GFP tagged PfNitA visualised using anti-GFP and HRP conjugated secondary antibodies

[29] www.plasmodb.org. Accessed on 01 January 2012.

(antibody details are provided in Table 2.3). The expected size of the fusion protein is approximately 64.8 kDa (ExPASy Compute pI/Mw tool). The results presented in Figure 4.3 confirm the protein was expressed well, and that it was found in the soluble fraction as well as the insoluble pellet. The band size of the blot is smaller than expected, however this is not unusual for hydrophobic proteins and the reason for discrepancy may be related to the large GFP tag (Rath *et al*, 2008). No dissociated GFP (approximately 30 kDa) was detected by the anti-GFP antibody in the western blots (data not shown). The blot was reprobed with an anti-BCKDH (branched chain α-keto acid dehydrogenase) antibody and HRP conjugated secondary antibody (Table 2.3) as a loading control.

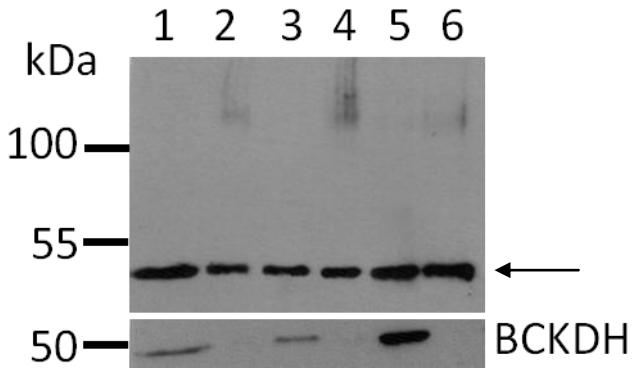


Figure 4.3: Expression of GFP tagged PfNitA_opt. Western blot of the insoluble and soluble parasite protein fractions using an anti-GFP antibody directed against the C-terminal tag confirms the full-length protein expression. Lanes 1, 3, and 5 contain the soluble parasite protein fraction and lanes 2, 4, and 6 the insoluble. Lanes 1 and 2 are D10$^{pfnita_opt\text{-}GFP}$-1, lanes 3 and 4 are D10$^{pfnita_opt\text{-}GFP}$ -2, and lanes 5 and 6 are 3D7$^{pfnita_opt\text{-}GFP}$. The blot was stripped and re-probed with an anti-BCKDH antibody as a loading control. 20 µg total protein was loaded in each lane, the arrow indicates the position of PfNitA-GFP.

BCKDH shows equal expression throughout the intraerythrocytic developmental cycle (IDC) and is a soluble mitochondrial protein (Günther *et al*, 2007). This showed that there was no contamination of the insoluble fractions with the soluble fraction, though the insoluble fraction may still be found in the soluble fraction. The fact that the PfNitA_opt-GFP protein was found in the soluble fraction as well as insoluble suggests the protein may not have folded correctly and may be found in the cell cytoplasm. All GFP transfected cell lines were subject to fluorescent live microscopy to determine the subcellular localisation of PfNitA-GFP. Cell lines were analysed by light microscopy using an AxioSkop-2 fluorescence microscope (Zeiss, Herts., UK).

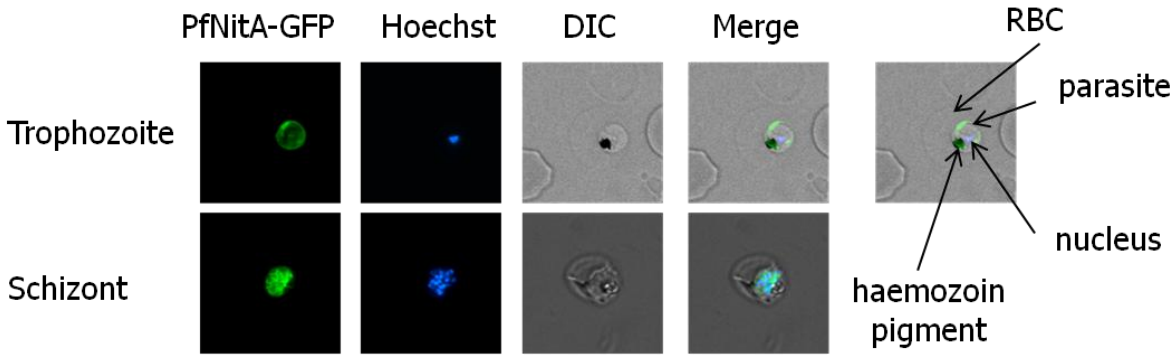


Figure 4.4: Localisation of epitope tagged PfNitA using live cell microscopy. To investigate the subcellular localisation of PfNitA the synthetic codon-optimised *pfnita_opt* gene was C-terminally tagged with GFP using Multi site Gateway® technology, and expressed episomally in D10 and 3D7 parasites. The images are from D10 parasites only. The images show the synthetic codon-optimised PfNitA protein tagged with GFP under the control of the Hsp86 promoter (PfNitA-GFP). Overlays of the images are provided to establish potential co-localisation (Merge). The merge 'Trophozoite' image has been duplicated to highlight the important components of the image: RBC, parasite, nucleus and haemozoin pigment (digestive vacuole). Hoechst 33258 was used as a nuclear stain. Images were obtained by DIC microscopy, or in the FITC (GFP) and the DAPI (Hoechst) fluorescent channels. DIC: differential interference contrast microscopy.

All lines exhibited GFP fluorescence at the periphery of the parasite cells indicative of a plasma membrane or parasitophorus vacuole membrane location of PfNitA-GFP. A low number of images showed good resolution of the brightfield (differential interference contrast microscopy (DIC)) parasites such as presented in Figure 4.4. Data obtained from the two parasites in Figure 4.4 suggest the PfNitA_opt protein is located in the parasite plasma membrane (PPM) as the individual merozoites show GFP fluorescence in the schizont image, although the individual merozoites cannot be discerned in the DIC image. Further means of microscopy were investigated to gain better overall images. D10$^{pfnita_opt\text{-}GFP}$-1 showed a stronger fluorescent signal when cells were viewed using the AxioSkop-2 microscope and were therefore selected for analysis by Delta Vision microscopy (Figure 4.5).

The merged images presented in Figure 4.5 clearly show the PfNitA_opt-GFP fusion protein localises to the PPM or parasitophorus vacuole membrane (PVM) of all parasite IDC stages. Evident from 12.9% (n = 31) of trophozoite and 31.6 % (n = 19) of schizont images was GFP fluorescence surrounding the haemozoin pigment and therefore digestive vacuole (DV).

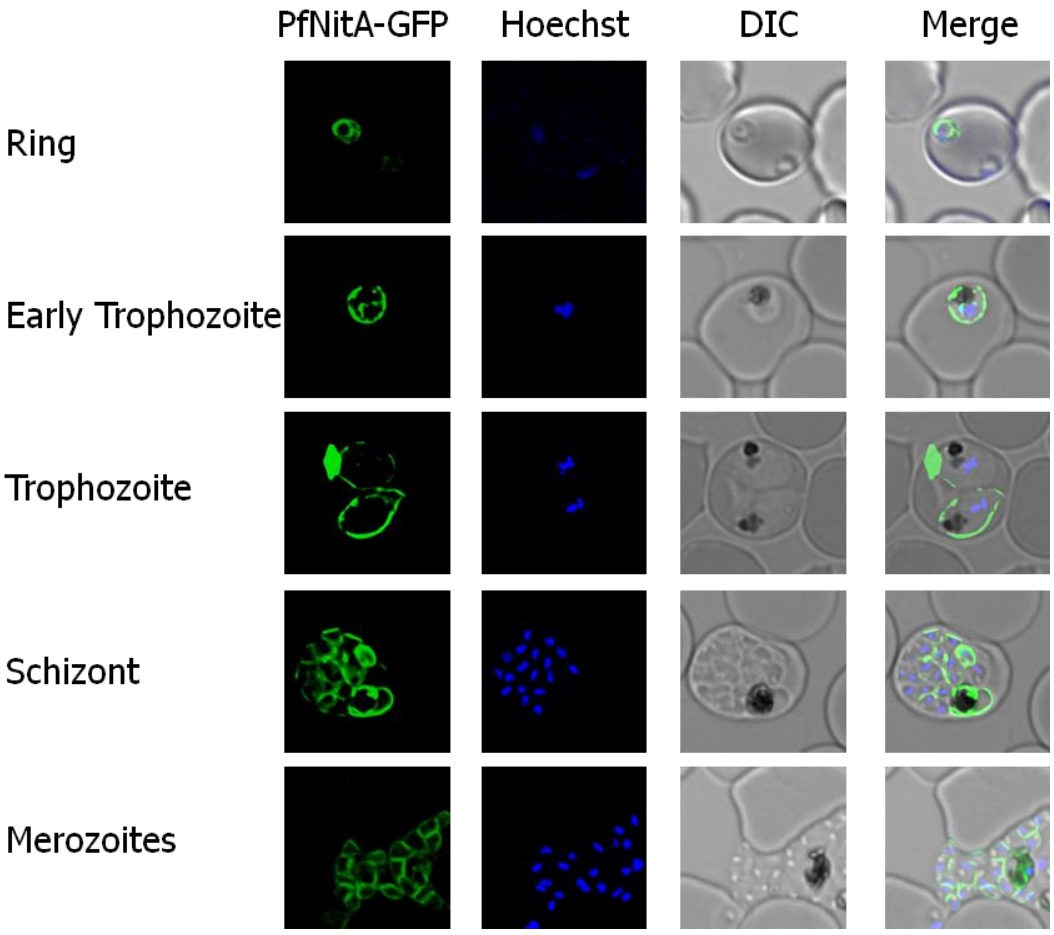


Figure 4.5: Localisation of PfNitA using Delta Vision microscopy. To investigate the subcellular localisation of PfNitA the synthetic codon-optimised PfNitA_opt was C-terminally tagged with GFP using Multi site Gateway® technology and expressed episomally in D10 parasites. The images show PfNitA_opt tagged with GFP under the control of the Hsp86 promoter. Images were taken at indicated stages of the intraerythrocytic developmental cycle. In the 'Trophozoite' image, two parasites are present in a single RBC. Overlays of the images are provided to show potential co-localisation (Merge). Hoechst 33258 was used as a nuclear stain. Images of infected RBCs were obtained by DIC microscopy, and by using the DAPI (Hoechst) and GFP (GFP) channels.

Furthermore, 41.4 % (n = 29) of ring stage parasites showed staining associated with two membranes (as presented in Figure 4.5 'Ring') compared to 31.0 % (n = 29) with a single (outer) membrane fluorescence (data not shown). Two images were captured showing merozoites recently released from the RBC, these showed PfNitA_opt-GFP localisation to the plasma membrane of free merozoites. The peri-nuclear stain recorded in the 'Early Trophozoite' stage was observed in 19.4 % (n = 31) of trophozoite images. This staining may be the endoplasmic reticulum (ER) and is possibly indicative of the transport of the fusion protein. A single bright fluorescent

area was recorded in certain images from both trophozoite and schizont stages (Figure 4.5 'Trophozoite'). It is thought that the punctate fluorescence could be due to vesicle transportation of the fusion protein to the cell membrane, or perhaps an artefact due to overexpression of the synthetic fusion protein.

4.2.2.2 Confirmation of subcellular localisation using immunofluorescence analysis of $(HA)_3$ tagged PfNitA

To confirm that the PfNitA-$(HA)_3$ fusion protein was being expressed, total protein was extracted from transfected parasites for western blot analysis. The total parasite protein was separated into insoluble and soluble fractions with the insoluble fraction being washed thoroughly to remove soluble proteins. Western blot analysis of these fractions using an anti-HA antibody and HRP conjugated secondary antibody revealed a band of approximately 37 kDa which is in line with the predicted size of the fusion protein at 41.7 kDa (ExPASy Compute pI/Mw tool) (Figure 4.6). The results in Figure 4.6, A illustrate the full-length protein with $(HA)_3$ tag is expressed well in all cell lines.

P. falciparum 3D7 transfected cell lines possess a protein band present in both the soluble and insoluble fractions. This implies there may be $(HA)_3$ tagged fusion protein present in the cytoplasm possibly due to the overexpression of the fusion protein. As the 3D7 cell lines display expressed $(HA)_3$ tagged protein in the soluble fractions, we decided to use only the *Plasmodium* D10 cell lines for further analysis. The western blot was reprobed with the anti-BCKDH antibody as a loading control. This experiment confirmed that there was a reasonable separation of soluble and insoluble fractions for D10. However, a certain level of contamination of the fractions was observed in the 3D7 samples. Therefore D10$^{pfnita_opt\text{-}HA}$-1 was chosen for further analysis by fluorescence microscopy.

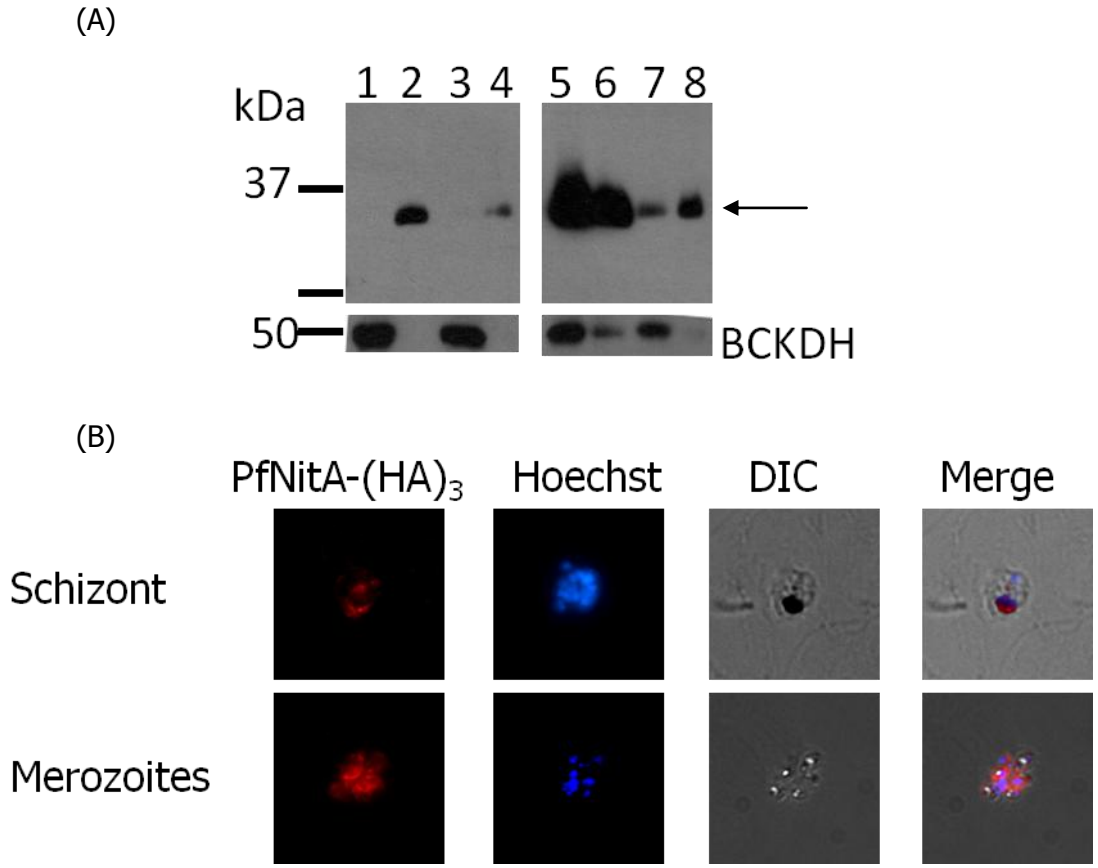


Figure 4.6: Localisation of epitope tagged PfNitA using immunofluorescence analysis (IFA). (A) Western blot of the insoluble pellet and soluble parasite extract using an anti-HA antibody directed against the C-terminal tag confirms the full-length protein expression. Lanes 1, 3, 5, and 7 contain the soluble parasite extract and lanes 2, 4, 6, and 8 the insoluble pellet. Lanes 1 and 2 are D10$^{pfnita_opt\text{-}HA}$-1, lanes 3 and 4 are D10 D10$^{pfnita_opt\text{-}HA}$-2. Lanes 5 and 6 are 3D7 D10$^{pfnita_opt\text{-}HA}$-1 and lanes 7 and 8 are 3D7 D10$^{pfnita_opt\text{-}HA}$-2. The blot was stripped and re-probed with an anti-BCKDH antibody as a loading control which ensured the fractions had been separated well. 20 µg total protein was loaded in each lane, the arrow indicates the position of PfNitA_opt-HA. (B) To investigate the subcellular localisation of PfNitA the synthetic codon-optimised PfNitA_opt protein was C-terminally tagged with $(HA)_3$ and expressed in D10 and 3D7 parasites. The above images are taken from D10 parasites. The images show the synthetic gene tagged with $(HA)_3$ under the control of the Hsp86 promoter (anti-HA) when probed with anti-HA and secondary Alexa594 antibodies. Overlays of the images are provided to establish potential co-localisation (Merge). Hoechst 33258 was used as a nuclear stain. Images were obtained by DIC microscopy, or in the TRITC (PfNitA-$(HA)_3$/Alexa594) and the DAPI (Hoechst) fluorescent channels.

The cell line D10$^{pfnita_opt\text{-}HA}$ was exploited for immunofluorescence analyses (IFA) with a mouse anti-HA antibody and an anti-mouse-Alexa594 antibody (Table 2.3). DNA was stained with Hoechst 33258. The images presented in Figure 4.6, (B) are typical of those for the IFA of strain D10$^{pfnita_opt\text{-}HA}$. In the 'Schizont' image, the staining of the

individual merozoite plasma membranes was clear. In the trophozoite stages the staining of the parasite plasma membrane was less apparent, but the DV staining was pronounced in some parasites.

The fluorescent live microscopy and IFA results suggests that PfNitA_opt is found in the PPM as fluorescence is observed encircling the parasites of ring, trophozoite, and merozoite stages. If the fluorescence had appeared around the schizont stage parasite and not the merozoites a PVM location would have been proposed. In the trophozoite and schizont stages where a clear DV is indicated by the haemozoin pigment, approximately 50 % of the parasites displayed fluorescence indicative of the DV membrane (DVM). Overall these results suggest a dual localisation for PfNitA to the parasite plasma membrane and digestive vacuole membrane.

4.2.2.3 Co-localisation studies of PfNitA with PfHT or PfCRT

Given the potential PPM and DVM location of PfNitA_opt described in Sections 4.2.2.1 and 4.2.2.2, co-localisation studies were carried out using antibodies raised to a PPM protein and a DVM protein. In this regard, the PPM antibody was directed against the *P. falciparum* hexose transporter (HT) (Woodrow *et al*, 1999) and was a kind gift from Prof. S. Krishna (University of London). The primary anti-PfHT antibody was used at a 1:500 dilution, and the secondary was an anti-goat-Alexa594. The DVM antibody was directed against the chloroquine resistance transporter (CRT) (Fidock *et al*, 2000) and was a kind gift from Dr. P. Bray (Liverpool School of Tropical Medicine). The anti-PfCRT antibody was used at a 1:400 dilution, and the secondary was an anti-rabbit-Alexa594 (antibodies and dilutions are given in Table 2.3).

Strain D10 $^{pfnita\_opt\text{-}GFP}$ was employed for the co-localisation experiments. The GFP signal was enhanced using an anti-GFP antibody to cross-react with the GFP fusion protein and secondary antibody conjugated to Alexa488 (Table 2.3). As in previous experiments (Sections 4.2.2.1 and 4.2.2.2) DNA was stained using Hoechst 33258. Synchronised late stage parasites were used for these experiments and not ring stages. The images presented in Figure 4.7 are of trophozoite parasites, no schizont stages were pictured during the analyses. Initially studies using only the anti-PfCRT or the anti-PfHT antibodies and their conjugates were carried out. The images from the anti-PfHT labelled experiments provided well defined images with clear fluorescence

observed around the parasites demonstrating the PPM location of PfHT. The anti-PfCRT antibody labelled the DV well, but unfortunately yielded non-specific binding to the RBC and parasite (data not shown).

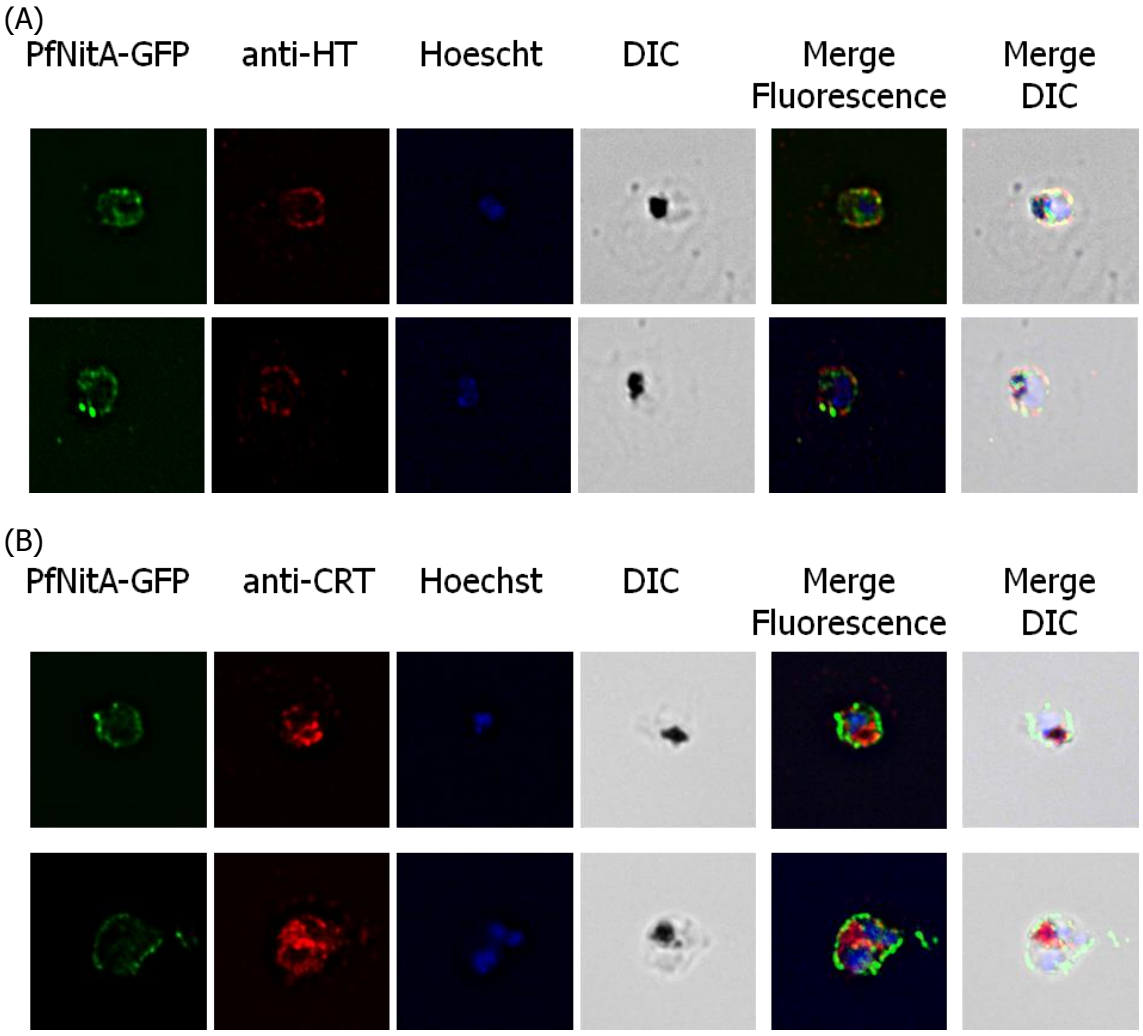


Figure 4.7: Co-localisation studies using immunofluorescence analysis. The subcellular localisation of PfNitA was investigated using antibodies to a known parasites plasma membrane protein, and a known digestive vacuole protein, both of which are transporters. (A) Immunofluorescence analysis using anti-HT (hexose transporter located in the parasite plasma membrane) denoted in red. The GFP tagged synthetic PfNitA is in green and fluorescence was boosted using an anti-GFP antibody. (B) Immunofluorescence analysis using anti-CRT (chloroquine resistance transporter found on parasite digestive vacuole membrane) is indicated in red. The GFP tagged synthetic PfNitA is presented in green. All images were taken from trophozoites. Overlays of the images are provided to establish potential co-localisation (Merge). Hoechst 33258 was used as a nuclear stain. IFA slides were viewed using DIC microscopy, and the DAPI (Hoechst), GFP (GFP and Alex488) and Alexa594 (Alexa594) channels.

The merged images presented in Figure 4.7, (A) show a co-localisation of PfHT (red) with PfNitA_opt (green) both of which surround the parasite. The results confirm the PPM location of PfNitA_opt. The merged images in Figure 4.7, (B) fail to demonstrate co-localisation between the PfCRT antibody (red) and the PfNitA_opt fusion protein (green). The results therefore do not support the DVM location of PfNitA_opt. The IFA experiments were limited in the parasite life cycle stage that was examined, e.g. trophozoite stage and not ring or schizont stages. The anti-PfCRT antibody did not provide a clear localisation to the DV only and was perhaps not the optimum for use in IFA as previous use in the Müller lab had been for western analyses only. The shape of the parasites and lack of clear RBC outline as observed in the DIC images of (B) suggests that the RBC and parasites are distorted perhaps relating to the anti-PfCRT antibody as all IFA experiments were treated the same. Taken together, although the IFA experiment does not support a DVM location for PfNitA_opt this possibility must not be excluded. To confirm the DVM localisation, additional IFA experiments including all the life cycle stages would be of practical value. Further to this, the parasite line encoding the GFP tagged PfNitA_opt could be transfected with a second plasmid encoding, for example, a PfCRT-mCherry fusion protein. Fluorescent live microscopy experiments should be carried out with the new cell line to analyse any co-localisation.

4.3 Knock-out of *pfnita* by Gene Replacement

As noted previously, many of the integral membrane proteins of *P. falciparum* have been demonstrated to be indispensable to the organism (e.g. Slavic *et al*, 2010, el Bissati *et al*, 2008, and, Waller *et al*, 2003). To genetically assess the *pfnita* gene product as a potential drug target, knock-out (KO) studies of the *pfnita* gene were attempted using the pCC1 and pCC4 plasmid constructs (Duraisingh *et al*, 2002 and Maier *et al*, 2006). Given the essentiality of other *Plasmodium* permeases an approach was taken to maximise the likelihood of obtaining KO cell lines after transfection. pCC1 and pCC4 vectors each encode a different drug selectable marker (DSM), namely human dihydrofolate reductase (*hdhfr*) and blasticidin-S-deaminase (*bsd*) conferring resistance to the drugs Walter Reed Institute anti-DHFR antifolate 99210 (WR 99210) and blasticidin-S-HCl (Bla), respectively (Table 2.4) (Fidock and Wellems 1997, and Ben Mamoun *et al*, 1999, respectively). In addition to a positive selection marker, the two pCC plasmids also encode a negative selectable marker. The negative selectable

marker is cytosine deaminase/uracil phosphoribosyl transferase (*cd*), a chimeric gene made from two genes of *S. cerevisiae* (Table 2.4) (Maier *et al*, 2006). The chimera is expressed by cells containing the plasmid and converts 5-fluorocytosine (5-FC) to the toxic 5-fluorouracil and 5-fluorouridine monophosphate. 5-fluorouracil elicits its toxic effects by causing an inhibition of RNA synthesis and 5-fluorouridine monophosphate inhibits thymidylate synthase (Maier *et al*, 2006). For this system to be successful for a gene KO, a double cross-over is required and the gene is replaced by the DSM and not merely disrupted, while the negative selectable marker is lost. A double cross-over is preferable to a single cross-over as in some instances the plasmid can 'loop out' which would leave a functional gene. The double cross-over event is rare in *Plasmodium* parasites, and this negative marker selects for parasites in which the rare event has taken place (Maier *et al*, 2006). The DSMs and their use during drug cycling are described in sections 2.5.3 and 2.5.7.1, respectively.

Fragment	Forward Primer		Reverse Primer		Construct Name
724 bp NitA 5'	NitA_KO_750_5_F	1	NitA_KO_750_5_R	2	*pfnita*750
756 bp NitA 3'	NitA_KO_750_3_F	3	NitA_KO_750_3_R	4	
508 bp NitA 5'	NitA_KO_500_5_F	5	NitA_KO_500_5_R	6	*pfnita*500
542 bp NitA 3'	NitA_KO_500_3_F	7	NitA_KO_500_3_R	8	

Table 4.2: Oligonucleotide primers used for amplification of *pfnita* inserts for KO constructs. Oligonucleotide sequences are given in Appendix 2. The numbers following each oligonucleotide primer correspond to the numbers in Figure 4.8, (A).

For cross-over locations, a 542 bp fragment homologous to a stretch of the *pfnita* 3' open reading frame (ORF), and a 508 bp fragment homologous to the 5' ORF were amplified from 3D7 gDNA. These PCR products were cloned into the pCC1 and pCC4 vectors. Additionally, a 756 bp fragment homologous to part of the *pfnita* 3' ORF and upstream non-coding region and a 724 bp fragment homologous to the 5' ORF and downstream non-coding region were amplified and likewise cloned into the pCC1 and pCC4. The oligonucleotide primer sets used for these amplifications are provided in Table 4.2 and Appendix 2.

The approximately 500 bp DNA stretches covered more of the gene itself, while the approximately 750 bp inserts covered more of the 5' and 3' non-coding regions. This is illustrated in Figure 4.8, (A) and (B). Cloning yielded the plasmids pCC1-*pfnita*500, pCC1-*pfnita*750, pCC4-*pfnita*500 and pCC4-*pfnita*750 (Figure 4.8, (C)).

3D7 and D10 parasites were transfected with the four KO constructs. After five to six weeks, parasites of both strains resistant to Bla or WR 99210 were detected in Giemsa stained thin blood smears.

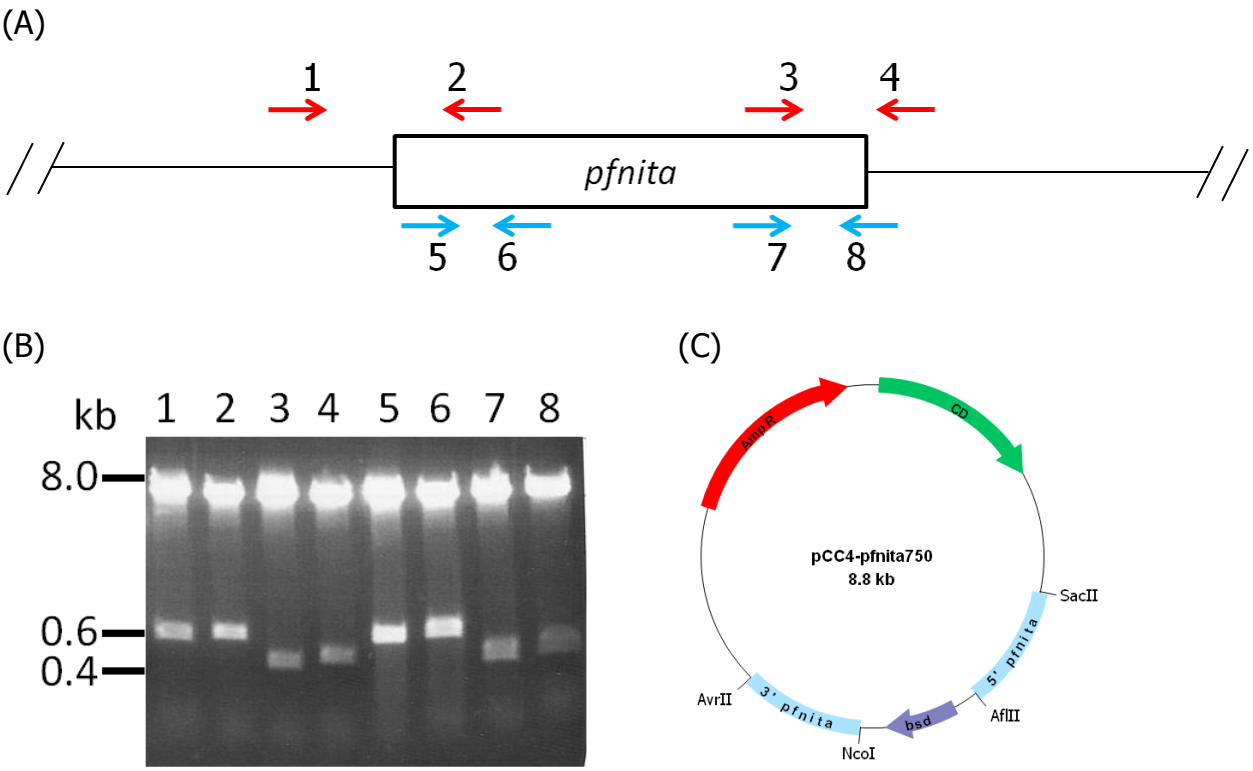


Figure 4.8: Knock-out by gene replacement strategy. (A) Schematic illustrating the PCR fragments generated from *pfnita* and non-coding regions targeted by the knock-out constructs. Red arrows show the 722 and 756 bp fragments and the blue arrows the 508 and 542 bp fragments. Image is not to scale. (B) Digest showing the approximately 500 bp (lanes 3, 4, 7, and 8) and appromimately 750 bp (lanes 1, 2, 5, and 6) 5' and 3' knock-out fragments were amplified from 3D7 gDNA and cloned into the pCC1 (lanes 1 - 4) and pCC4 (5 - 8) vectors encoding the drug selectable markers *hdhfr* and *bsd*, respectively. (C) Example plasmid construct showing the 750 bp 5' and 3' *pfnita* gene fragments cloned into the pCC4 vector and the drug selectable marker blasticidin-S-deaminase (*bsd*). The restriction sites used to clone in the *pfnita* gene fragments are shown, *Nco*I and *Avr*II for the 5' fragment and *Sac*II and *Afl*II for the 3' fragment. The ampicillin resistance (AmpR) cassette is a selectable marker for cloning and selection in bacteria. CD acted as a negative selectable marker during the transfection and integration process. Diagrams are not to scale.

Ring and late stage parasites were observed at a very low parasitaemia (0.1 – 0.2 %) for approximately two weeks after the first parasites were observed, but the parasitaemia never reached higher levels. The parasites maintained abnormal morphology and were always surrounded by a large quantity of debris suggesting

continuous parasite death. The reason most of the transfected cell lines did not form stable cultures is unknown, though it could be speculated that the *pfnita* gene was was being targeted by the KO construct leading to eventual cell death. Only one of these lines, transfected with pCC1-*pfnita*500 (designated pCC1-*pfnita*), was able to establish a WR 99219 resistant culture named D10$^{pCC1\text{-}pfnita}$. Stabilates were made and gDNA generated for Southern blot analysis of the genotype of transfected parasites.

The D10 $^{pCC1\text{-}pfnita}$ cell line was taken through drug cycling to promote integration of the plasmid (summarised in Figure 4.9). The parasite culture was divided into two and drug treatments carried were out as described in Section 2.5.7.1 and illustrated in Figure 2.6. Removing WR 99210 from the culture should result in the loss of the episome and thus the *hdhfr* gene. After three weeks in the absence of drug pressure, the culture was subjected to renewed drug pressure, ensuring that only those parasites that still contained plasmid DNA where the selectable marker had integrated into their genome would survive. It is believed that this drug cycling procedure enriches the culture for parasites where a recombination event has taken place.

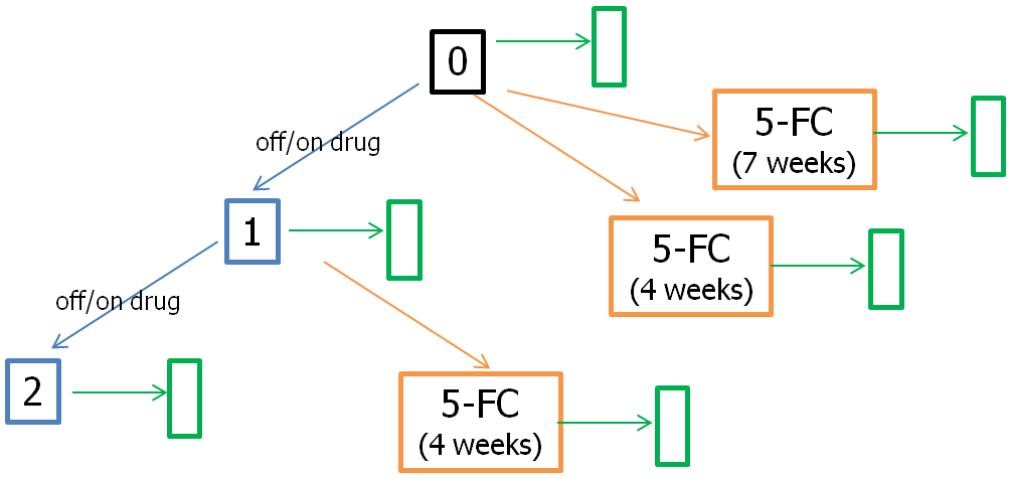


Figure 4.9: Overview of the drug cycling process for parasites transfected with pCC1-*pfnita*. Parasites after transfection, growing on WR 99210 were termed cycle 0 (black box). Drug cycling of transfected parasites was carried out to select for integration of plasmids into the parasite genome. Off/on WR 99210 cycling is displayed by blue boxes/arrows and treatment with 5-FC is indicated in orange. At each stage gDNA was taken for analysis using PCR and/or Southern blot, and stabilates were made for the preservation of the parasite line, as indicated by a green box/arrow combination.

The second subculture was treated with 5-FC. 5-FC is metabolised to the toxic product 5-fluorouracil by the negative selectable marker cytosine deaminase/uracil phosphoribosyl transferase (CD). The product inhibits RNA synthesis and thymidylate synthase (Maier *et al*, 2006). Using the negative selection drives double homologous recombination which leads to replacement of the gene of interest (GOI) by the positive selectable marker and the loss of the negative selectable marker. If a double cross-over event has taken place parasites should be resistant to both WR 99210 and 5-FC.

The D10$^{pCC1\text{-}pfnita}$ parasites that were cycled off WR 99210 for three weeks and subjected to WR 99210 again underwent a slight growth defect for 3-4 days after the drug was added, but were able to establish a typical culture. As soon as normal growth was resumed gDNA was recovered for a diagnostic Southern blot. An overview of the expected banding pattern from a Southern blot when gDNA is digested with *Nde*I is provided in Figure 4.10, (A). The culture was divided into two again for further on/off WR 99210 cycles and 5-FC treatment.

The on/off drug cycling was carried out twice. The parasites recovered from each cycle were termed D10 $^{pCC1\text{-}pfnita\text{-}cyc0}$, D10 $^{pCC1\text{-}pfnita\text{-}cyc1}$and D10 $^{pCC1\text{-}pfnita\text{-}cyc2}$. It took approximately seven weeks for the WR 99210 and 5-FC resistant D10$^{pCC1\text{-}pfnita\text{-}cyc0}$ parasites to appear in a Giemsa stained thin blood smear. This is a relatively long time of exposure to 5-FC as resistance to 5-FC can occur after prolonged exposure to the drug. Subsequently, if parasites did not establish a culture after four weeks they were terminated. Parasites D10$^{pCC1\text{-}pfnita}$ from drug cycles zero, one and two were treated with 5-FC. After four weeks no parasites could be detected in a thin blood smear. An overview of the drug cycling process that was carried out for this parasite line is given in Figure 4.9.

A Southern blot of gDNA isolated from a WT D10 strain and D10$^{pCC1\text{-}pfnita}$ at each cycle was carried out after digestion with the restriction enzyme *Nde*I. The blot was hybridised using a probe homologous to the 3' 542 bp *pfnita* insert, and a second probe was generated against *hdhfr*. The restriction sites and areas covered by the probe are shown in Figure 4.10, (A). The results of Southern blotting are presented in Figure 4.10 (B) shows a 4.46 kb fragment is revealed and confirms the presence of the endogenous gene in all cell lines. A fragment of 2.5 kb would have been present if the

plasmid had integrated at the correct locus. The 0.68 kb fragment shows that the transfected line in cycle 0 does contain the knock-out plasmid. The presence of the additional fragment, approximately 700 bp, on the blot could be due to incomplete digestion of the plasmid or may suggest a random integration event has taken place if pCC1-*pfnita* has integrated into the *P. falciparum* genome at an unknown locus. The Southern blotting experiments revealed that by cycle 2 the plasmid appeared to have been lost when using the gene specific probe, but may still have been present as an approximately 8 kb fragment is present when using the *hdhfr* specific probe corresponding to the plasmid.

Interestingly, the results from the Southern blot containing gDNA from the 5-FC treated culture D10$^{pCC1\text{-}pfnita\text{-}cyc0}$ indicated that no genotypic change had occurred when compared to the pre-5-FC cultures suggesting parasites were resistant via some other means than loss of the KO plasmid (Figure 4.10, (B), lane 3). If integration had occurred, a band of 3 kb would have been present with the hDHFR probe. Using the gene specific probe, a band of 2.5 kb would be expected.

A PCR experiment using primers CD F and CD R (Appendix 2) against the negative selectable marker did not amplify a product in any of the parasites treated with 5-FC (Figure 4.10, (C)) suggesting loss of the negative selectable marker. An overview of the results is given in Table 4.3. A PCR experiment was carried out using primer sets that would amplify a correct sized gene product only if integration had taken place at the intended locus. The 5′ forward primer (F 5′) (5′ KO check int F) is present in the downstream non-coding region of the gene and the 5′ reverse primer (R 5′) (5′ KO check int R) is present on the KO plasmid between the 5′ gene fragment and DSM. Conversely, the 3′ forward primer (F 3′) (3′ KO check int F) is present on the KO plasmid between the DSM and 3′ gene fragment and the 3′ reverse primer (R 3′) (3′ KO check int R) is present in the upstream non-coding region.

The primers are indicated by red arrows in the data presented in Figure 4.10, A. Oligonucleotide primers are provided in Appendix 2. The expected product sizes of 1555 bp for the 5′ product and 1505 bp for the 3′ product would have been amplified only if integration had taken place. No products were amplified from a gradient PCR (data not shown).

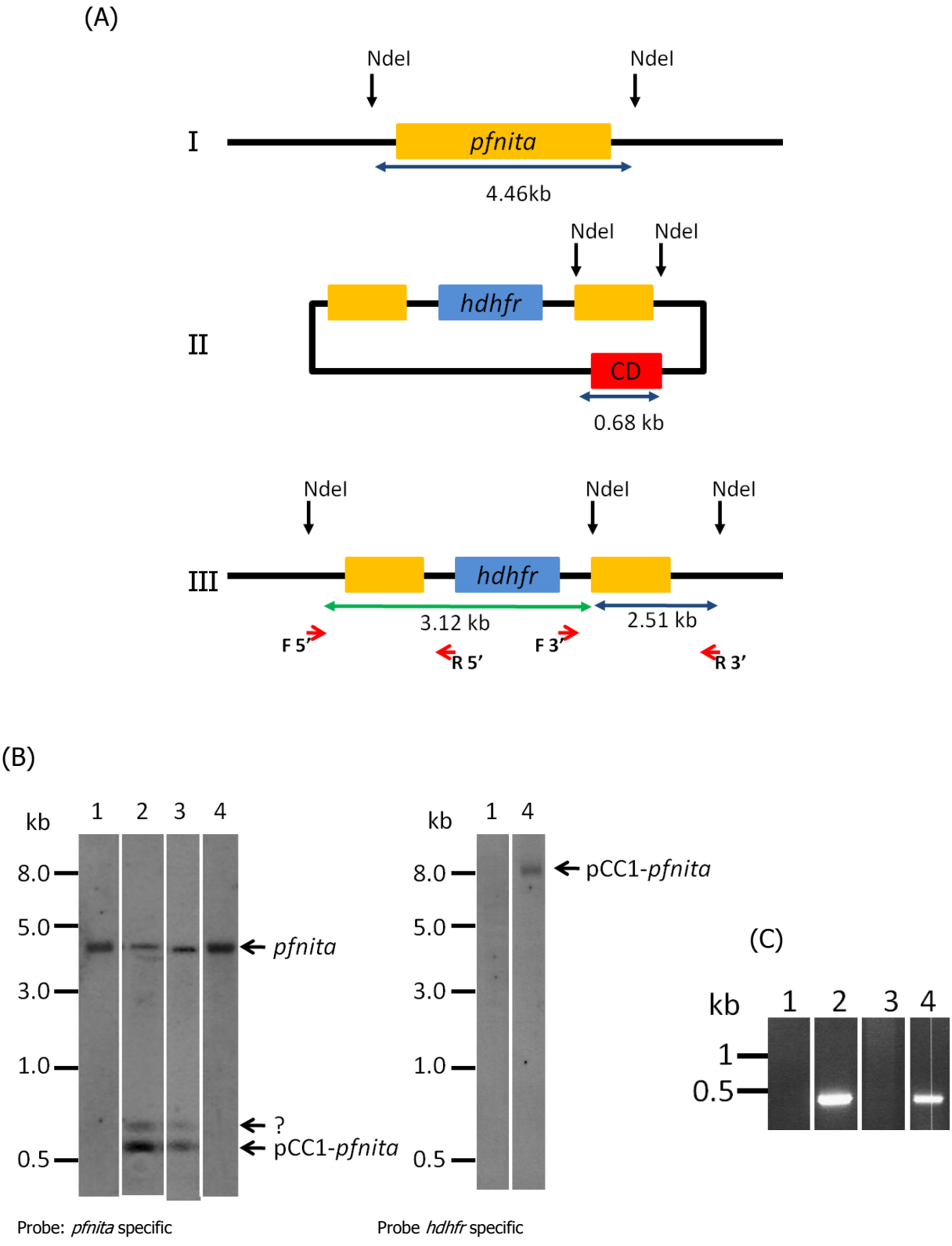


Figure 4.10: Knock-out by double cross-over recombination. (A) A schematic diagram of the *pfnita* wild-type locus (I), the pCC1-*pfnita* plasmid (II) and the *pfnita* gene locus following a double cross-over recombination event (III). *Nde*I restriction sites are indicated by black arrows. The lengths of the

expected fragments detectable with *pfnita* specific probe homologous to the 5' insertion of the plasmid are indicated by blue arrows, and those homologous to the drug selectable marker *hdhfr* are shown in green. The oligonucleotide primers used to test for integration using PCR are indicated by small red arrows in part III. The expected sizes are 1555 bp for the 5' product (F 5' + R 5') and 1505 bp for the 3' product (F 3' + R 3'). (B) Southern blot analysis after digestion of gDNA with *Nde*I, of cell lines transfected with pCC1-*pfnita*, probed with gene specific and *hdhfr* probes. Lane 1 contains D10 WT gDNA, lanes 2 and 3 contain gDNA from D10$^{pCC1\text{-}pfnita}$ in WR 99210 selection cycle 0, lane 3 after 5-FC treatment, and lane 4 contains gDNA in WR 99210 selection cycle 2. (C) PCR using primers against the negative selectable marker CD was carried out to check for loss of the negative selectable marker. Lane 1 contains D10 WT gDNA, lanes 2 and 3 contain gDNA from D10$^{pCC1\text{-}pfnita}$ in WR 99210 selection cycle 0 Lane 3 after 5-FC treatment, lane 4 contains gDNA in WR 99210 selection cycle 2.

Parasite line	PCR	Southern Blot	
	cd	*pfnita*	*hdhfr*
Cycle 0	positive	endogenous gene, KO plasmid	positive
Cycle 0 + 5-FC (seven weeks)	negative	endogenous gene, KO plasmid	positive
Cycle 0 + 5-FC (four weeks)	----------no parasites----------		
Cycle 1	positive	endogenous gene, KO plasmid	positive
Cycle 1 + 5-FC (four weeks)	----------no parasites----------		
Cycle 2	positive	endogenous gene only	negative

Table 4.3: Overview of Southern blot and PCR results from cell line D10$^{pCC1\text{-}pfnita}$. PCRs are either positive or negative for the amplification product from *cd*. Southern blots using the *hdhfr* probe are either negative or positive for the presence of *hdhfr* on the pCC1-*pfnita* plasmid.

A second set of transfections of vectors pCC4-*pfnita*500 and pCC4-*pfnita*750 into 3D7 wild-type parasites was performed in duplicate, no parasites were observed after two months of selection on Bla. Thus, from 20 transfections with four different plasmids and two different parasite strains, a single parasite line containing the KO plasmid was generated. This single parasite line, denoted D10$^{pCC1\text{-}pfnita}$, was taken through drug cycling to promote integration of the plasmid. However, drug cycling resulted in the loss of the plasmid and no integration was achieved.

As lack of integration was observed, it can be inferred that either the gene is refractory to recombination events, or that it is essential to the intraerythrocytic stage of the parasite. An approach where a complementary copy of the gene was co-transfected with the KO plasmid was therefore attempted (Section 4.5).

4.4 "Conditional" Knock-out of *pfnita*

As the previous strategy to produce a knock-out of *pfnita* in *P. falciparum* did not give rise to a viable line lacking the *pfnita* gene, another plan was developed and pursued. In this strategy the codon-optimised synthetic *A. nidulans pfnita_opt* gene was used to complement the knock-out of the endogenous GOI (for example, Slavic *et al*, 2010). It was anticipated that by using the synthetic gene to restore the function of *pfnita*, the endogenous gene should be targetable by a knock-out construct. The *A. nidulans* synthetic codon-optimised *pfnita* gene was chosen as both organisms are eukaryotes. The pCHD-Hsp86-*pfnita_opt*-$(HA)_3$ plasmid was used as this vector encodes a C-terminal tag smaller than GFP and may thus circumvent possible problems of interference with the activity or oligomerisation of the transporter protein. The outcome of this approach relies on the synthetic codon-optimised *pfnita_opt* gene being processed into the active protein. The results presented in Section 4.2 show that the codon-optimised synthetic *pfnita* gene is processed into the full-length protein when expressed in strains D10 and 3D7 *P. falciparum* (Figure 4.6, (A)), and is transported to the PPM (Figure 4.6, (B)).

The pCHD-Hsp86-*pfnita_opt*-$(HA)_3$ plasmid was used for the complementary copy of *pfnita* which encodes a *hdhfr* gene as a DSM. For the KO plasmid pCC4-based constructs were chosen as pCC4 encodes a different DSM, *bsd*. The transfected parasites were therefore selected for using Bla and WR 99210.

Two procedures were tested. (i) the KO plasmid was transfected into cells already harbouring pCHD-Hsp86-*pfnita_opt*-$(HA)_3$. Thus D10 $^{pfnita\_opt\text{-}HA}$ parasites were transfected with pCC4-*pfnita*500 or pCC4-*pfnita*750, on two separate occasions. No parasites were detected in any of the four transfections after eight weeks incubation. (ii) D10 and 3D7 parasite strains were transfected with the two plasmids pCHD-Hsp86-*pfnita_opt*-$(HA)_3$ and pCC4-*pfnita*500 or pCC4-*pfnita*750 simultaneously, and in duplicate. The transfected parasites were selected using Bla and WR 99210. After six weeks, parasites were detected in a single culture from a total of eight co-transfections which had been transfected with the pCC4-*pfnita*500 KO plasmid (referred to as pCC4-*pfnita)* and pCHD-Hsp86-*pfnita_opt*-$(HA)_3$ (designated pCHD-*pfnita_opt*-HA). The single transfected line was designated D10 $^{pfnita\_opt\text{-}HA/pCC4\text{-}pfnita}$.

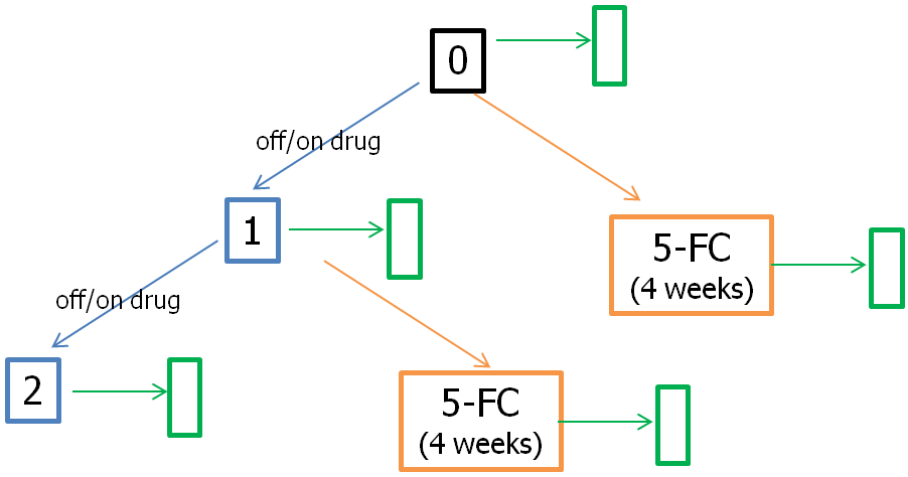


Figure 4.11: Overview of the drug cycling process for parasites transfected with pCHD-*pfnita_opt*-HA and pCC4-*pfnita*. Parasites after transfection, growing both WR 99210 and Bla are termed cycle 0 (black box). Drug cycling of transfected parasites was carried out to select for integration of plasmids into the parasite genome. Off/on Bla cycling is indicated by blue boxes/arrows and treatment with 5-FC is indicated in orange. At each stage gDNA was taken for analysis using PCR and/or Southern blot, and stabilates were made for the preservation of the parasite line. This is indicated by a green box/arrow combination.

Transfected parasites, D10 *pfnita_opt*-HA/pCC4-*pfnita*, were taken through drug selection cycles as described previously (Sections 2.5.7.1 and 4.3) and an overview is provided in Figure 4.11. The parasites were taken through two drug selection cycles where Bla drug pressure was removed and WR 99210 drug pressure was maintained to preserve the episomal copy of the gene. When parasites were growing as normal under drug pressure, gDNA was generated for a diagnostic Southern blot.

The band sizes expected after gDNA restriction with *Nde*I are given in Figure 4.12, (A). In addition a western blot was carried out on soluble and insoluble protein extracted from D10 *pfnita_opt*-HA/pCC4-*pfnita* at cycle 0 (B). The result showed that the synthetic codon-optimised PfNitA protein was expressed well and was present in the membrane containing insoluble fraction as anticipated.

D10 *pfnita_opt*-HA/pCC4-*pfnita* parasites in cycles 0 and 1 were treated with 5-FC to promote the integration of the KO plasmids. After two weeks of continuous culture under the selective pressure of Bla, WR 99210 and 5-FC parasites were observed to be growing normally. Therefore gDNA was generated and stabilates were prepared. In

comparison to the results presented in Section 4.3, the recovery of parasites treated with 5-FC was rapid, two weeks compared to seven weeks. This finding suggested the *cd* gene had been lost and the KO plasmid may have integrated by double cross-over recombination.

A PCR amplification was carried out to amplify the *cd* gene from each of the parasite cycles with and without 5-FC treatment. The results displayed in Figure 4.12, (C) show the *cd* gene is still present during cycles 0, 1, and 2. A very faint product is observed for cycle 0 + 5-FC (lane 4) with no band in cycle 1 + 5-FC (lane 5) illustrating the loss of the *cd* gene from cycle 1 + 5FC parasites. Furthermore, a PCR experiment to check for 5' integration and 3' integration was carried out on the 5-FC treated samples only. The primers were placed up and downstream of the native gene, and up and downstream of the drug selectable marker (Section 4.3 and Appendix 2). These PCR's did not give any products when applied to an agarose gel (data not shown). Without a positive control it is impossible to establish if (i) the PCR is simply not working or if (ii) integration has not occurred. However, at low PCR primer annealing temperatures (40 ^{o}C) it might be anticipated that some products would have amplified due to non-specific binding of the primers to the plasmid construct.

A diagnostic Southern blot was consequently carried out using the restriction enzyme *Nde*I (Figure 4.12). The DNA fragments detected by the *pfnita* gene specific probe are indicated by blue arrows, while fragments expected with a probe specific for the *hdhfr* selectable marker are specified by green arrows Figure 4.12, (A). The *pfnita* probe detected the WT fragment of around 4.5 kb in each of the D10 *pfnita_opt*-HA/pCC4-*pfnita* cell lines, and plasmid of approximately 0.7 kb in the transfected cell lines Figure 4.12, (D). In the event of a double cross-over having occurred, a 2.5 kb fragment would have been expected, but this was not observed in any of the samples. Two high molecular weight fragments (around 6 and 10 kb) were observed in the 5-FC treated samples only. This pattern may indicate recombination between the two plasmids.

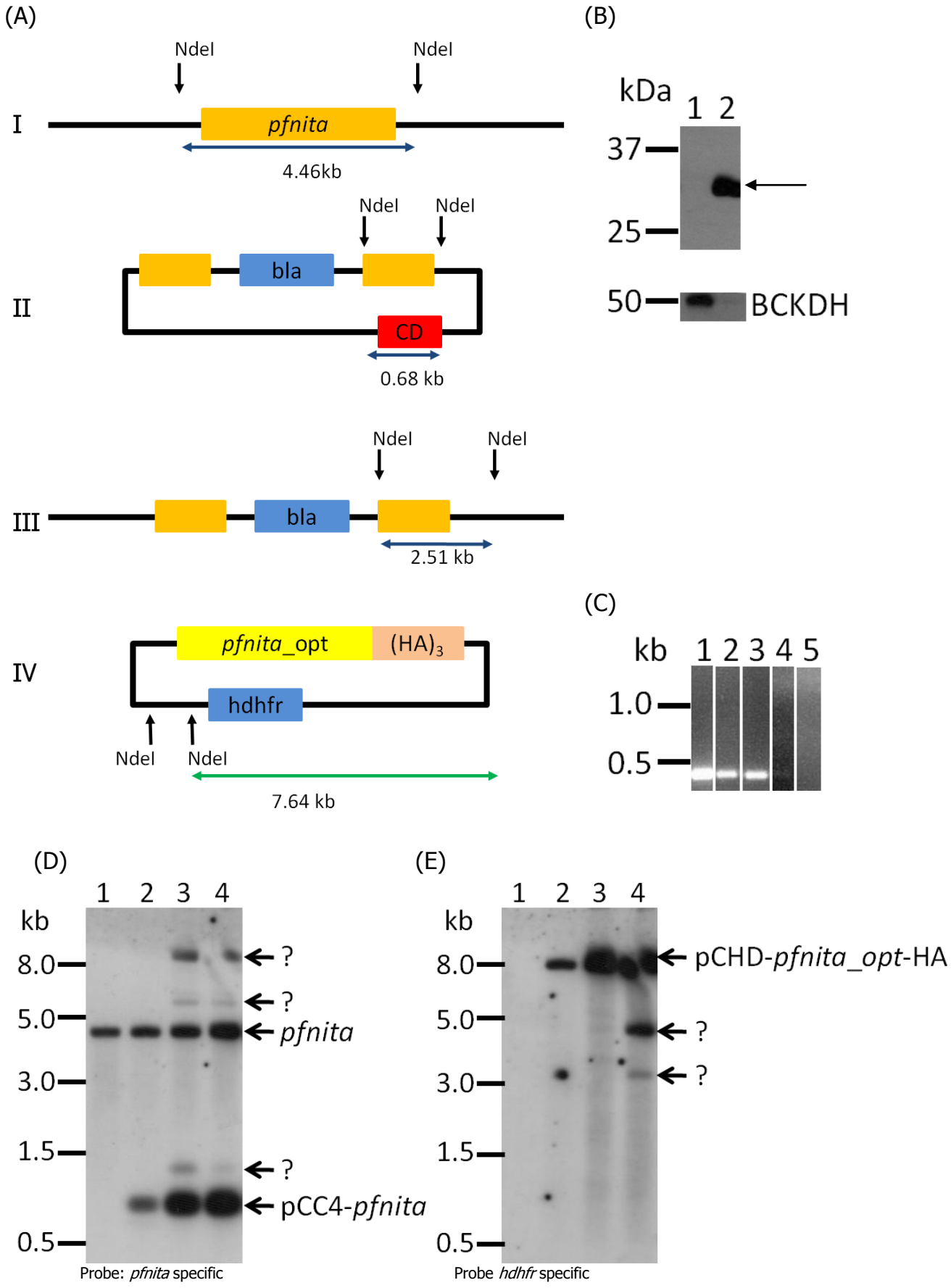

(A)
NdeI
NdeI
I
pfnita
4.46kb
NdeI
NdeI
bla
II
CD
0.68 kb
NdeI
NdeI
III
bla
2.51 kb
IV
pfnita_opt
(HA)3
hdhfr
NdeI
NdeI
7.64 kb
(B)
kDa
1 2
37
25
50
BCKDH
(C)
kb
1 2 3 4 5
1.0
0.5
(D)
1 2 3 4
kb
8.0
5.0
3.0
1.5
0.5
?
?
pfnita
?
pCC4-pfnita
Probe: pfnita specific
(E)
1 2 3 4
kb
8.0
5.0
3.0
1.5
0.5
pCHD-pfnita_opt-HA
?
?
Probe hdhfr specific

Figure 4.12: Knock-out in the presence of an episomal copy of *pfnita*. (A) Knock-out of *pfnita* by gene replacement. A schematic diagram (not to scale) of the *pfnita* wild-type locus (I), the pCC4-*pfnita* plasmid (II), the *pfnita* gene locus following a double cross-over recombination event (III), and the pCHD-*pfnita_opt*-HA plasmid (IV). Diagnostic digests to analyse the genotype of transfected parasites were performed with *Nde*I. Restriction sites are indicated by black arrows. The lengths of the expected DNA fragments detectable with *pfnita* specific probe homologous to the 5' insertion of the plasmid are indicated by blue arrows, and those homologous to the drug selectable marker of the synthetic gene plasmid, hDHFR, are indicated in green. (B) A western blot using an anti-HA antibody directed against the 3' HA tag. The soluble (lane 1) and insoluble pellet (lane 2) are shown. The blot was stripped and re-probed with an anti-BCKDH antibody as a loading control. 20 µg total protein was loaded in each lane, the arrow indicates the position of PfNitA_opt-HA. (C) A PCR amplification was carried out to check for the loss of the cytosine deaminase/uracil phosphoribosyl transferase chimera at cycles 0 (lane 1) 1 (lane 2) and 2 (lane 3) and after treatment with 5FC at cycles WR 99210 cycles 0 (lane 4) and 1 (lane 5). (D) Southern blot using a gene specific probe after restriction of gDNA with *Nde*I, of cell lines co-transfected with pCC4-*pfnita* and pCHD-*pfnita_opt*-HA. The blue arrows in (A) indicate the expected sizes with a gene specific probe and the green arrow the *hdhfr* specific probe. Lane 1 contains D10 WT genomic DNA, lanes 2 and 3 contained gDNA from D10 *pfnita_opt*-HA/p^CC4-*pfnita* in WR 99210 selection cycle 0, lane 3 was after 5-FC treatment, and lane 4 contained gDNA in WR 99210 selection cycle 1 after 5-FC treatment. (E) Is as (D) using a *hdhfr* specific probe.

The *hdhfr* probe should detect only the plasmid containing the synthetic gene and produce a single 7.6 kb DNA fragment on the Southern blot. The 7.6 kb fragment was indeed present in the blot, as well as two lower molecular weight fragment bands of around 3 and 4.5 kb were also detected (lane 4) corresponding to parasites in WR 99210 selection cycle 1 after treatment with 5-FC. The two lower molecular weight fragments may suggest recombination between the KO and complementing plasmids.
To extend these data, probes were generated to the *bsd* and *cd* genes contained in the KO plasmid. The *bsd* DNA probe detects a 7.3 kb fragment in the KO plasmid and a 3 kb fragment after an integration event. The *cd* gene is only present on the plasmid and is lost during double cross-over recombination. Figure 4.13 displays Southern blots carried out using *bsd* (A) and *cd* (B) probes. The *bsd* probe hybridises to a band diagnostic to the presence of the pCC4 knock-out plasmid in each of the D10 *pfnita_opt*-HA/pCC4-*pfnita* cell lines, and as expected no signals were detected in the WT sample. In addition to the 7.3 kb DNA fragment, a low molecular weight DNA fragment of around 500 bp was observed. The size expected by the *cd* probe is 7.3 kb present on the KO plasmid. The *cd* probe detected a very faint band in each of the lanes containing the transformed parasites. The PCR analysis presented in Figure 4.12, (C) lane 1 confirms

the presence of the *cd* gene in the non- 5-FC treated cycle 0. Thus an obvious band should have been present in the Southern blot (Figure 4.13, (A) lane 2). In the event of a single cross-over either at the 5' or 3' insert only (i.e. a single cross-over), the *cd* gene should remain. If a 3' only integration had occurred, the *bsd* and *cd* probes should hybridise to a 7.3 kb DNA fragment, the equivalent to the plasmid. The *pfnita* gene specific probe would detect two fragments of similar molecular weight, 2.5 and 2.6 kb. However, a band or bands corresponding to these sizes was not observed in Figure 4.12, (D) ruling out a 3' only integration. On the other hand, a 5' only integration may also have taken place. The fragment sizes detected by the *bsd* and *cd* probes should be 3 and 8.8 kb, respectively. There was no 3 kb fragment detected by the *bsd* probe in Figure 4.13, (A), and the fragment in Figure 4.13, (B) is too small to correspond to the 8.8 kb fragment detected by the *cd* probe.

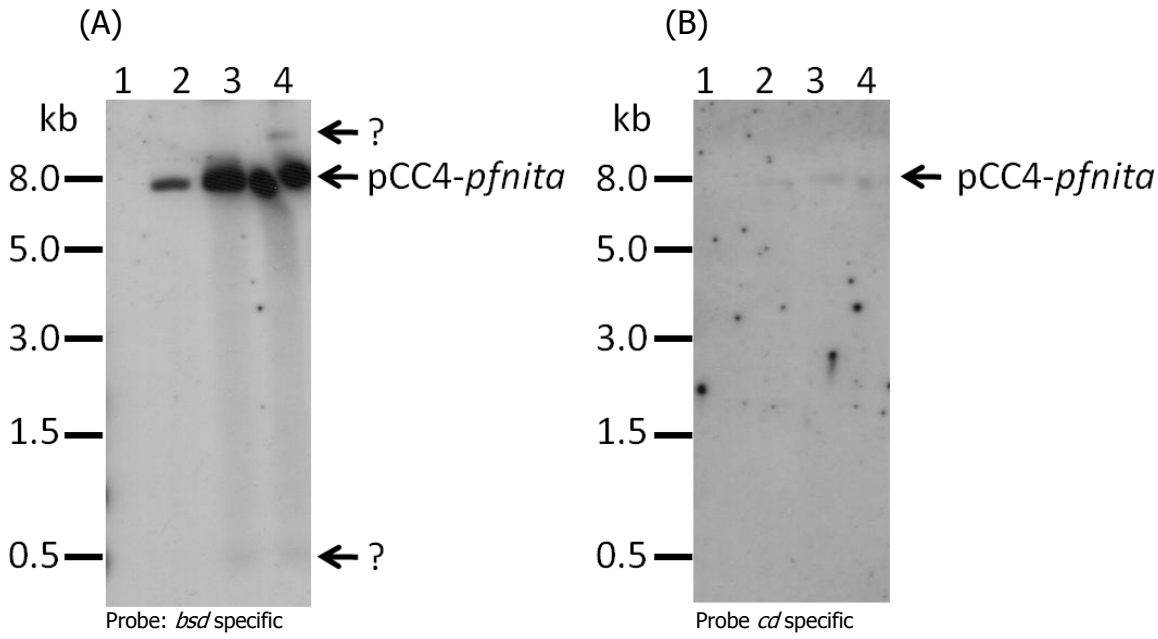


Figure 4.13: Southern analysis of D10 *pfnita_opt*-HA/pCC4-*pfnita* using *bsd* and *cd* probes. Southern blots after digestion of gDNA with *Nde*I, from cell lines transfected with pCC4-*pfnita* and pCHD-*pfnita_opt*-HA, probed with (A) *bsd* and (B) *cd* probes. The *bsd* probe detects a 7.3 kb fragment for the plasmid and a 3kb fragment after integration, the *cd* probe detects a 7.3 kb fragment on the plasmid only. Lane 1 contains D10 WT genomic DNA, lanes 2 and 3 contain gDNA from D10 *pfnita_opt*-HA/pCC4-*pfnita* in WR 99210 selection cycle 0, lane 3 is after 5-FC treatment, and lane 4 contains gDNA in WR 99210 selection cycle 1 after 5-FC treatment.

As the additional bands cannot be explained by plasmid recombination with the parasite genome it is possible the two plasmids have recombined leading to the resistance of parasites to both Bla and 5-FC without integration of the pCC4-*pfnita* KO plasmid. This is somewhat speculative and in order for this idea to be verified the parasite lines would have to be analysed in greater detail in further experiments. Thus, the attempt to knock-out PfNitA whilst in the presence of a supplementary copy of the gene was not successful. This failure indicates that the gene may be essential to asexual stage *P. falciparum* parasites. The information in Table 4.4 provides an overview of the results from this section.

Parasite	PCR	Southern Blot			
	cd	*pfnita*	*hdhfr*	*bla*	*cd*
Cycle 0	positive	WT, KO plasmid	positive	positive	positive
Cycle 0+ 5-FC	negative	WT, KO plasmid	positive	positive	positive
Cycle 1	positive	WT, KO plasmid	positive	positive	positive
Cycle 1+ 5-FC	negative	WT, KO plasmid	positive	positive	positive
Cycle 2	positive	WT, KO plasmid	positive	positive	positive

Table 4.4: Overview of Southern blot and PCR results from cell line D10 *pfnita_opt*-HA/pCC4-*pfnita* after drug cycling. PCR reactions are either positive or negative for the amplification product from *cd*. Southern blots using the *hdhfr* probe are either negative or positive for the presence of *hdhfr* on the pCHD-*pfnita_opt*-HA plasmid and blots using the *bla* or *cd* probes are either positive or negative for the corresponding genes from the pCC4-*pfnita* plasmid.

4.5 Tagging the Native Gene by Allelic Exchange

In order to assess the susceptibility of the *pfnita* locus to recombination following the failure to obtain null mutants using double homologous crossover recombination a tagged functional copy of the C-terminal *pfnita* region was created using the MultiSite Gateway® system. To attempt to reconstitute the gene function of *pfnita*, a single homologous crossover reaction should yield a functional, tagged copy of the *pfnita* gene, while demonstrating that the region can be targeted by exogenous DNA through the integration of the plasmid backbone. To target the locus of *pfnita* an 859 bp C-terminal region of the gene (710-1569 bp) was amplified from 3D7 gDNA using the primers PfNitACtml GW F and PfNitA GW R (Appendix 2). The forward primer contains a CACC overhang for directional cloning into the TOPO ENTR/D vector. The C-terminal *pfnita* (*ctml*) gene fragment was cloned in frame with 3xHaemagglutinin ($(ha)_3$) or green fluorescent protein (*gfp*) genes using the MultiSite Gateway® technology (Figure

4.14). The final plasmid contains the C-terminal *pfnita* gene fragment controlled by the Hsp86 promoter. Since only the C-terminus is present on the plasmid, no expression should occur from this construct.

Both constructs pCHD-Hsp86-*ctml*-GFP and pCHD-Hsp86-*ctml*-$(HA)_3$ were transfected on two independent occasions into *P. falciparum* D10 strains. Parasites containing the plasmid were selected by applying WR 99210 since the plasmid contains the *hdhfr* resistance gene. Parasites were observed four to five weeks after transfection. Transfected lines were termed D10$^{ctml\text{-}GFP}$-1, D10$^{ctml\text{-}GFP}$-2, D10$^{ctml\text{-}HA}$-1 and D10$^{ctml\text{-}HA}$-2. To obtain expression of the full-length gene with a GFP or $(HA)_3$ tag, a single cross-over must take place. The single cross-over forms one full-length copy of the gene with the GFP or $(HA)_3$ tag and a non-functional truncated copy further downstream (A of 4.17 and 4.18, respectively). After a single cross-over the endogenous gene would therefore be tagged with GFP or $(HA)_3$.

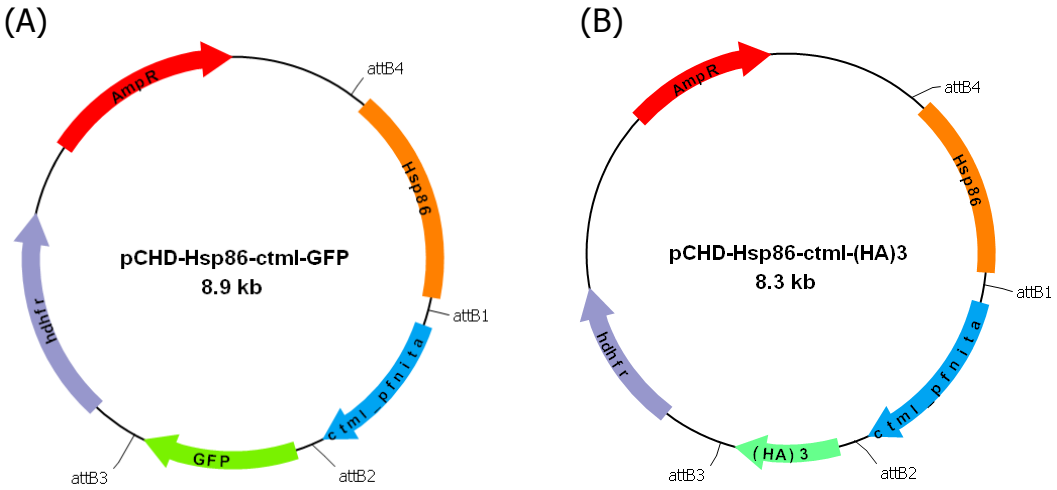


Figure 4.14: C-terminal *pfnita* integration transfection plasmids. The final plasmids (A) pCHD-Hsp86-*ctml*-GFP and (B) pCHD-Hsp86-*ctml*-$(HA)_3$ encode the codon-optimised *pfnita* gene and a GFP or $(HA)_3$ tag, respectively. These plasmids were transfected into 3D7 and D10 parasites. The attB sites that bring the constructs together during the Gateway cloning LR reaction are indicated. The ampicillin and hDHFR selectable markers used in bacterial cloning and parasite transfection/selection, respectively, are also shown.

gDNA was isolated from WR 99210 resistant parasites, analysed by diagnostic Southern blot and diagnostic PCR for integration events. Parasites from both GFP and $(HA)_3$ cell lines were taken through two on/off drug cycles, (see Section 2.5.7.1 and

Figure 2.6 for an explanation). The data shown in Figure 4.15 provides an overview of the drug cycling process for GFP tagged parasites and Figure 4.17 for $(HA)_3$ tagged parasites. The GFP and $(HA)_3$ tagged cell lines will be described separately.

4.5.1 GFP tagged PfNitA

After each drug selection cycle for D10$^{ctml\text{-}GFP}$-1, gDNA was isolated from parasites for a diagnostic Southern blot. The expected fragment sizes using a *pfnita* gene specific and *hdhfr* probes is presented in Figure 4.16, (A). gDNA was digested with *Nde*I and *Cla*I and probed with the *pfnita* gene specific probe and *hdhfr* specific probe against the DSM of the plasmid. After integration of the plasmid by single cross-over, the DSM will remain. Parasites transfected with D10$^{ctml\text{-}GFP}$-1 appeared to have lost the plasmid after two off/on drug cycles, as demonstrated by lack of the diagnostic fragment of 4.55 kb in the Southern blot using the gene specific probe and the absence of the *hdhfr* fragment at 3.65 kb using the *hdhfr* probe (Figure 4.16 (B) lane 3). The Southern blot also demonstrated that no integration of pCHD-Hsp86-*ctml*-GFP had occurred.

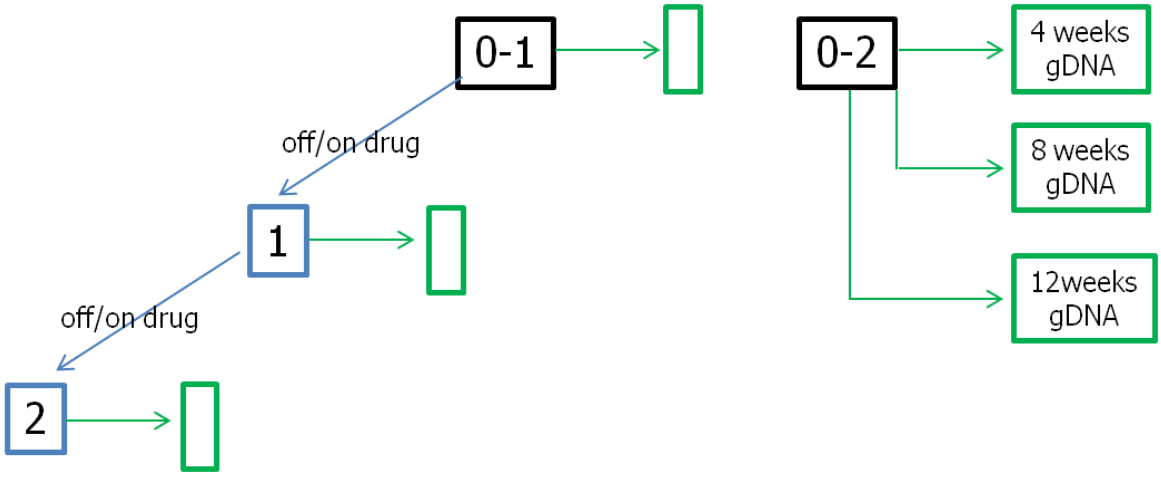


Figure 4.15: Overview of the drug cycling process for parasites transfected with pCHD-Hsp86-*ctml*-GFP. Parasites after transfection, growing on WR 99210 were termed cycle 0-1 and cycle 0-2 (black boxes). Drug cycling of transfected parasites was carried out to select for integration of plasmids into the parasites genome. Off/on drug WR 99210 cycling is indicated by blue boxes/arrows. At each stage gDNA was taken for analysis using PCR and/or Southern blot, and stabilates were made for the preservation of the parasite line. This is indicated by a green box/arrow combination.

DNA fragment sizes of 1.3 and 6.2 kb were expected using the gene specific probe if construct integration had taken place. The *hdhfr* probe should detect a fragment of

3.65 kb which is identical to the plasmid fragment after integration. The results presented in Figure 4.16 (B) lanes 2 and 3 shows that the endogenous gene has not been affected by drug cycling as it is present in cycle 0 and cycle 2.

Consequently, a second parasite line, D10$^{ctml\text{-}GFP}$-2, was thawed since it had not been taken through any drug selection cycles (cycle 0). This parasite line was maintained in culture in the presence of selectable drug pressure and no drug cycling was carried out.

At four week intervals gDNA was isolated and analysed by a diagnostic PCR using three forward primers and a single reverse primer from the *gfp* gene (Table 4.5 and Appendix 2) to test for the presence of an integration event. The three forward primers were made from different areas of the 5' *pfnita* gene to maximise the chances of obtaining a product from the PCR as no positive control for integration was available. The oligonucleotide primer locations are indicated by red arrows in Figure 4.16 (A). The expected sizes of the PCR products using the integration primers are as given in Table 4.5. PCR amplification was also carried out to detect the presence of an approximately 500 bp *hdhfr* fragment using primers hDHFR F and hDHFR R (Appendix 2). The results from the integration PCR are included in Figure 4.16 (C).

Primer Name		Product size
PfNitA KO 750 5 F	A	2445 bp
PfNitA KO 500 5 F	B	1787 bp
PfNitA F1	C	1314 bp
GFP mut2 R	R	n/a

Table 4.5: Primers used to check for integration of pCHD-Hsp86-*ctml*-GFP. The primer name denotes the designation of the primer as in Appendix 2. The single letter after the primer corresponds to Figure 4.16. The product size is the size of the product amplified if integration has taken place when PCR was carried out using each of the forwards primers in combination of the single reverse primer.

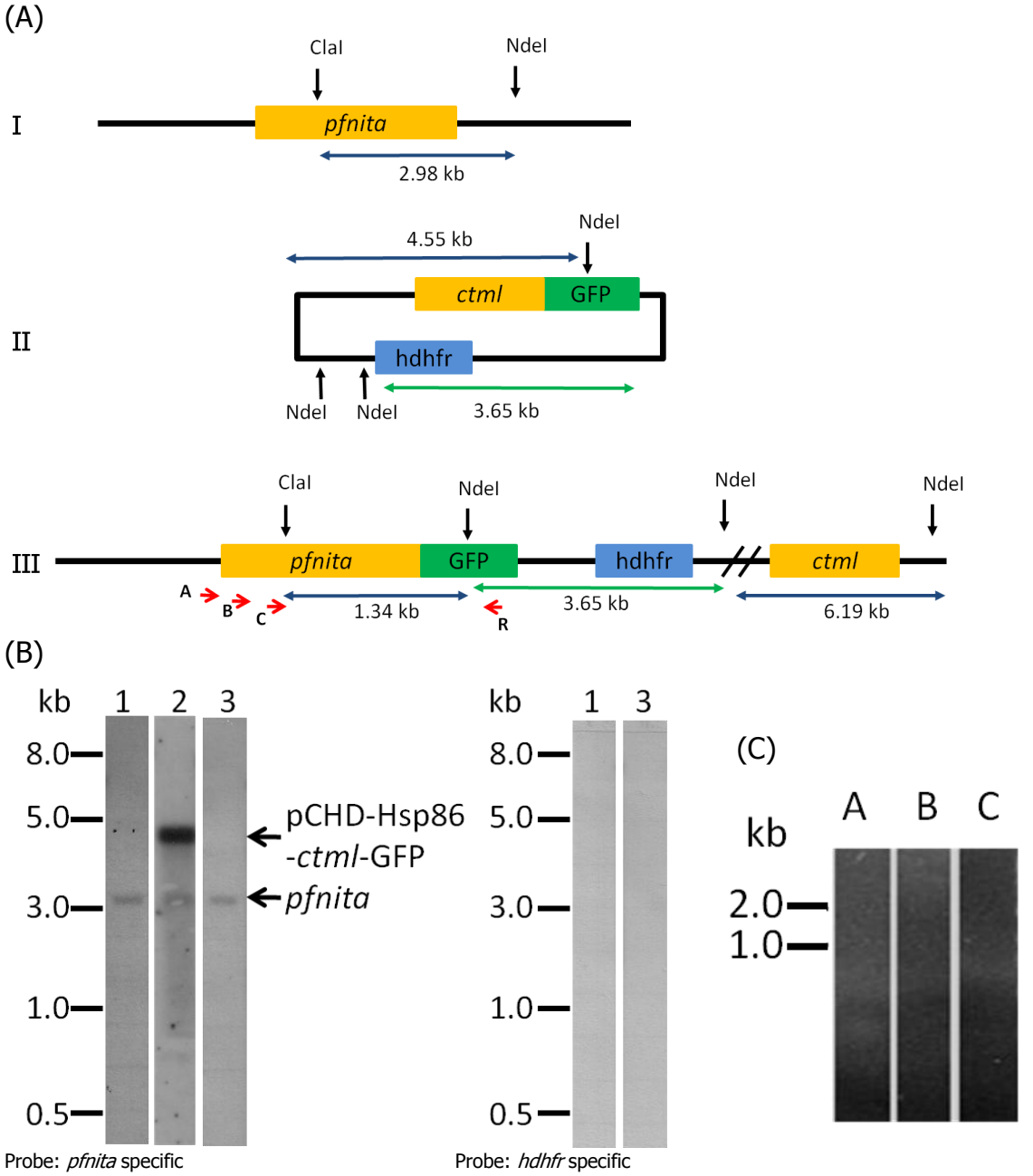


Figure 4.16: Tagging the native gene with GFP using allelic exchange. (A) A schematic diagram (not to scale) of the *pfnita* wild-type locus (I), the pCHD-Hsp86-*ctml*-GFP plasmid (II) and the *pfnita* gene locus following a single cross-over recombination event (III). Diagnostic digests to analyse the genotype of transfected parasites were performed with *Nde*I and *Cla*I. Restriction sites are indicated by black arrows. The lengths of the expected fragments detectable with *pfnita* specific probe homologous to the 5' insertion of the plasmid are indicated by blue arrows, and those homologous to the drug selectable marker hDHFR are shown in green. Primers used to detect integration are shown by small red arrows. (B) Southern blots of cell lines transfected with pCHD-Hsp86- *ctml*-GFP after digestion of gDNA with *Nde*I and *Cla*I, and probed with a gene specific probe and *hdhfr* probe. The blue arrows in (A) indicate the expected sizes with a gene specific probe and the green arrows the hDHFR specific probe. Lane 1 contains D10 WT genomic DNA, Lanes 2 and 3 contain gDNA from D10$^{pCHD\text{-}Hsp86\text{-}ctml\text{-}GFP}$ in WR 99210

selection cycles 0 and 2, respectively. (C) A PCR amplification using primers to detect integration was carried out after 12 weeks growth. Forward primer A is specific to the downstream non-coding region and primers B and C are specific to the 5' *pfnita* sequence, and the reverse primer is found in the GFP sequence. The expected sizes were as follows: A + R 2445 bp, B + R 1787 bp and C + R 1314 bp.

None of the primer sets amplified a product from D10$^{ctmI\text{-}GFP}$-2 gDNA after 12 weeks of continuous growth in culture subject to WR 99210 selection pressure. The hDHFR primers amplified an approximately 500 bp product indicating the *hdhfr* gene was still present (data not shown). No Southern analysis was carried out on this cell line. When I left the laboratory no integration event was detectable for this plasmid after 12 weeks.

4.5.2 $(HA)_3$ tagged PfNitA

An overview of the drug cycling process for strain D10$^{ctmI\text{-}HA}$-1 is presented in Figure 4.17. After each off/on drug selection cycle for D10$^{ctmI\text{-}HA}$-1, gDNA was isolated from parasites for a diagnostic Southern blot. The expected fragment sizes using gene specific and *hdhfr* probes is presented in Figure 4.18, (A). gDNA was digested with *Nde*I and *Cla*I and probed with a gene specific probe and *hdhfr* specific probe against the DSM of the plasmid. After integration of the plasmid by single cross-over recombination, the DSM should remain in the parasite genome. In the Southern blot results of D10$^{ctmI\text{-}HA}$-1 a dark smear was repeatedly observed at cycles 0, 1 and 2, suggesting potential non-specific degradation of gDNA. The plasmid construct (7.5 kb) and endogenous gene (3 kb) were also detected in each of the cycles (Figure 4.18 (B) lane 2). After cycle 2 parasites were maintained in culture under drug pressure. gDNA was generated from these parasites and the possibility of integration was investigated using a diagnostic PCR amplification using the same forwards primers as described above (Section 4.5.1, Table 4.5 and Table 4.6) and a reverse primer (R2) found between the *$(ha)_3$* sequence and the *hdhfr* gene (GW int R1, Table 4.6 and Appendix 2). The expected product sizes are outlined in Table 4.6. Each of the primer sets returned a positive result implying integration had taken place in at least a subset of the parasite population and the results are presented in Figure 4.18 (C). A second Southern blot was carried out and the results are presented in Figure 4.18 (B) lane 3.

Primer Name		Product size
PfNitA KO 750 5 F	A	2081 bp
PfNitA KO 500 5 F	B	1743 bp
PfNitA F1	C	1270 bp
GW int R1	R2	n/a

Table 4.6: Primers used to check for integration of pCHD-Hsp86-*ctmI*-(HA)$_3$. The primer name denotes the name of the primer from Appendix 2. The single letter after the primer corresponds to Figure 4.16. The product size is the size of the product amplified if integration has taken place when the PCR reaction was carried out using each of the forwards primers in combination of the single reverse primer.

The results of the Southern blot using the gene specific probe showed that the endogenous gene (3 kb) was still present as was the plasmid pCHD-Hsp86-*ctmI*-(HA)$_3$ of 7.5 kb. The dark smear was still present but additional bands at approximately 10 kb, 6 kb, 5 kb and 4 kb were distinguishable (Figure 4.18 (B) lane 3). Following an integration event hybridising DNA fragments would be expected at 4.3 kb and 6.2 kb, possibly corresponding to some of the detected fragments. When the blot was reprobed with a *hdhfr* specific probe a fragment was observed at 7.5 kb which is consistent with the plasmid, however, additional fragments were visible at 1.75 kb and 4 kb.

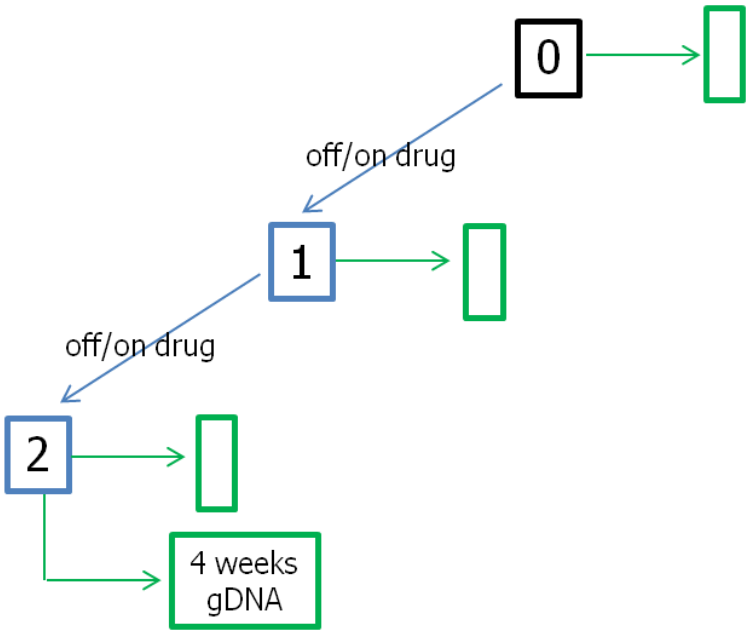
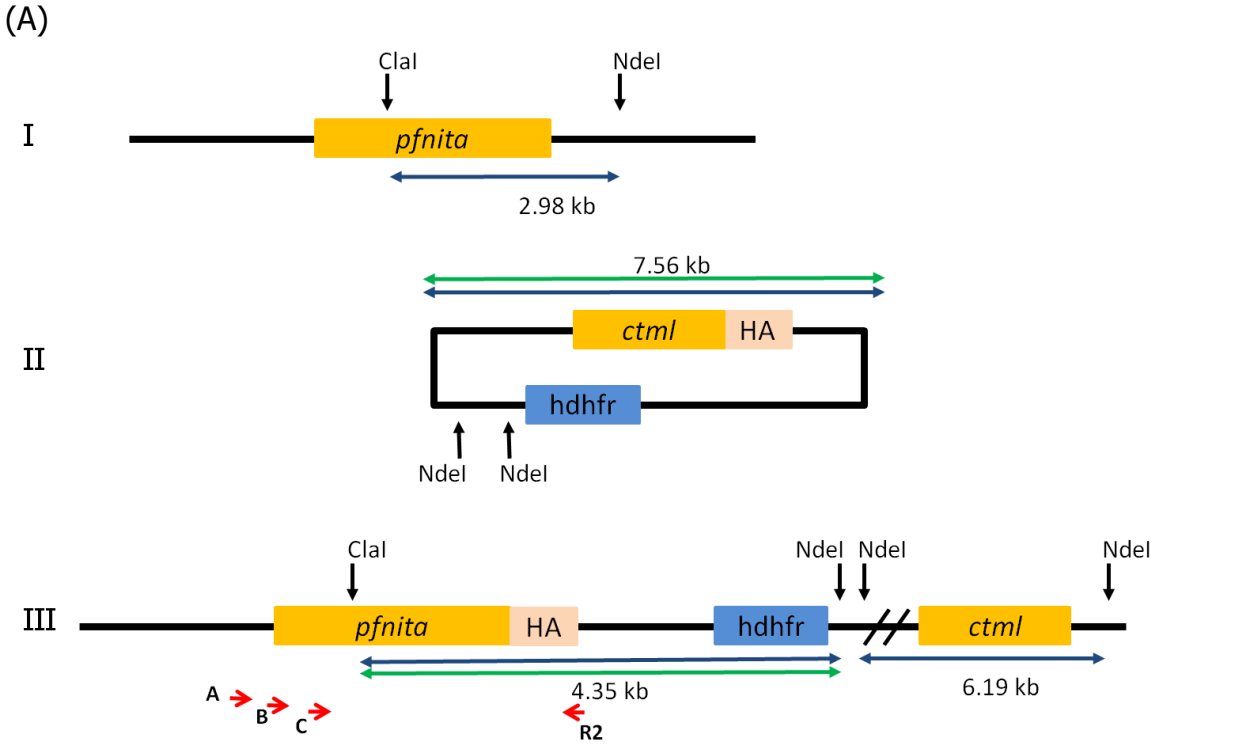


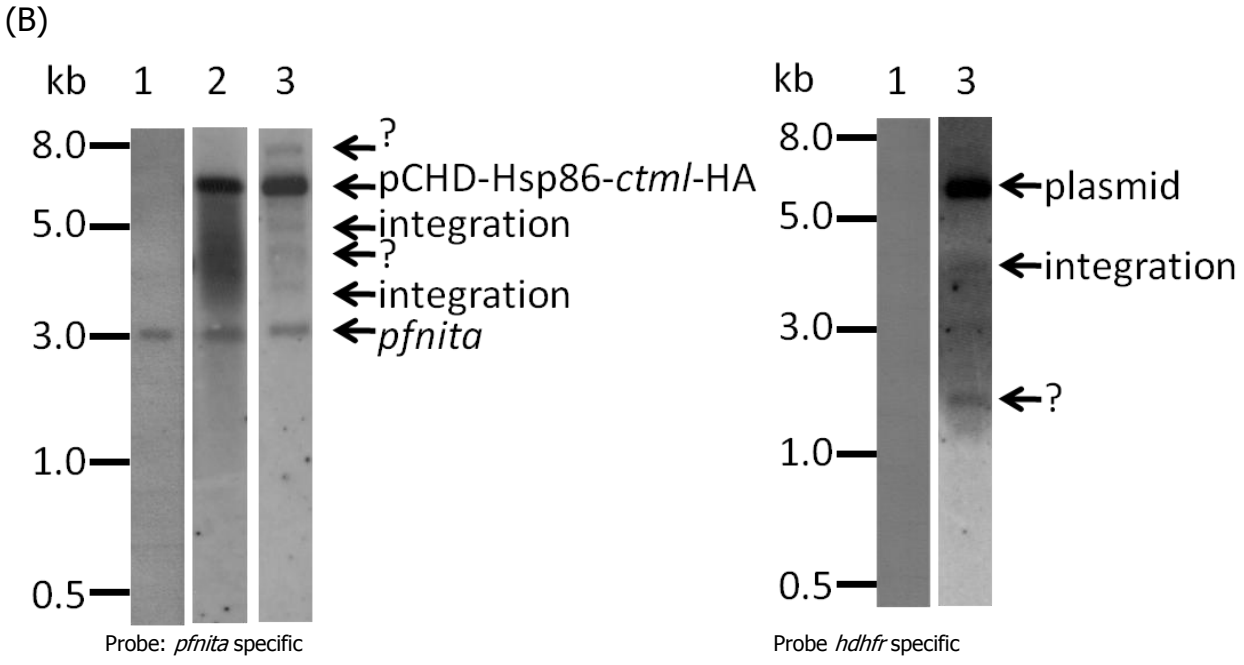
Figure 4.17: Overview of the drug cycling process for parasites transfected with pCHD-Hsp86-*ctmI*-(HA)$_3$. Parasites after transfection, growing on WR 99210 were termed cycle 0 (black box). Drug cycling of transfected parasites was carried out in order to select for integration of plasmid constructs into the parasite genome. Off/on drug WR 99210 cycling is indicated by blue boxes/arrows. At each stage gDNA was taken for analysis using PCR and/or Southern blot, and stabilates were made for the preservation of the parasite line. This is indicated by a green box/arrow combination.

After an integration event a DNA fragment would be expected at 4.3 kb when using the *hdhfr* probe. The fact that the additional bands observed with both the gene specific and *hdhfr* probes were faint indicated only a small population of parasites contained the integrated plasmid at the correct locus, and that integration elsewhere in the genome may also have taken place. Further to this, a western blot of the insoluble and soluble protein from D10$^{ctmI\text{-HA}}$-1 was carried out using an anti-HA antibody (as described previously). Insoluble protein from D10$^{pfnita_opt\text{-HA}}$ was used as a positive control for size reference.

The results presented in Figure 4.18, (D), lane 1 shows the presence of the insoluble protein and no protein cross-reacted with the antibody. However, three distinct bands were observed in the insoluble fraction (lane 2), one of which corresponded to the positive control (lane 3) at approximately 35 kDa. This result implied a full-length tagged PfNitA protein was expressed by the parasites. Moreover, the other band on the blot is perhaps related to non-specific integration of the plasmid and subsequent protein expression. The blot was stripped and re-probed with an anti-BCKDH antibody to verify separation of the soluble and insoluble fractions.

Both protein fractions from D10$^{ctmI\text{-HA}}$-1 parasites were positive for BCKDH at different intensities (Figure 4.18, (D)). Visualisation of a band on the western blot may be due to the large concentration of protein applied to the SDS-PAGE gel, approximately 1 mg. If only a relatively small number of parasites contained the correctly integrated pCHD-Hsp86-*ctmI*-$(HA)_3$ plasmid then by extension, only a small number of parasites would contain the full-length $(HA)_3$ tagged PfNitA protein, thus the volumes of insoluble and soluble proteins that were electrophoresed was titrated between 20 µg and 2.5 mg. At 1.2 mg, the anti-HA antibody gave the banding pattern present in lane 3 of Figure 4.18, (D). At concentrations lower than 1.2 mg no bands were detected using the anti-HA antibody, though bands were detected in all soluble protein lanes when probed with anti-BCKDH as well as insoluble protein in lanes with 160 µg plus of total protein.


(A)
I
ClaI
NdeI
pfnita
2.98 kb
II
7.56 kb
ctml
HA
hdhfr
NdeI
NdeI
III
ClaI
NdeI
NdeI
NdeI
pfnita
HA
hdhfr
ctml
4.35 kb
6.19 kb
A
B
C
R2


(B)
kb
1
2
3
8.0
5.0
3.0
1.0
0.5
?
pCHD-Hsp86-ctml-HA
integration
?
integration
pfnita
Probe: pfnita specific
kb
1
3
8.0
5.0
3.0
1.0
0.5
plasmid
integration
?
Probe hdhfr specific

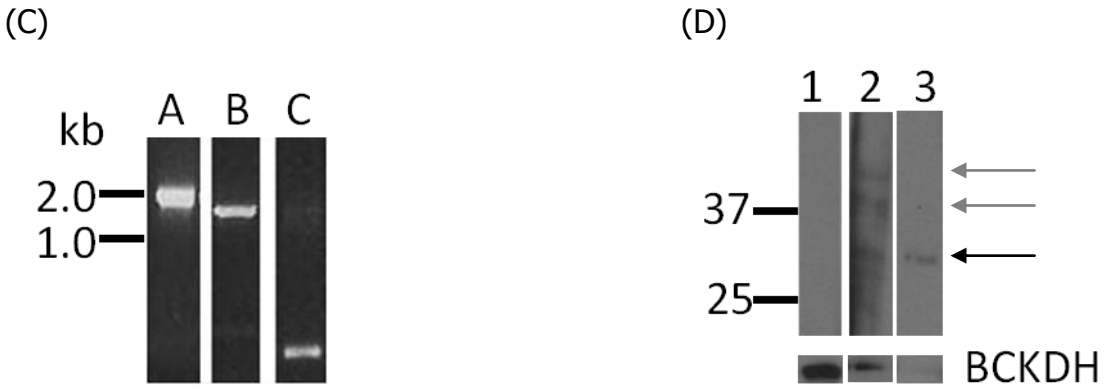


Figure 4.18: Tagging the native gene with $(HA)_3$ using allelic exchange. (A) A schematic diagram (not to scale) of the *pfnita* wild-type locus (I), the pCHD-Hsp86-*ctmI*-$(HA)_3$ plasmid (II) and the *pfnita* gene locus following a single cross-over recombination event (III). Diagnostic digests to analyse the genotype of transfected parasites were performed with *Nde*I and *Cla*I. Restriction sites are indicated by black arrows. The lengths of the expected fragments detectable with *pfnita* specific probe homologous to the 5' insertion of the plasmid are indicated by blue arrows, and those homologous to the drug selectable marker hDHFR are shown in green. Primers used to detect integration are shown by small red arrows. (B) Southern blots after digestion of gDNA with *Nde*I and *Cla*I, of cell lines transfected with pCHD-Hsp86-*ctmI*-$(HA)_3$, and probed with a gene specific probe and *hdhfr* probe. The blue arrows in (A) indicate the expected sizes with a gene specific probe and the green arrows the hDHFR specific probe. Lane 1 contains D10 WT genomic DNA, Lanes 2 and 3 contain gDNA from $D10^{ctmI\text{-}HA}$ in WR 99210 selection cycles 0 and 2, respectively. (C) A PCR using primers to detect integration was carried out. Forward primer A is specific to the downstream non-coding region and primers B and C are specific to the 5' *pfnita* sequence, and the reverse primer, R, is found in the plasmid only. The expected sizes are as follows: A 2081 bp, B 1743 bp and C 1270 bp. (D) A western blot was carried out against insoluble and soluble parasite extract using an anti-HA antibody directed against the C-terminal tag. Lane 1 contains the soluble parasite extract and lane 2 the insoluble pellet. Lane 3 provides a control sample for size (the synthetic PfNitA insoluble pellet). The blot was stripped and re-probed with an anti-BCKDH antibody as a loading control which examined whether or not the fractions had been separated efficiently. The black arrow indicates the position of PfNitA-HA and the grey arrows the additional bands.

Cells of $D10^{ctmI\text{-}HA}$-1 should therefore be cloned by limiting dilution to select for a clonal cell line which has undergone the integration event at the intended locus. However, due to technical reasons this work is yet to be completed.

4.6 Knock-out by Gene Disruption and Knock-out Control by Allelic Exchange

An alternative method to achieve loss of function of a gene product is to generate a gene disruption by single homologous recombination into the gene locus of interest.

pHH1 based plasmid constructs are employed successfully in knock-out and knock-out control studies to prove that the gene locus is targetable without the addition of any tags as described in Section 4.5. The approach is depicted in Figure 4.19.

The pHH1 plasmid contains the *hdhfr* gene for selection of transfected parasites with WR 99210, and an ampicillin resistance cassette for selection during bacterial cloning (Reed *et al*, 2000) (Figure 4.19, A and B I). The gene of interest (GOI)fragments are cloned between *Bgl*II and *Xho*I restriction sites, and these must be incorporated into the primers.

The GOI is truncated at the N-terminus and is missing the ATG start codon. The GOI is also truncated at the C-terminus and contains an artificial stop codon (A) II. For knock-out control constructs the GOI fragment to be cloned into the vector is truncated at the N-terminus as in the knock-out, but contains the full C-terminal sequence (B) II. A single cross-over recombination event between a knock-out construct and the GOI leads to two truncated and non-functional copies of the gene (A) III. Recombination between a knock-out control construct and the GOI leads to the formation of one full-length copy and one truncated copy of the gene (B) III. The knock-out control does not contain any N-terminal or C-terminal tag that may disrupt the function of the protein making this a viable option as previous attempts altered the C-terminal of the protein with a tag.

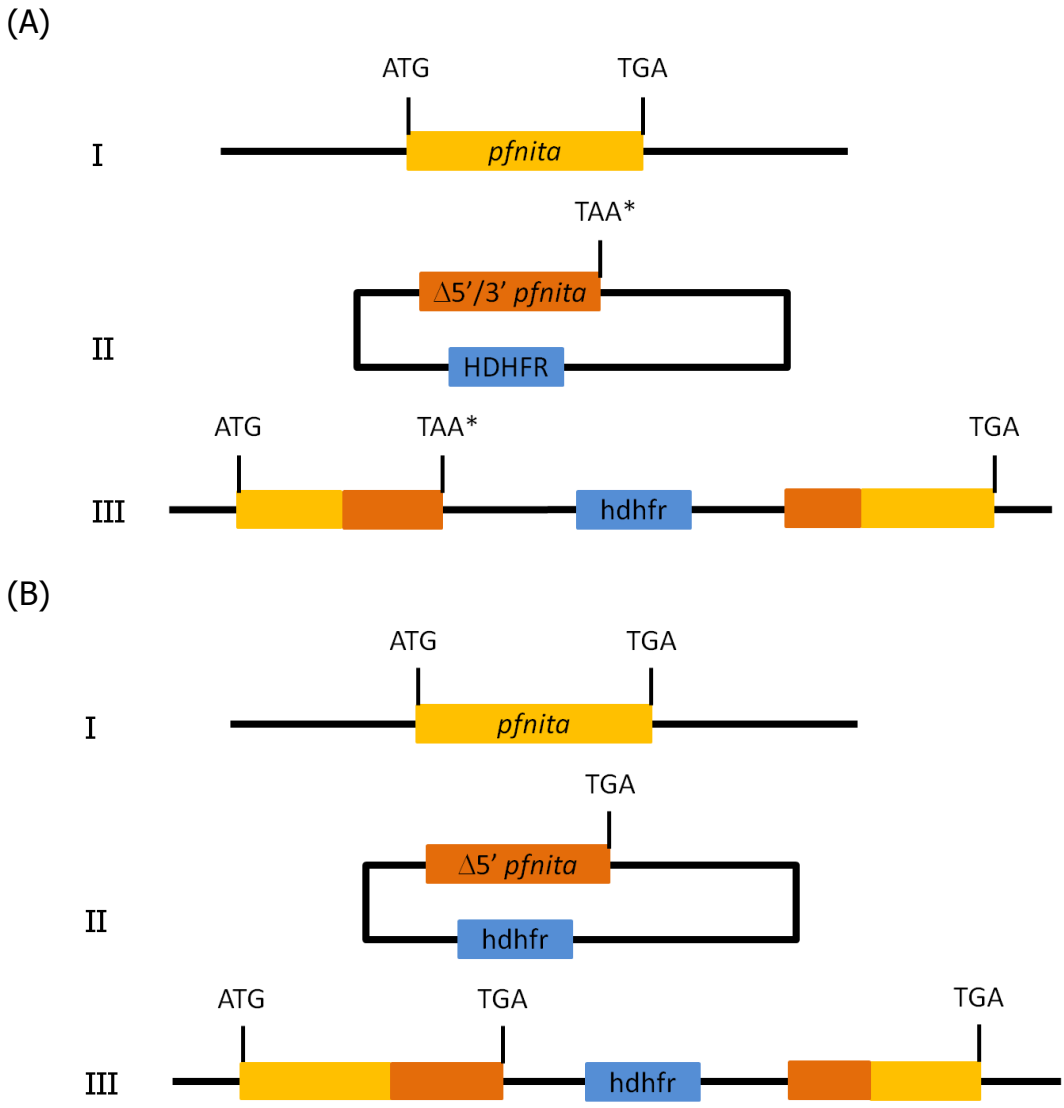


Figure 4.19: Future Strategy; knock-out and knock-out control by allelic exchange. (A) A schematic diagram of the *pfnita* WT locus (I). The knock-out fragment is truncated at the 5' and 3' of the gene, leaving approximately 1000 bp available for recombination (II). A single cross-over at the correct locus should result in an interruption of the locus producing two non-functional copies of the gene (III). (B) A control integration event. The knock-out control fragment is, as with the knock-out construct truncated at the 5', but is in contrast full-length at the 3' retaining the endogenous stop codon (II). Integration of the knock-out control construct, by single cross-over, into the gene locus should thus result in the disruption of the locus generating a single functional copy of the gene and a truncated copy (III). Diagrams are not to scale.

Therefore, as a knock-out control, a 1375 bp insert homologous to the *pfnita* C-terminal region and containing the endogenous stop codon was amplified using the primers PfNitA pHH1 KOcon F and PfNitA pHH1 KOcon R (Appendix 2) from 3D7 gDNA

and cloned into the intermediate cloning vector pSCB (StrataCloneBlunt, Agilent Technologies, Cheshire, UK) for sequencing. A knock-out construct was also prepared truncated at the 5' and 3' ends containing a TAA stop codon. The construct was amplified using the primers PfNitA pHH1 KO F and PfNitA pHH1 KO R (Appendix 2) from 3D7 gDNA, and cloned into pSCB for sequencing. The sequences of the two products were confirmed. However, due to time constraints, the KO and KO control inserts have not yet been cloned into the final pHH1 vectors and sequenced, and the transfections with these new plasmids await further experimental work.

4.7 Summary

- Subcellular prediction programs do not predict a localisation for *Plasmodium falciparum* NitA, or any *Plasmodium* orthologues. Therefore the synthetic codon-optimised *pfnita* gene was cloned into an expression plasmid containing sequences encoding GFP or $(HA)_3$ tag to allow biochemical analysis. The synthetic *pfnita* gene was expressed episomally, and the fusion protein expression was confirmed by western blot exploiting antibodies made to the C-terminal epitope tag. Using fluorescence light microscopy, the GFP or $(HA)_3$ tagged PfNitA protein was located to the parasite plasma membrane but was also observed in the digestive vacuole membrane.

- Co-localisation studies of PfNitA were carried out using immunofluorescence analysis comparison with two known transporters. One of these transporters is known to be located in the parasite plasma membrane (hexose transporter) whilst the second in the digestive vacuole membrane (chloroquine resistance transporter). In IFA's co-localisation of PfNitA was observed exclusively with the hexose transporter in the plasma membrane. However, the location in the digestive vacuole membrane localisation was not confirmed.

- Attempts to knock-out the *pfnita* gene were unsuccessful using a double cross-over recombination approach. Out of 20 individual transfections into two different parasite strains only one parasite line was generated in which the KO plasmid was maintained episomally. This suggests that *pfnita* is either essential for parasite survival or that the locus is unable to be genetically targeted.

- A synthetic codon-optimised *pfnita* expression plasmid (pCHD-*pfnita_opt*-HA) was co-transfected with a pCC4-*pfnita* knock-out vector to complement for the potential loss of the *pfnita* gene. Both plasmids were maintained episomally as shown by Southern blot. Western blot analysis confirmed expression of the complementing PfNitA-$(HA)_3$ protein. Treatment with 5-FC caused loss of the negative selectable marker and potentially led to the recombination between the two constructs as suggested by Southern blot analysis.

- A C-terminal portion of the *pfnita* gene was cloned into a plasmid harbouring a GFP or $(HA)_3$ tag. The purpose was to tag the endogenous gene and to express the tagged gene under control of its native promoter to show whether the gene locus is targetable, *per se*. Recombination was not achieved using the GFP tagged gene construct. In contrast, the $(HA)_3$ tagged gene construct, on the other hand, underwent recombination at the correct gene locus as confirmed by diagnostic Southern blot and PCR. The cloning of this parasite line has not yet been achieved due to technical difficulties.

5 *Plasmodium chabaudi* Cir2 Expression and Structural Analyses

5.1 Introduction

The PcCir2 protein is one of a large family of proteins encoded by rodent, simian and human malarias. The proteins of the family have been postulated to play a role in antigenic variation due to the presence of a hypervariable region (HVR) (Figure 1.6), but no evidence of such a role has been determined. The objectives of this chapter were (i) to overexpress PcCir2 in a bacterial expression system, (ii) to produce purified protein for crystallography trials, (iii) develop a 2D structural model, and, (iv) determine the 3D crystal structure of the protein. It was hoped that with structural models, progress could be made in the determination of a role.

5.1.1 Membrane protein expression

Membrane proteins (MP's) are particularly difficult to work with and special considerations need to be taken when it comes to overexpression and purification. However, the production of proteins, and in particular membrane proteins, is important to enable the study of membrane proteins as they comprise the largest class of drug targets (Arinaminpathy *et al*, 2009). The overexpression of membrane proteins can be toxic to the cells they are produced in. There is a finite surface area that can be utilised by membrane proteins, too high an expression may prevent the hosts' endogenous, and in some cases essential, membrane proteins from reaching their destination. Nonetheless, membrane proteins must be produced at high levels to reach a sufficient concentration to be purified for further study. In this regard, purified protein for crystallography is required in milligram amounts. The large hydrophobic domains contained within membrane proteins can result in instable proteins that have a tendency to aggregate when removed from the phospholipid bi-layer (Junge *et al*, 2008). Such aggregation may lead to the packaging of proteins into inclusion bodies (IBs). Protein can be extracted from IBs. To guarantee a protein has been correctly re-folded, an activity assay is normally required to test the proteins' activity is as expected.

During the purification of a membrane protein, a detergent is used to replace the lipid bi-layer and maintain the structural conformation of the protein. Choosing the correct detergent is fundamental for membrane protein stabilisation and for retaining the protein in its native state. The removal of a protein from its membrane may cause unfolding and other structural changes. Likewise, the detergent must be compatible with purification and related procedures such as crystallographic approaches (Newstead *et al*, 2007). Detergents are amphiphilic compounds with polar and nonpolar domains. Solubility is provided by the hydrophilic domain of a detergent. The hydrophobic domain allows access to the lipid bi-layer. Detergents are able to disrupt the hydrophobic associations that hold membrane proteins in place (Newstead *et al*, 2007). The small amphipathic molecules bind to the hydrophobic regions of the membrane proteins and dislodge the lipid molecules. The side chains of detergents come in a variety of lengths and can be ionic, anionic and zwitterionic. Using strong ionic detergents, such as SDS, causes the protein to become denatured and are therefore unsuitable to the purification of membrane proteins which need to be functionally active when purified. Milder anionic detergents such as DDM have been proven to be useful (e.g. Beckham *et al*, 2010). The "membrane proteins of known structure database" [30] contains a list of all membrane proteins crystallised and the detergents used.

E. coli was chosen as the expression host for the *pccir2* gene rather than more complicated and expensive systems such as yeast, insect or mammalian cells (Junge *et al*, 2008). *E. coli* has many advantages over other expression systems, despite the fact that it has drawbacks when used in the production of a eukaryotic protein, especially one from *Plasmodium* due to the high AT content of its genome. Yeasts such as *Pichia pastoris* (e.g. Beckham *et al*, 2009) are used in the expression of membrane proteins. Yeasts are eukaryotic organisms and therefore carry out a range of posttranslational modifications (PTM), though these are not always suited to the protein of interest (Junge *et al*, 2008). The cost of culturing yeasts is low, and the processes are reasonably easy to carry out. The culture and transfection of insect cells is more difficult, due to the complexity in creating protein expression vectors and more complex media. However, PTMs are possibly more similar to those in *Plasmodium* and

[30] http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html. Accessed on 16 January 2010.

the larger size of cells provides more membrane for the overexpressed proteins (Junge *et al*, 2008). Mammalian cells carry out PTMs typical of human cells, though potentially not *Plasmodium* cells. The culture of mammalian cells, such as human embryonic kidney cells and human epitheloid cells, is the most complicated of the eukaryotic cell types and transfection/expression can take many days to occur (Junge *et al*, 2008). For each of the expression hosts, the protein expression must be optimised to maximise the protein production with minimal cost to the host.

E. coli was chosen as the host for this project due to the low cost of culturing and relative ease of use. Most of the structural information known about membrane proteins has come from those expressed in bacterial expression systems, 85 % compared to 7.2 % (yeast, insect and mammalian cells combined) (Junge *et al*, 2008). There are many plasmid vectors for *E. coli* and each of these can be used to fine tune the production of a given membrane protein. These vectors can also add tags such as GFP to act as a reporter protein or a polyhistidine tag to aid in purification. The downside of using a prokaryote system to express a eukaryotic protein is the absence of post-translational modifications such as palmitoylation and proteolytic cleavage (Junge *et al*, 2008) which may prove to be essential for the correct folding and therefore protein function.

5.1.2 *P. chabaudi* as a model

P. falciparum, the causative agent of most human malarial deaths, can be propagated in the laboratory. However, as it infects only humans and a small number of monkeys, which raises ethical issues, no *in vivo* model is readily available. To study *Plasmodium in vivo* researchers turn to rodent malarias as these share a conservation of their genetic and biochemical processes with the human malarias (Waters, 2002). The whole of the rodent malaria life cycle can be studied *in vivo* under laboratory conditions, and the parasites are able to be genetically manipulated *in vitro* (Reece and Thompson, 2008). In this respect *P. chabaudi* is recognised as a useful model for the investigation of aspects such as drug resistance (Afonso *et al*, 2010), testing of drugs, and research on antigenic variation (Cunningham *et al*, 2005) as rodent models allow *in vivo* investigations of host-parasite relationships. Rodent malaria models are not without their disadvantages such as validity of extrapolating host-parasite interactions from the rodent model to the human. However the study of a process such as

antigenic variation requires *in vivo* investigation and we are thus concentrating on a protein from the rodent malaria *P. chabaudi*, the only rodent malaria known to undergo antigenic variation under laboratory conditions (Fischer *et al*, 2003).

5.2 Sequence Considerations

We chose to study the Cir2 protein of *P. chabaudi* as a representative PIR protein. This protein was known to be expressed in the parasites at the time of selection (Dr. C. Janssen, personal communication). An an antigenically variable protein, the PcCir2 protein is expected to be present at the surface of the RBC, as has been shown for other members of the CIR and PIR family (Janssen *et al* 2004, and Cunningham *et al*, 2005).

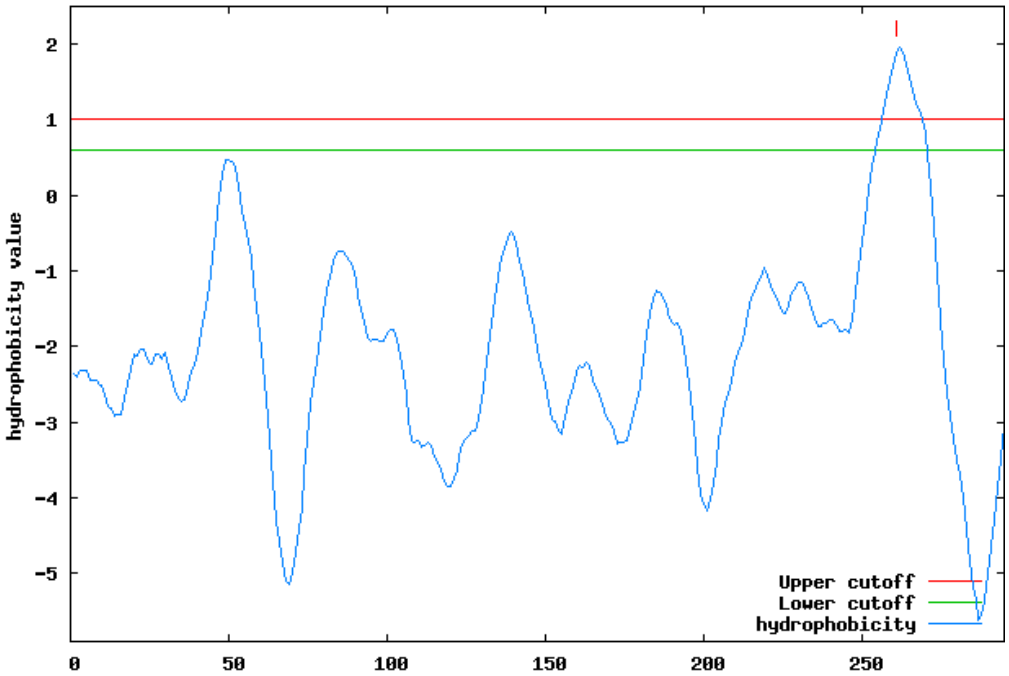


Figure 5.1: Hydropathy plot of the PcCir2 protein. The y-axis is the hydrophobicity value, the higher the value the more hydrophobic a set of amino acids is deemed to be. The x-axis displays the amino acid residue number. A single predicted TMD for Cir2 is observed approximately at residues 260 – 280. Default settings for upper and lower hydrophobicity value were used and can be seen as a red and green line running across the images, respectively. Peaks above the red line are taken as TMDs, and those above the green line are putative TMDs. 'start position of window in sequence' denotes the sliding window of 21 amino acids, which can be found in the 312 amino acid PfNitA sequence. Diagram taken from Mobyl@Pasteur v1.0.

The SignalP and IPSORT (Section 2.6.6 and Table 2.6) computer programs were used to identify any signal peptides the PcCir2 protein sequence may contain. Both programs returned a negative result, indicating they do not predict a signal sequence in PcCir2. The PIR protein family have been postulated to have 1 or 2 transmembrane domains (TMDs). To investigate the number of TMDs present within the PcCir2 protein, the amino acid residue sequence was subject to *in silico* analyses using membrane protein topology predictions, summarised in Table 5.1. The hydropathy plots obtained from these analyses indicated that the protein contains one TMD (Figure 5.1) between residues Leu 263 and Tyr 281. The amino acid sequence (LIPGLLIFAAIPVFLGIAY) is flanked by two lysine residues Lys 262 and Lys 282. These positively charged side chains are unlikely to be located within the putative TMD. However this possibility cannot be ruled out.

TMD Predictor	TMDs	Orientation
TMHMM	1	N-out
HMMTOP	1	N-out
SOSUI	1	N-out
TMpred	1	N-out
TopPred	1	N-out
Consensus	1	N-out

Table 5.1: Membrane topology predictions. Multiple membrane protein topology predictions were used to predict the number of transmembrane segments in PcCir2 and the orientation through the membrane. Computer program URLs are given in Section 2.6.2.

The hypervariable region (HVR) was established using a Plotcon (EMBOSS) program as reported by Janssen *et al* (2004). The program positions the HVR before the predicted TMD and covers around 41 residues given as: 213 NKKKGESCDFPSLPQISPKKSFAQNSLESPGHTSGHNSEDI 253. The study also established 5 sequence motifs conserved between members of the *pir* and *rif* gene families (Janssen *et al*, 2004). Two of these conserved motifs are present in PcCir2. The first motif (Motif 3) [KV]-L-C-I-[YF]-[LA]-[YI]-[LIY]-W-L, is present in PcCir2 with the residue sequence as 85 KLAQYAILWL 94 and the second (Motif 4) [CF]-D-K-[ED]-I-Q-K-[IQ]-[IY]-L-[KR]-[DE] is present as 287 FDKQRHRQYLRE 298. The schematic PcCir2 protein presented in Figure 5.2, A highlights the positioning of these key features. The predicted 2D model is presented with the key features (Figure 5.2, B).

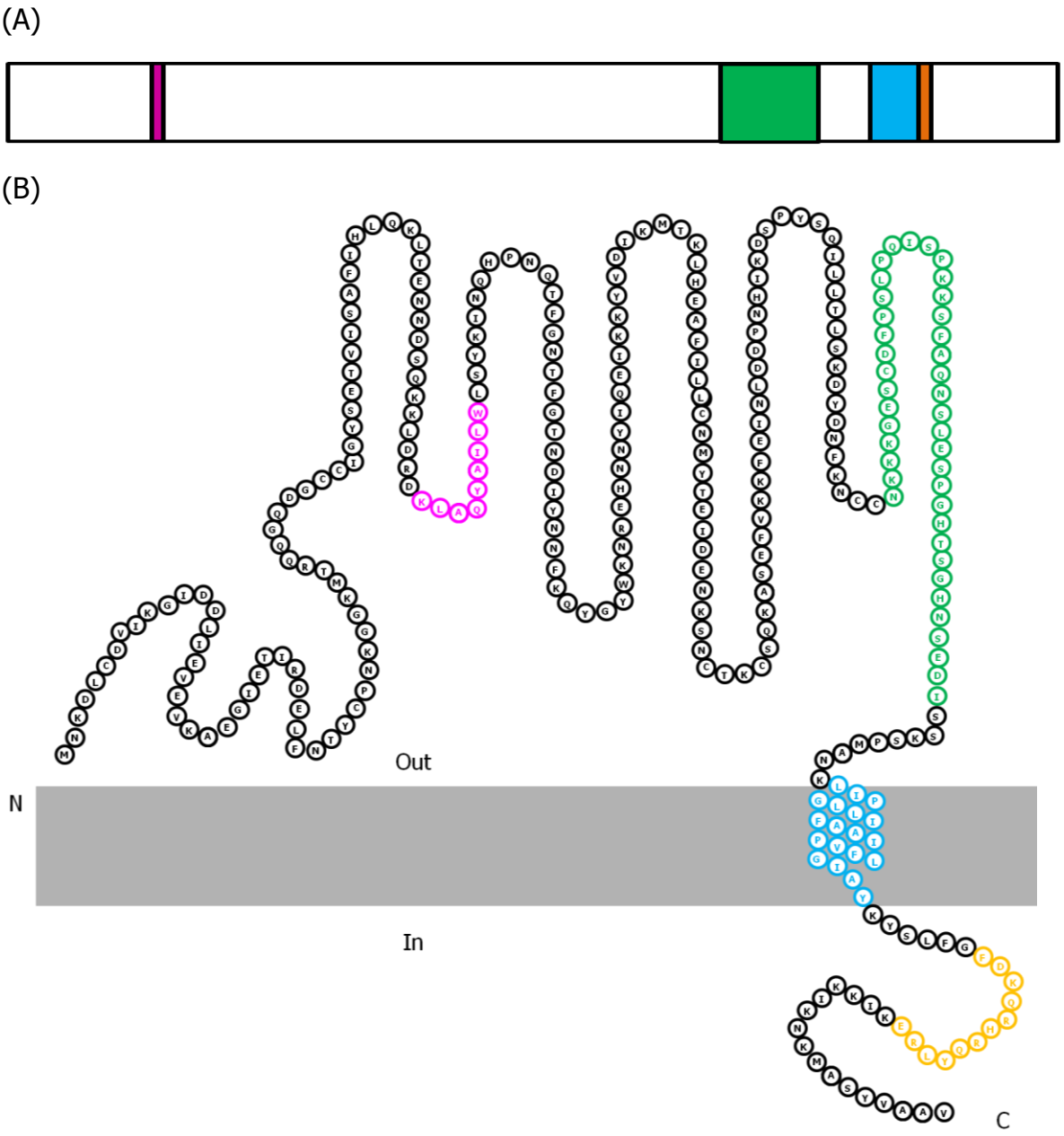


Figure 5.2: PcCir2 predicted 2D membrane topology. (A) A schematic of the PcCir2 residue sequence and (B) predicted 2D topology, illustrating the key features in different colours. The HVR is green, the TMD in blue, the grey bar represents the lipid bi-layer. The two *pir* family motifs observed in the PcCir2 residue sequence are presented in pink and orange. N and C denote the termini. Diagram is not to scale.

5.3 Recombinant Expression of Full-length PcCir2 in *E. coli*

To facilitate expression in *E. coli*, the *pccir2* coding sequenced was codon-optimised to reduce the A+T richness from 71 % to 56 %. Whilst redesigning the gene sequence,

it was important to ensure (i) the removal any Shine-Dalgarno motifs (AGGAGG) which could be misread as ribosomal binding sites, and (ii) to omit repeats of A's that could be mistaken for polyA tails and degraded. The synthetic codon-optimised *P. chabaudi cir*2 gene and protein are herein referred to as *pccir*2 and PcCir2, respectively.

5.3.1 Optimisation of PcCir2 expression in pTTQ18 and pTTQ10H

The synthetic codon-optimised *pccir*2 gene was inserted into the IPTG inducible vectors pTTQ18 and pTTQ10H to give the PcCir2 protein expression plasmids pTTQ18_*pccir*2 and pTTQ1OH_*pccir*2 (Figure 5.3, A and B). These plasmids encode a hexa-histidine and deca-histidine His-tag, respectively, and the linking sequence RGS (Stark, 1987). Expression of the *pccir*2 coding region results in a protein with a C-terminal His-tag fusion.

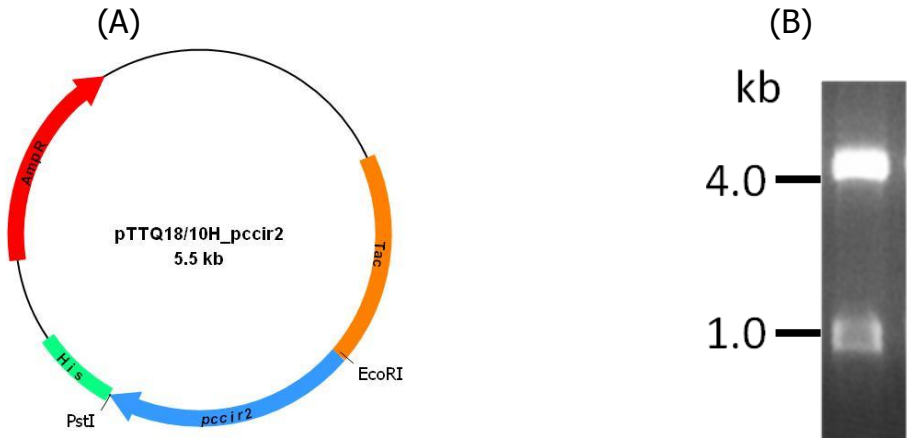


Figure 5.3: Cloning of *pccir*2 into the TTQ vectors for expression in *E. coli*. (A) Plasmid construct of *pccir*2 upstream from a His-tag. pTTQ10H encodes 10 histidine residues, and pTTQ18 encodes 6. The restriction sites used to insert the *pccir2* gene are indicated as well as the ampicillin resistance (AmpR) gene which acted as a selectable marker during cloning. The *tac* promoter is upstream of the inserted gene for inducible expression using IPTG. Diagram is not to scale. (B) The 938 bp *pccir*2 gene was cloned into the 4500 bp TTQ vectors using the restriction sites *Eco*RI and *Pst*I.

Small scale expression trials using two *E. coli* strains, BL21 (DE3) and C43 (DE3), transformed with either plasmid were carried out with a number of conditions tested. The C43 (DE3) strain was demonstrated to be particularly effective in the production of membrane proteins through two uncharacterised mutations (Miroux and Walker, 1996). Bacterial cells were grown in two types of media, LB or 2xTY (Appendix 1), with the inducer IPTG used at 0.1 mM, 0.5 mM or 1 mM. The temperatures tested

after induction were 37 °C or 22 °C, the amount of time after induction varied with cells grown for 3 or 6 h. Bacterial samples were harvested and crude membranes prepared. The preparations were analysed by SDS-PAGE and staining with a His-tag stain.

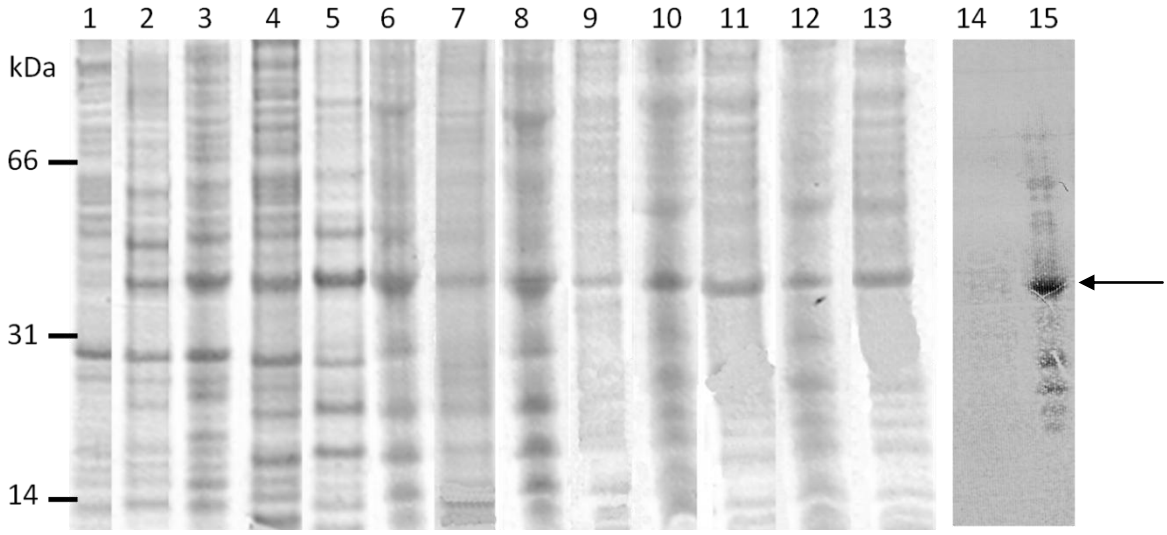


Figure 5.4: Small scale expression trials of PcCir2 from pTTQ18 and pTTQ10H. Crude membrane preparations were run on 4–12% NuPAGE gels and stained with CBB (lanes 1 – 13). Lane 1 is an uninduced control. Protein was expressed in C43 (DE3) (lanes 2–3) and BL21 (DE3) (lanes 4 – 13) *E. coli* host strains transformed with pTTQ18-*pccir*2 (lanes 2, 4, 6, 8, 10 and 12) and pTTQ1OH-*pccir*2 (lanes 3, 5, 7, 9, 11 and 13). The bacterial cells were cultured in different growth media: LB (lanes 2 – 7, 10 and 11) and 2xTY medium (lanes 8, 9, 12 and 13), and were induced with 1 mM (lanes 2 – 5) and 0.1 mM (lanes 6 – 13). All samples were obtained from a growth time of 3 h after induction with IPTG, cells were incubated at 37 °C (lanes 2 – 9) and 22 °C (10 – 13). Lanes 14 and 15 display the His-tag stained gel, lane 14 corresponding to lane 1 and lane 15 corresponding to lane 6. 15 µg of total protein extract was loaded in each lane. An arrow indicates the position of PcCir2. Molecular weight markers are indicated on the left hand side.

The data presented in Figure 5.4 displays proteins from crude membrane preparations of cells grown under the various conditions following electrophoresis in NuPAGE gels and staining using Coomassie brilliant blue (CBB) (lanes 1 – 13) or the His-tag stain (lanes 14 and 15). The PcCir2 protein is indicated with an arrow and conforms to the expected molecular weight of 37.4 kDa (hexahistidine tag) and 38.0 kDa (decahistidine tag) (ExPASy Compute pI/Mw tool[31]). Lane 1 is an uninduced control with pTTQ18-*pccir*2 in which the 37.4 kDa protein is not observed.

[31] http://www.expasy.ch/tools/pi_tool.html. Accessed on 26 March 2009.

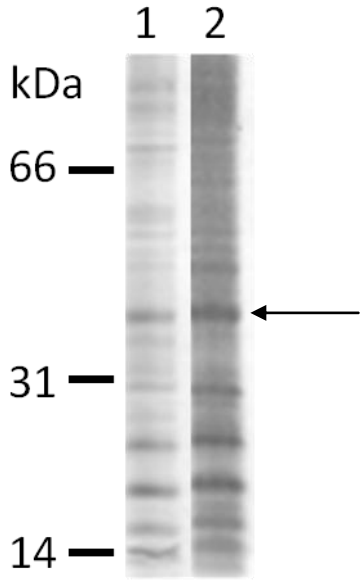


Figure 5.5: Large scale membrane preparations. Lane 1 contains the PcCir2 protein expressed in a crude membrane preparation from the large scale expression. This was prepared to verify that PcCir2 protein is well expressed and confirm the expected size. Lane 2 displays the *E. coli* membranes from the small scale expression trial, and confirms protein expression in each preparation. 15 µg of total protein was loaded in each lane. An arrow indicates the position of PcCir2. Molecular weight markers are indicated on the left hand side.

Several conditions produced good expression and the conditions chosen for further studies were the BL21 (DE3) strain transformed with pTTQ18-*pccir*2 and grown in LB media, cells induced with 0.1 mM IPTG and grown for a further 3 h at 37 ^{o}C (Lane 6). The expression trials incubated for 6 h produced less protein than the 3 h samples and are not presented. The decrease in observed protein after 6 h was possibly due to degradation of the protein as there were a number of smaller bands present on the gel when stained with the His-tag stain. These low molecular weight products imply degradation of the protein, premature termination of protein translation or non-specificity of the His-tag stain.

Large scale attempts to obtain the PcCir2 protein in milligram quantities were carried out using the optimised conditions (detailed above) and 12 L LB broth. The reason for such a large culture volume is that membrane proteins are expressed at relatively low quantities as they must be expressed in bacterial membranes without toxicity to the cell. Recombinantly expressed membrane proteins may result in adverse effects to the host bacteria for a number of reasons including the formation of perforations in the

membrane causing an osmotic disparity and interfering with targeting of native membrane proteins.

A small aliquot equivalent to 15 µg of protein from the membrane preparation was analysed for presence of PcCir2. The results presented in Figure 5.5 demonstrate that the PcCir2 protein was expressed in the membrane. The membrane preparation provided 2.25 g of total protein, from 24.6 g of cells.

5.3.2 Purification of PcCir2

PcCir2 is a transmembrane protein, therefore detergents are required to release the protein from the lipid bi-layer.

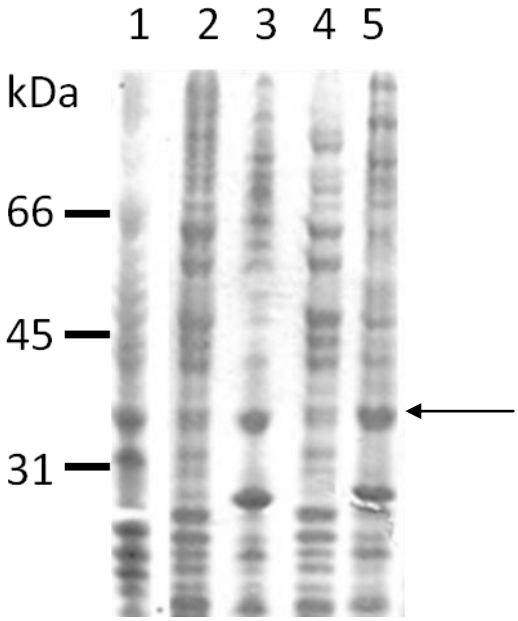


Figure 5.6: Solubilisation of PcCir2 using DDM and OG. A CBB stained gel with PcCir2 in lane 1, the DDM (n-Dodecyl-ß-D-maltopyranoside) soluble and insoluble fractions are in lanes 2 and 3, respectively. The soluble and insoluble fractions from incubation with OG (n-Octyl-ß-D-glucopyranoside) are in lanes 4 and 5, respectively. Equal volumes of the soluble and insoluble fractions were loaded in each lane. An arrow indicates the position of PcCir2. Molecular weight markers are indicated on the left hand side.

The PcCir2 protein is thought to be monotopic. Thus, the detergents used for the crystallisation of monotopic proteins[32] were taken into account. Only a few monotopic

[32] http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html. Accessed on 16 January 2010.

proteins have been crystallised, consequently no consensus detergent was discovered. For this reason, two detergents DDM and OG available in our laboratory were trialled first. Total protein extract were solubilised in 1 % or 2 % of the detergents for 1 h or overnight, and incubated at 4 ^{o}C on ice.

Soluble and insoluble fractions were separated by high speed centrifugation and separated on a Nu-PAGE gel. The results obtained were not encouraging as most of the protein remained in the insoluble fractions (Figure 5.6, lanes 3 and 5). Further detergents of short or long chain length and of non-ionic, ionic or zwitterionic nature were obtained and tested at different percentages (Table 5.2).

Abbreviation	Chemical	% Tested	CMC mM (%)	Properties
CH	CHAPS*	2, 4	8.0 (0.49)	zwitterionic
DA	Deoxycholic Acid	2, 4	6.0 (0.24)	anionic
DDM	n-Dodecyl-ß-D-maltopyranoside	1, 2	0.17 (0.0087)	long, non-ionic
DM	n-Decyl-ß-D-maltopyranoside	n/a	1.8 (0.087)	non-ionic
FC-12	Fos Choline-12	0.5, 1	1.5 (0.047)	harsh, zwitterionic, short
FC-14	Fos-Choline-14	0.5, 1, 2, 4	0.12 (0.0046)	harsh, zwitterionic, long
OG	n-Octyl-ß-D-glucopyranoside	1, 2	18 (0.53)	short, non-ionic
OT	n-Octyl-ß-D-thioglucopyranoside	2, 4	9 (0.28)	non-ionic
TX	Triton X -100	2 ,4	0.23 (0.015)	mild, anionic
ZG	Zwittergent 3-14	2, 4	0.01 - 0.06	zwitterionic

Table 5.2: Detergent properties. Details the various detergents tested in the solubilisation of PcCir2 and the various percentages they were tested at. The CMC and associated chemical properties are also given. *3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS).

Detergent 1	Detergent 2
FC-14	DDM
FC-14	ZG
FC-14	TX
DDM	DA
DDM	TX
DDM	ZG
TX	ZG
TX	DA
ZG	DA

Table 5.3: Detergent combinations. The combination of detergents that were tested, each detergent was used at 1 %.

To increase the chances of obtaining soluble protein after incubation with the detergents combinations were also used (Table 5.3). Each of the solubilisation trials was incubated on ice at 4 ^{o}C overnight. The critical micelle concentration (CMC) is

given in Table 5.2 which is the minimum concentration at which micelles will be spontaneously formed in a liquid. It is important not to use a concentration of detergent that falls below its CMC.

The soluble and insoluble membrane protein fractions of the detergent tests were separated as before and samples applied to a NuPAGE gel. The solubilised proteins from a selection of the detergents tested are presented in Figure 5.7, A. The combination of DA and DDM gave the most solubility. However, it was determined that FC-14 and FC-12 were the most effective at releasing protein from the *E. coli* plasma membrane. Purification trials were continued using FC-14 as this has the lower CMC, and latterly FC-12.

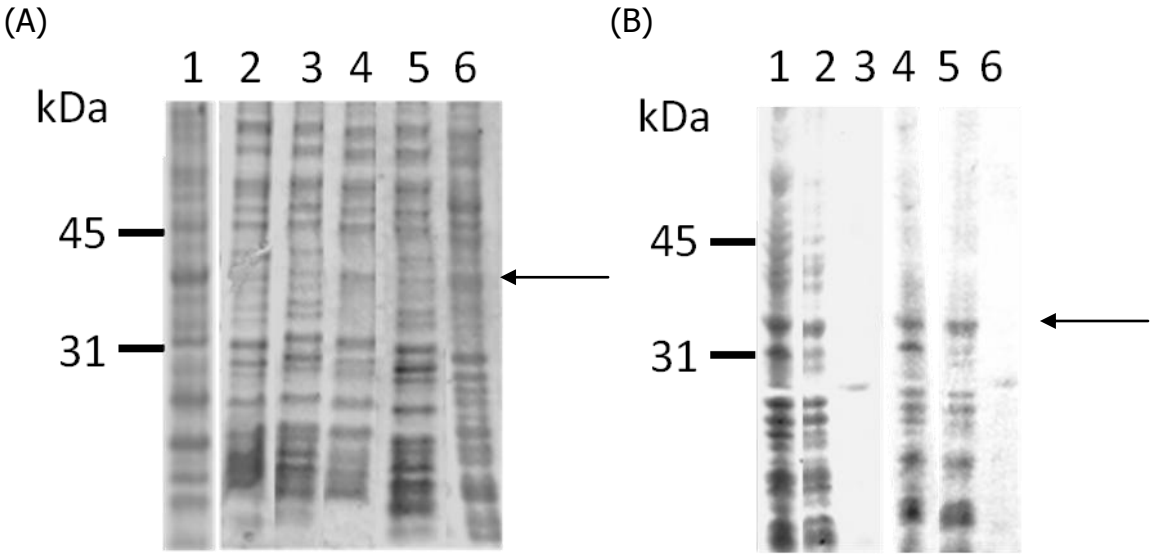


Figure 5.7: Solubilisation of PcCir2 using further detergents. (A) Lane 1 contains the optimised PcCir2 protein from a small scale expression trial. Proteins solubilised from incubation with the detergents ZG, CH, DA, ZG + DA and DDM + DA are given in lanes 2 – 6, respectively. (B) NuPage gel stained with CBB (lanes 1 – 3) and the His-tag stain (lanes 4 – 6). Lanes 1 and 4 are optimised PcCir2 from a small scale expression trial. The other lanes contain protein solubilised using 0.5 % FC-14 in an overnight incubation experiment. Detergent solubilised membrane protein is observed in lanes 2 and 5, and insoluble in lanes 3 and 6. Equal volumes of the soluble and insoluble fractions were loaded in each lane. The arrows indicate the position of PcCir2. Molecular weight markers are indicated on the left hand side.

One drawback of these detergents is that they solubilise most of the proteins in the membrane, and many of proteins are present when stained with the His-tag stain as can be seen in Figure 5.7, B. As the recombinant protein was insoluble in most of the detergents trialled, it is possible the protein was packaged into inclusion bodies (IB's)

and that is the reason for the lack of solubility, this possibility was not examined. However it should be noted that in section 5.5.3 and Figure 5.31, the PcCir2 protein is expressed (albeit from a different vector) and shown not to be present in IB's.

Despite the possible contaminating protein bands, the PcCir2 protein was purified using a cobalt IMAC column to bind the His-tag after solubilisation overnight using FC-14. Protein was eluted in fractions 1 – 13 and a total of 20 x 0.25 ml fractions were collected for analysis. During elution, the concentration of imidazole was increased incrementally when the OD_{280} of samples began to decrease. After 8 eluted fractions the imidazole was increased from 200 mM to 300 mM, after elution fraction 12 the imidazole was increased to 400 mM and for elutions 17, 18, and 19 the concentration was increased to 500 mM, this can be seen by the peaks and troughs on the chart displayed in Figure 5.8, A . Protein content in the fractions was assessed by reading the OD_{280} on a nanodrop as Bradford's reagent reacted with the FC-14 and gave a false positive result. The chart in Figure 5.8, A was generated by measuring the OD_{280} of each of the eluted samples. The extinction coefficient was calculated as an OD_{280} of 0.850 = 1 mg/ml protein (Pepstats, EMBOSS[33]).

To verify the correct protein had been eluted, a small volume was analysed on a NuPAGE gel. The preparation was not particularly clean as numerous extra bands of could be detected on the gel one of these particularly close to the PcCir2 band Figure 5.8, B. Once the protein sequence had been confirmed by mass spectrometry (MS) analysis, the fractions were applied to Vivaspin columns with molecular weight cut-offs (MWCO) of 50 and 30 kDa for concentration by diafiltration. However, many of the smaller protein bands could still be observed and a few larger bands were also present when the concentrated proteins were applied to a NuPAGE gel. In addition, the 50 kDa MWCO device allowed some of the protein to be eluted in the flow through. The larger fragments gave a ladder-like appearance, inferring the protein may have aggregated during concentration.

Since the detergent can play a role in the stability of the protein and its monodispersity, FC-12 was also tested. As before, during elution the concentration of

[33] http://emboss.bioinformatics.nl/cgi-bin/emboss/pepstats. Acessed on 20 August 2012.

imidazole was increased incrementally. After 8 eluted fractions the imidazole was increased from 200 mM to 300 mM, after elution 12 increased to 400 mM and for elutions 17 and 18 the concentration was increased to 500 mM, this can be seen by the peaks and troughs on the chart present in Figure 5.8, C (protein concentration was calculated as before).

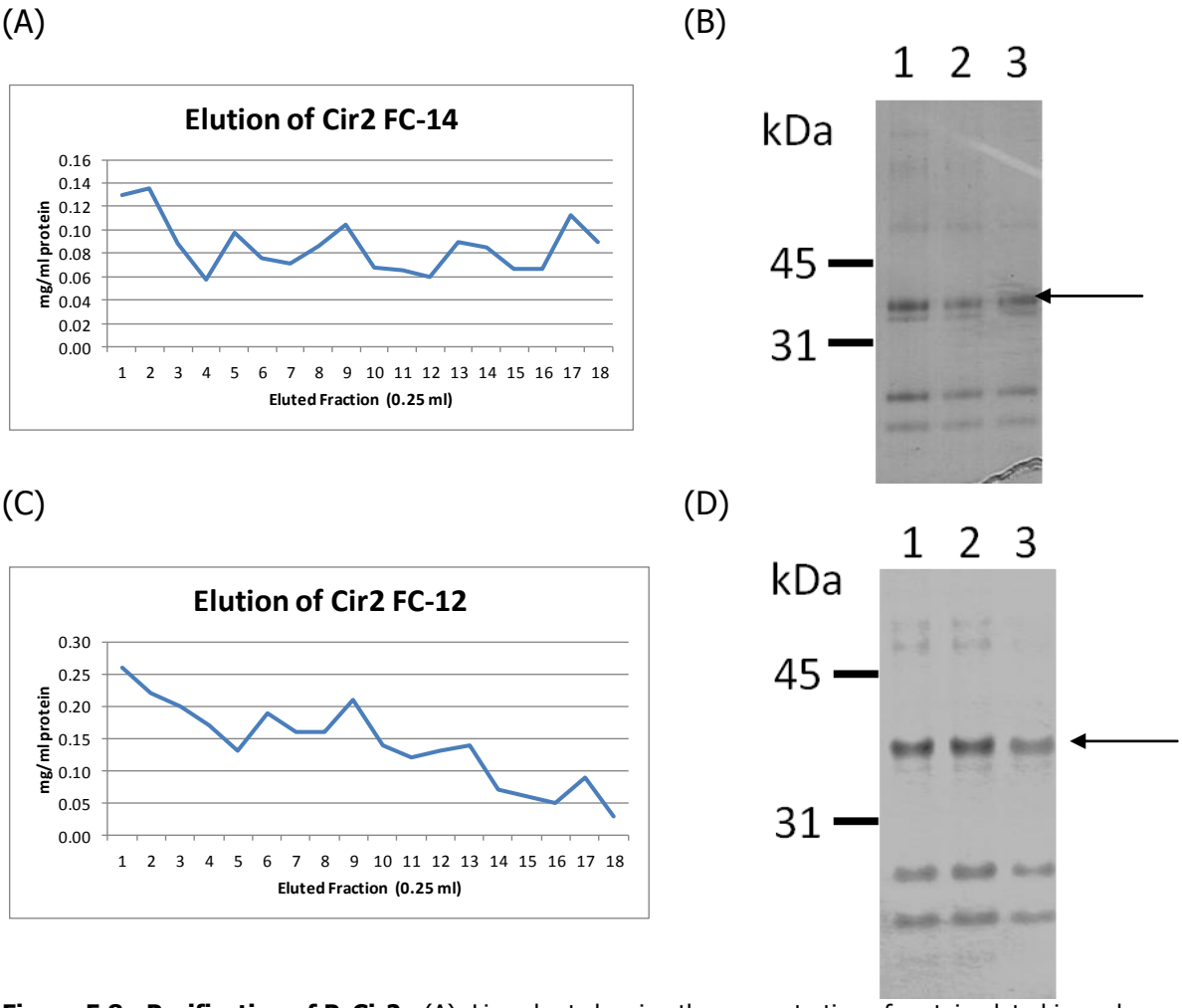


Figure 5.8: Purification of PcCir2. (A) Line chart showing the concentration of protein eluted in each fraction, after solubilisation and purification of PcCir2 using FC-14. (B) A CBB stained gel, lanes 1 – 3 are eluted fractions 2, 5 and 7. (C) The chart illustrates the concentration of protein eluted in each fraction, after solubilisation and purification of PcCir2 using FC-12. (D) A CBB stained gel, lanes 1, 2 and 3 display the elution fractions 2, 5 and 7. The same volumes of elutate were loaded on the gel. An arrow indicates the position of PcCir2. Molecular weight markers are indicated on the left hand side.

The results are presented in Figure 5.8, D and many additional bands were still present in the samples when proteins were separated using a Nu-PAGE gel. The major lower contaminating bands appear to be the same sizes in the gels from FC-12 and FC-14

solubilisation. From each preparation three - four higher molecular weight protein bands are also observed. In particular, the protein band closest to the PcCir2 protein is present in both purifications.

Approaches to evaluate the quality of a protein and to characterise a membrane protein require a highly pure protein. As expression in this vector did not yield a pure protein it was decided not to pursue this line of investigation further.

5.3.3 Optimisation of PcCir2 expression using pGFPe

Given the problems in purifying the protein from the TTQ18 vector, another vector was chosen. This vector, pGFPe, encodes a GFP reporter gene, TEV protease site and His-tag. The *pccir2* gene was amplified with the oligonucleotide primers PcCir GFPe F and PcCir GFPe R (Appendix 2). The PCR product was cloned into the vector upstream of these coding regions so as to give a fusion protein with a C-terminal cleavable GFP and His-tag, the resulting plasmid was termed pGFPe_*pccir2* (Figure 5.9, A and B).

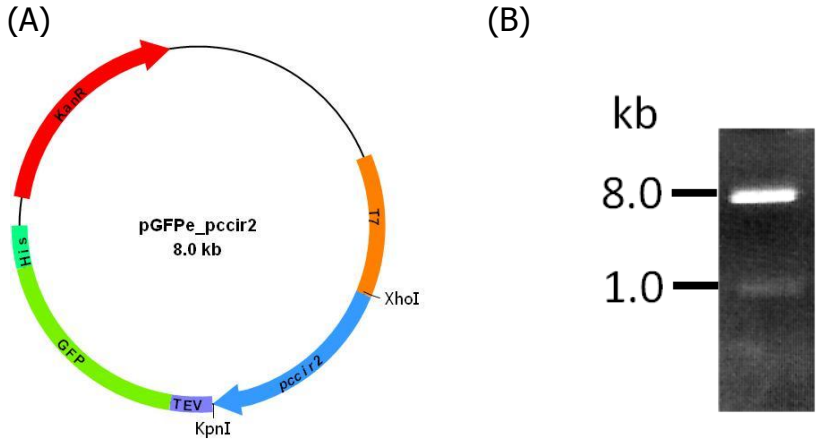


Figure 5.9: Cloning of *pccir2* into the pGFPe vector for expression in *E. coli*. (A) Plasmid construct showing the *pccir2* gene upstream from a 8xHis-tag, *gfp* gene, and TEV protease site. The restriction sites used to clone in the *pccir2* gene, *Xho*I and *Kpn*I are given as well as the kanamycin resistance (KanR) gene which acted as a selectable marker during cloning. The T7 promoter is present upstream from the gene. Diagrams are not to scale. (B) The 938 bp *pccir*2 gene was cloned into the 7000 bp pGFPe vector using the restriction sites *Xho*I and *Kpn*I.

The *gfp* reporter gene is useful in protein expression research since its addition as a fusion to other genes leads to production of proteins where the protein expression can be monitored by measuring GFP fluoresence. This implies that the fluorescent tag and

its fused protein both fold correctly and that the targeting of the fusion protein is similar to that of the native protein without the fluorescent tag. A second benefit of this plasmid is that a two-step purification process can take place. The expressed fusion protein is initially purified and eluted using a metal affinity column. The eluted protein is then applied to a metal affinity column and bound using the His-tag, a TEV protease is added to the column and the desired protein cleaved from the His-tag and is eluted.

Two colonies were tested for each condition to check for variability of expression between colonies. The results from one colony are present as protein production was similar in both. In all trials the media used was LB broth containing 50 µg/ml kanamycin. Figure 5.10 displays samples from cell lysate (lanes 2 – 10) and membrane preparations (lanes 11 – 14) with good expression observed in each of the conditions tested.

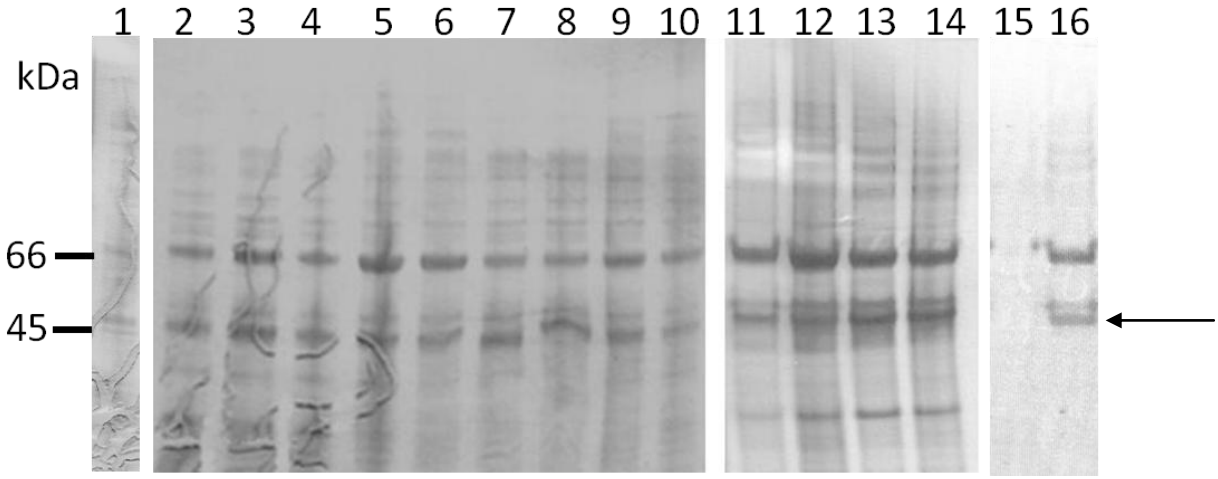


Figure 5.10: Small scale expression trials of PcCir2-GFP. Samples run on 4–12% NuPAGE gels stained with CBB. Lanes 1 – 10 contain cell lysate and lanes 11 – 14 are membrane preparations. Lane 1 is an uninduced control. The cells were induced with 0.05 mM (lanes 3, 5, 7, 9, 12 and 14) and 0.1 mM (lanes 2, 4, 6, 8, 10, 11 and 13). All samples were grown in LB, and after induction with IPTG, cells were grown at 37 ^{o}C (lanes 5, 6, 9, 10, 11 and 12) and 22 ^{o}C (lanes 2, 3, 4, 7, 8, 13 and 14). Growth time after induction was 3 h (lane 2), 6 h (lanes 3 – 6 and 11 – 14) and 8 h (lanes 7 – 10). Lanes 15 and 16 contain an uninduced and expressing (same as lane 5) culture which has been stained against the C-terminal His tag. The same volumes of cell lysate or 15 µg membrane protein was loaded on the gel. An arrow indicates the position of PcCir2. Molecular weight markers are indicated on the left hand side.

The expected molecular weight of the protein is 66.3 kDa. After MS analysis, the lower of the two prominent bands was confirmed to be PcCir2-GFP, the protein running at a lower molecular weight than expected, approximately 45 kDa. This is not unknown for membrane proteins, possibly though the reduced binding of SDS to hydrophobic areas (Rath *et al*, 2008). MS analysis indicated the approximately 66 kDa molecular weight band to be an *E. coli* protein, namely mannose-1-phosphate guanyltransferase. Protein production was deemed to be better at 6 h, and thus the experiments were repeated and crude membranes prepared (Figure 5.10 lanes 11 – 14). It was hoped that the larger MW weight band would be removed in the cytoplasmic or periplasmic fractions, but it was not. Large scale expression of PcCir2-GFP was carried out in 12 L LB broth using the following conditions, 0.05 mM IPTG and a growth time following induction of 6 h at 22 °C. The cells from this were harvested and produced a total wet weight of 46.2 g. Cells were used to prepare cell membranes. The large scale membrane preparations produced 446 mg of total protein.

5.3.4 Purification of PcCir2-GFP

Solubilisation trials were carried out for PcCir2-GFP using a single concentration (2.5 %) of each detergent. The detergents tested were DDM, ZG, OG, OT, FC-12, TX, DA and DM (Table 5.2).

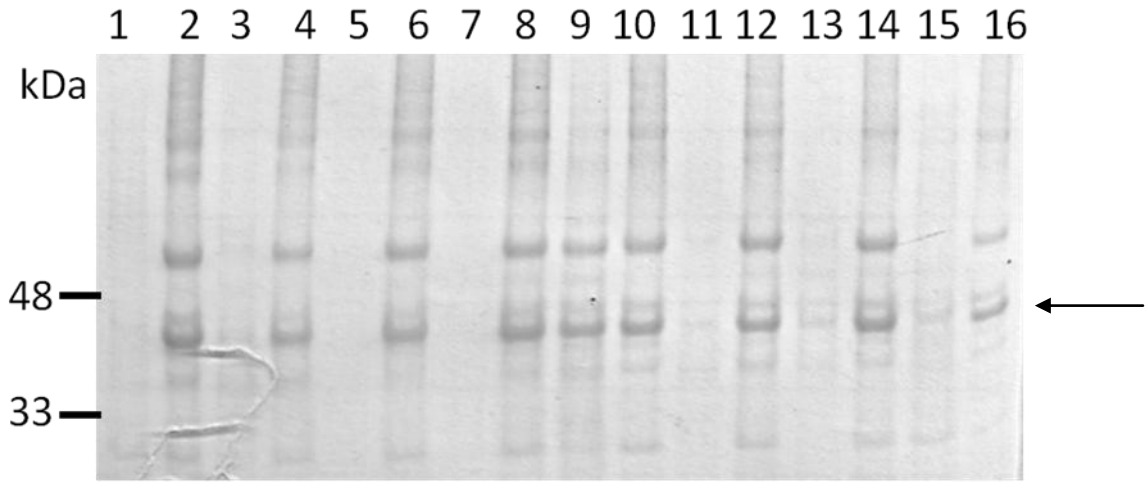


Figure 5.11: Solubilisation of PcCir2-GFP using different detergents at 2.5 %. Soluble fractions are in lanes 1, 3, 5, 7, 9, 11, 13 and 15. Insoluble fractions are in lanes 2, 4, 6, 8, 10, 12, 14 and 16. Membranes were solubilised with DDM (1 and 2), ZG (3 and 4), OG (5 and 6), OT (7 and 8), FC-12 (9 and 10), TX (11 and 12), DA (13 and 14) and DM (15 and 16). The same volumes of soluble and insoluble fractions were loaded on the gel. The arrow indicates the position of PcCir2-GFP. Molecular weight markers are indicated on the left hand side.

Solubilised and insoluble proteins from membranes solubilised with detergents were electrophoresed using a NuPAGE gel to separate the proteins by weight. The results in Figure 5.11 display the gel after staining with CBB.

Again, FC-12 provided the best solubilisation of the protein (Figure 5.11, lanes 9 and 10). However, unlike the solubilisation of PcCir2 expressed from the pTTQ18 plasmid, the solubilised fraction contains merely around 50 % of the PcCir2 protein. It could be that the GFP moiety increasing the size of the protein is making it less accessible to the detergents.

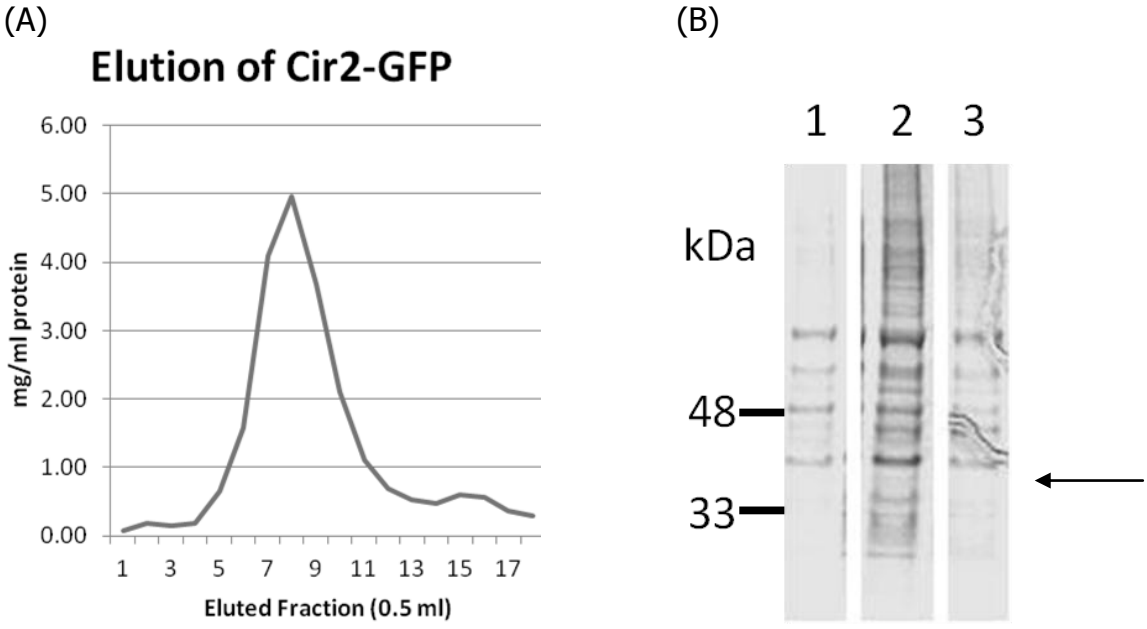


Figure 5.12: Purification of PcCir2-GFP. (A) The chart illustrates the concentration of protein eluted in each fraction, after solubilisation and purification of PcCir2-GFP in FC-12. (B) A CBB stained gel with eluted fractions 5, 8 and 11 (in lanes 1, 2 and 3 respectively). The same volumes (5 µl) of eluted fractions were loaded on the gel. The arrow indicates the position of PcCir2-GFP. Molecular weight markers are indicated on the left hand side.

The PcCir2-GFP protein was purified using an IMAC column. The detergent-membrane/protein mixed was loaded into the coloum, and protein eluted by adding 200 mM imidazole, the imidazole concentration was not increased during the elution. The purification of PcCir2-GFP was monitored visually, by the colour of the elutate. After elution fraction 14, the column was stopped to determine if more protein would elute after a short static incubation of the beads in buffer containing imidazole. The

chart in Figure 5.12, A was generated by measuring the OD_{280} of each of the eluted samples. The extinction coefficient was calculated as an OD_{280} of 0.800 = 1 mg/ml protein (Pepstats, EMBOSS). Samples of eluted fractions 5, 8, and 11 were separated on a NuPAGE gel to ensure the correct protein had been eluted (Figure 5.12, B).

Eluted fractions 5 through 16 were combined and concentrated using a Vivaspin centrifugation device with a 30 kDa MWCO. After the first centrifugation step, the flow through and 100 µl concentrate were both a bright green colour suggesting some of the protein had passed through the column, or that the GFP had somehow detached from the PcCir2 fusion. The centrifugation step was repeated by adding 6 ml of elution buffer (without imidazole) to reduce the imidazole concentration by diafiltration. This had the effect of eluting and diluting the previously concentrated protein. When eluted fractions were applied to a gel the dilution of the protein with elution buffer made the bands very faint and thus it is difficult to state if the GFP had dissociated or if the protein had passed through the concentration column. Some protein was concentrated as observed on the CBB stained gel, however, after repeated attempts at concentration the protein was lost in the flow through (data not shown). Other means of protein concentration were not investigated.

5.3.5 Optimisation of PcCir2 expression in pTTQ18-TEV

The TTQ18 vector was engineered by Dr. S.E Unkles to contain a TEV protease site between the cloned gene and His-tag to allow a two step purification process to be applied. The *pccir2* gene was cloned into the TTQ18-TEV plasmid using the *Psf*I and *Eco*RI sites as described previously (Section 5.3.1) to produce TTQ18-TEV_*pccir2*. The construct was transformed into BL21 (DE3) cells for analysis of protein expression. Two colonies were chosen for small scale expression trials of PcCir2 in the TTQ18-TEV vector. Conditions tested were temperature after induction of 37 ^{o}C and 22 ^{o}C, and a growth time of 3 and 6 h. Concentrations of IPTG used were 0.1 mM, 0.5 mM and 1 mM. An HRP conjugated anti-His tag antibody was used to demonstrate the expression of PcCir2 (Figure 5.13). Approximately three mins after detection solution had been added to the membrane, the membrane began to develop yellow-brown bands where the PcCir2 protein would be expected. This was most likely due to using a concentration of antibody that was too high. At these areas, a white band is observed on the film after exposure using the X-ray processor. The western blot shows a great deal of background staining, however the blot was not repeated at a

higher antibody dilution. Both colonies reacted similarly in the western blot and the results from one colony are displayed. Cells grown at 22 °C and 37 °C induced with 1 mM IPTG for 6 h contain a two proteins at a similar size (38 kDa) on the western blot, one of which may correspond to PcCir2 (Figure 5.13). Therefore these conditions were not chosen for large scale expression of the PcCir2 protein. The greatest expression was thus observed in Lane 1 using the conditions 0.1 mM IPTG, 3 h incubation after induction at 37 °C, the same as the expression conditions using the TTQ18 vector.

The large scale production of PcCir2 from the pTTQ18-TEV was carried out in 12 L LB broth, using the optimum conditions determined from the small scale expression trials: 0.1 mM IPTG was added after initial growth at 37 °C, and cells continued to incubate for 3 h at 37 °C (Figure 5.13). The cells from the large scale growth were harvested and produced a total wet weight of 39.0g.

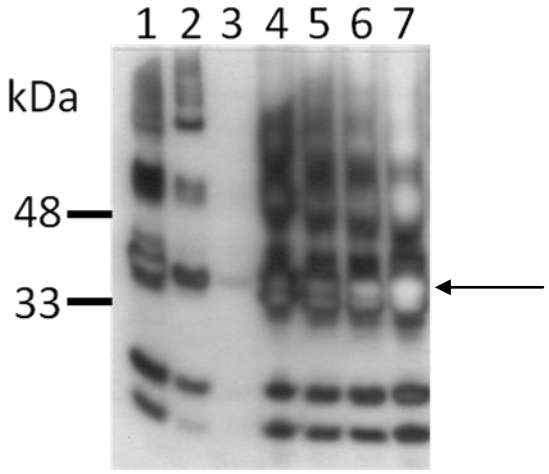


Figure 5.13: Small scale expression trials of PcCir2 from TTQ18-TEV. A western blot to illustrate the expression of PcCir2, using an antibody against the C-terminal hexahistadine tag. The cells were induced with 0.1 mM (lanes 1, 2, and 5), 0.5 mM (lane 7) and 1 mM (lanes 3, 4, and 6). All samples were grown in LB, and after induction with IPTG, cells were grown at 37 °C (lanes 1– 4) and 22 °C (5 – 7). Growth time after induction was 3 h (lanes 1 and 3) and 6 h (2, 4 – 7). 1 µg protein was loaded in each lane. The arrow indicates the expected position of PcCir2. Molecular weight markers are indicated on the left hand side.

5.3.6 Purification of PcCir2 with TEV cleavage site

Six detergents DDM, ZG, FC-12, TX, DA, and DM, (Table 5.2) with mixed properties were used to test for the solubilisation of PcCir2 when expressed in the pTTQ18-TEV vector. Cells were grown in 200 ml LB and expressed using the conditions optimised

above followed by crude membrane preparation. Each of the detergents was used at a concentration of 2.5 %, and incubated with the membranes overnight at 4 °C on ice. The solubilised fractions were separated from the insoluble fractions by centrifugation.

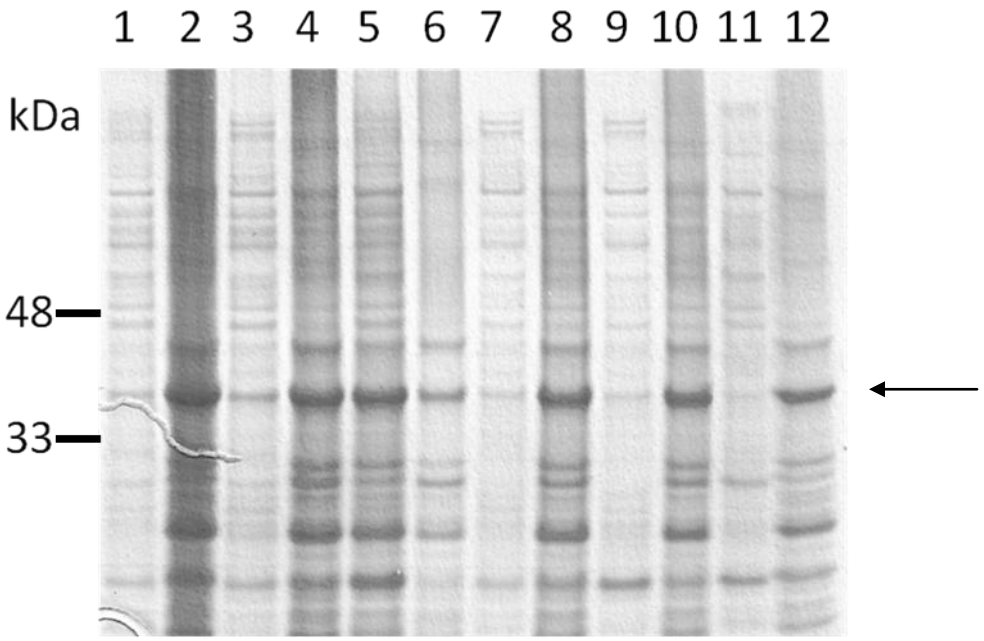


Figure 5.14: Solubilisation using different detergents at 2.5 %. Soluble fractions are in lanes 1, 3, 5, 7, 9, and 11. Insoluble fractions are in lanes 2, 4, 6, 8, 10, and 12. Membranes were solubilised with DDM (1 and 2), ZG (3 and 4), FC-12 (5 and 6), TX (7 and 8), DA (9 and 10) and DM (11 and 12). The same volumes of soluble and insoluble fractions were loaded on the gel. An arrow indicates the position of PcCir2. Molecular weight markers are indicated on the left hand side.

Samples were run on a NuPAGE gel to determine in which fraction the PcCir2 protein was present in (Figure 5.14) and therefore the best detergent for solubilisation of PcCir2 from membranes. A sample was extracted from the gel and sent for MS analysis which confirmed the presence of the PcCir2 protein. FC-12 was the best at solubilising PcCir2 (lanes 5 and 6) with slightly more protein detected in the soluble fraction.

Cells from the large scale culture were used to prepare membranes. The large scale membrane preparations produced a total of 572 mg protein. Protein was solubilised overnight in 0.5 % FC-12 at 4 °C on ice. The following morning the soluble fraction was separated by centrifugation and protein purified using a cobalt IMAC column to bind the hexahistidine tag. Protein was eluted in 0.5 ml fractions and the OD_{280}

measured. The column was stopped after the elution of fraction 6 to allow the imidazole to elute more protein. After elution fraction 14, the imidazole concentration was increased to 300 mM and allowed equilibrate. No more protein was eluted at this stage.

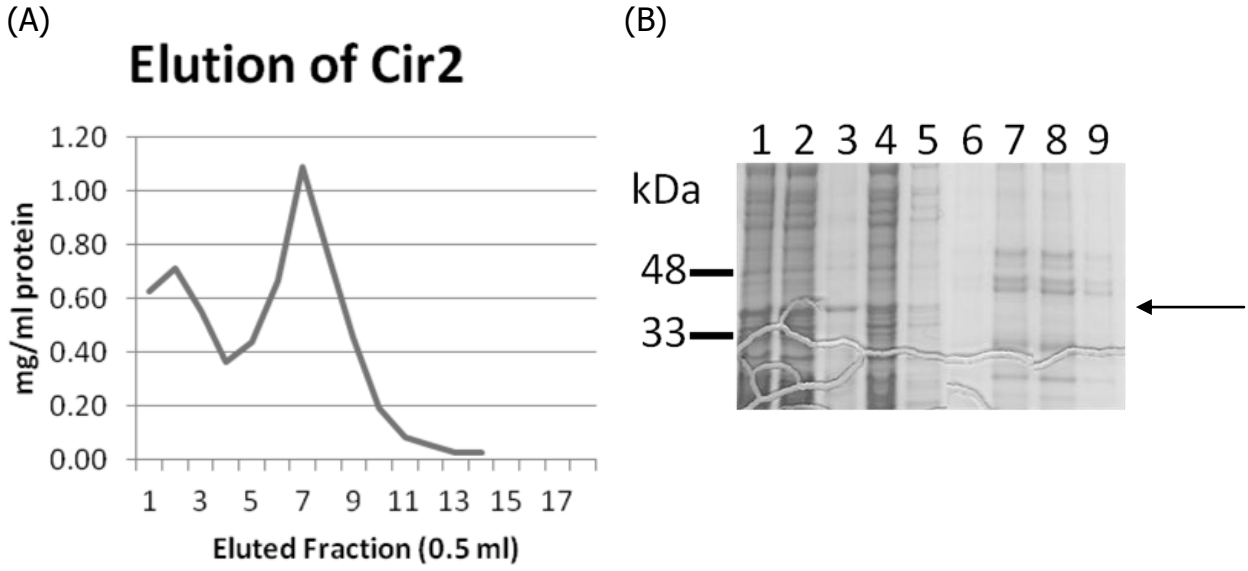


Figure 5.15: Purification of PcCir2 with TEV cleavage site. (A) The chart presents the concentration of protein eluted in each fraction, after solubilisation and purification of PcCir2 in FC-12. (B) Eluted fractions have been separated on a NuPAGE gel, lane 1 displays the protein and detergent suspension; lanes 2 and 3 contain the soluble and insoluble fractions after overnight incubation in FC-12, respectively; lanes 4 and 5 show the flow through and wash of the metal affinity beads, respectively. The eluted fractions 4, 6, 8 and 10 are in lanes 6 – 9, respectively. The same volumes of each of the samples (2.5 µl) were applied to the gel. The arrow indicates expected position of PcCir2 as no MS analysis was carried out. Molecular weight markers are indicated on the left hand side.

The extinction coefficient was calculated as an OD_{280} of 0.860 = 1 mg/ml protein (Pepstats, EMBOSS) and the information used to prepare the chart in Figure 5.15, A. Figure 5.15, B displays proteins separated on a NuPAGE gel. The PcCir2 protein can be seen in the first 5 lanes but not 6 – 9 which are eluted fractions 4, 6, 8 and 10. Protein from a large scale culture is provided as a comparison in lane 1. In previous experiments (Sections 5.3.1 and 5.3.3) and the western blot in Figure 5.13, the PcCir2 protein was found to be well expressed and it was therefore surprising that in this instance the quantity of PcCir2 was less than anticipated. The soluble and insoluble fractions after the high speed centrifugation step are presented in lanes 2 and 3, respectively. As was expected from the small scale trials, the protein is present in both

fractions. What was not expected was for the protein to be contained in the flow through after incubation with the IMAC beads and in the wash fraction. This implies the protein was unable to bind to the metal affinity column. The eluted fractions (lanes 6 – 9) demonstrate the presence of some proteins, but none corresponding to that of PcCir2.

Time permitting, experiments using different metal affinity columns such as Ni^{2+} or Zn^{2+} to find out if these ions would provide a purer PcCir2 protein. Further to this, gel filtration (size exclusion chromatography) would take place to separate the eluted proteins on size. Gel filtration should elute the protein with fewer contaminants and remove any large aggregates which may have formed. The gel filtration method can be monitored to analyse the monodispersity of the protein especially as a monodisperse protein is required for crystallography for example flouresence size exclusion chromatography (FSEC).

Given the problems with expression and purification of PcCir2 using each of the expression plasmids, it was decided to concentrate on expressing domains of PcCir2 which lacked the TMD and should be soluble.

5.4 Recombinant Expression of PcCir2 Domains in *E. coli*

A putative ordered domain, ending just before the HVR and not including the TMD, was identified by Dr. R. Russell (University of St Andrews), where he studied the secondary predictions for the PcCir2 protein, looking for ordered regions (sheets/helices) and disordered regions. As this domain is lacking the TMD and HVR it should be soluble and therefore would not require optimisation of detergent solubilisation, and purification might be easier. It was thought that any structural data acquired from the domain would corroborate any structural data gained from the full-length protein. A schematic illustrating the placement of the domain with regards the full-length gene is given in Figure 5.16.

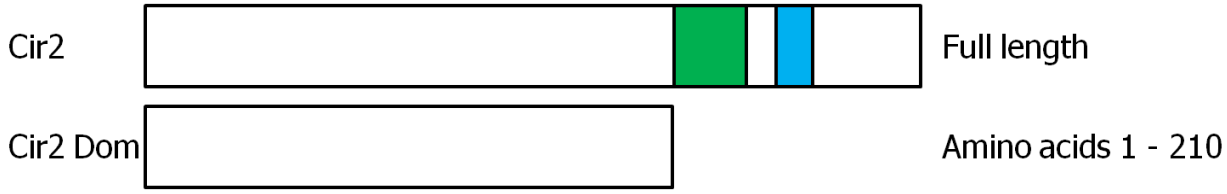


Figure 5.16: PcCir2 protein domain 1. A schematic of the protein and protein domain, full-length PcCir2 (Cir2) and PcCir2Domain (Cir2Dom). Indicated by the blue block is the TMD, the green block shows the position of the HVR. Diagrams are not to scale.

5.4.1 Optimisation of PcCir2 Domain expression in pEHISGFPTEV and pEHISTEV

DNA encoding the putative soluble domain, *pccir2domain* (*pccir2dom*) was amplified, using primers PcCir2Dom F and PcCir2Dom R (Appendix 2), from the synthetic codon-optimised *pccir2* gene. The 630 bp domain was inserted into plasmid vectors pEHISGFPTEV and pEHISTEV producing pEHISGFPTEV_ *pccir2dom* and pEHISTEV_ *pccir2dom*, respectively (Figure 5.17 A and B).

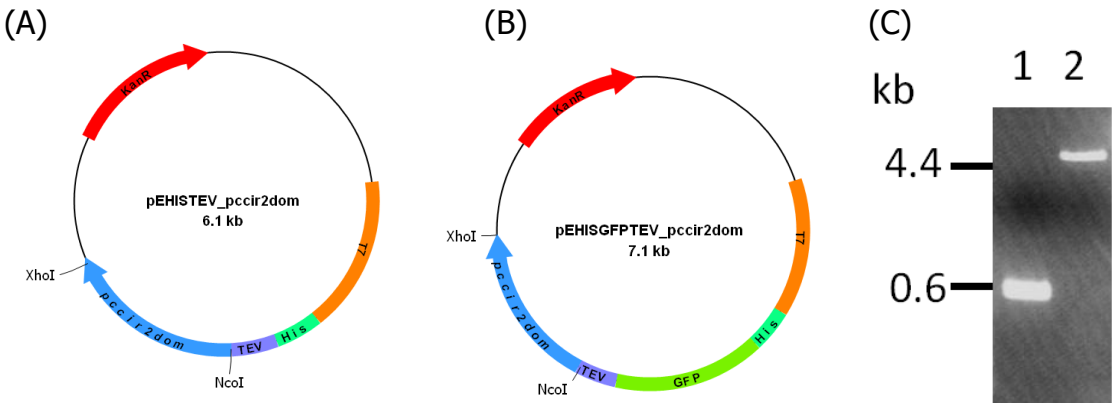


Figure 5.17: Cloning of *pccir2dom*. (A) Plasmid constructs displaying the *pccir2dom* downstream from a His-tag and TEV protease site. (B) Plasmid constructs showing *pccir2dom* downstream from a His-tag, GFP gene and TEV protease site. The restriction sites used to clone in the *pccir2dom* gene are given as well as the kanamycin resistance (KanR) gene which acted as a selectable marker during cloning. The T7 promoter is upstream from the gene. (C) The 630 bp *pccir2dom* (lane 1) was cloned into the 6500 bp pEHISGFPTEV vector (lane 2) using the restriction sites *Xho*I and *Nco*I.

The pEHISGFPTEV plasmid contains a *gfp* gene situated N-terminally to the cloned gene and a TEV protease site to allow cleavage of the GFP protein. In addition both plasmids contain a His-tag situated at the N-terminus of the cloned gene to allow purification of the expressed protein. The *gfp* reporter gene is useful in protein

expression research as fluorescence is observed only when the GFP polypeptide has been folded correctly and implies the downstream protein will also be correctly folded.

The GFP containing vector has the added bonus that only a small (100 µl) sample of bacterial cell suspension is required to observe fluorescence using a standard UV transilluminator. This meant that a large number of conditions could be screened in a relatively short time without the need of protein gels and western blotting. Samples with GFP fluorescence were applied to a NuPAGE gel and all provided good expression of the expected 54.6 kDa band (ExPASy). Induced cells grown at a lower temperature (22 ^{o}C, lanes 1 – 4) and for longer time periods (6 – 8 h, Lanes 2 and 3) showed better expression (Figure 5.18) possibly due to the longer growth time additional protein was produced. For each condition giving rise to GFP fluorescence, the whole cell soluble and insoluble fractions were separated and analysed by exposure of the fraction to UV light source. GFP fluorescence of the samples was assessed by comparison with a strain expressing GFP alone (the empty vector pEHISGFPTEV).

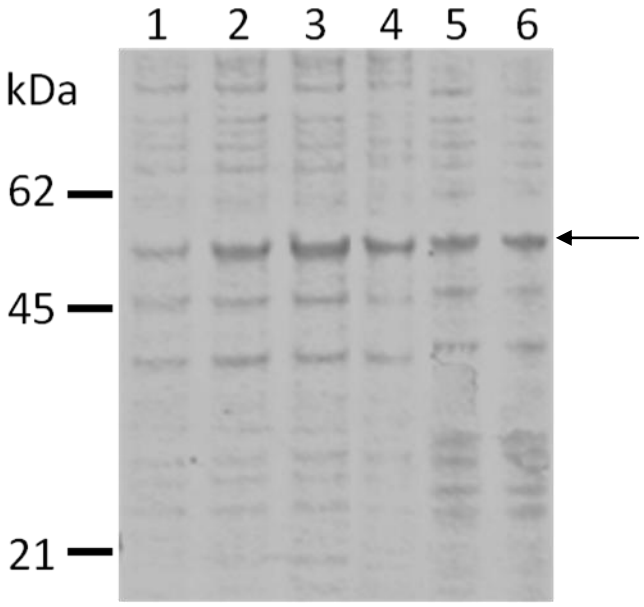


Figure 5.18: Expression of PcCir2Dom-GFP. CBB stained Nu-PAGE showing the expression of pEHISGFPTEV-*pccir2dom* under different conditions. The cells from lanes 1 – 4 were incubated at room temperature with rotary shaking of 180 rpm for 3, 6, 8 and 16 h respectively, and lanes 5 – 6 were grown at 37 ^{o}C with 250 rpm and 3 and 6 h expression time, respectively. All were grown in LB broth and induced with 0.5 mM IPTG. The same volume of cell lysate was loaded in each lane. The arrow indicates the position of the PcCir2Dom-GFP fusion protein. Molecular weight markers are indicated on the left hand side.

Figure 5.19, A displays the separated soluble and insoluble fractions from cells expressing the vector pEHISGFPTEV alone and cells expressing pEHISGFPTEV_*pccir2dom*. GFP is observed in the soluble fraction in the sample from pEHISGFPTEV alone (Figure 5.19, A) but not in the soluble fraction of cells expressing pEHISGFPTEV_*pccir2dom* as shown in B.

In each of the conditions trialled no soluble PcCir2 protein was obtained (Figure 5.19, C). In all conditions trialled the GFP tagged PcCir2 domain was well expressed. As the expression levels were high (as judged by the band size on CBB stained gel) it was questioned that if too much protein was produced, was it packaged into IBs. Attempts were therefore made to slow down the growth of the expressing bacteria by limiting the concentration of IPTG to 0.05 mM and by growing the cells in a M9 minimal salts medium (Appendix 1). Figure 5.20 (lanes 1 – 4) demonstrates that M9 media had little to no effect on the solubility of the PcCir2Dom protein.

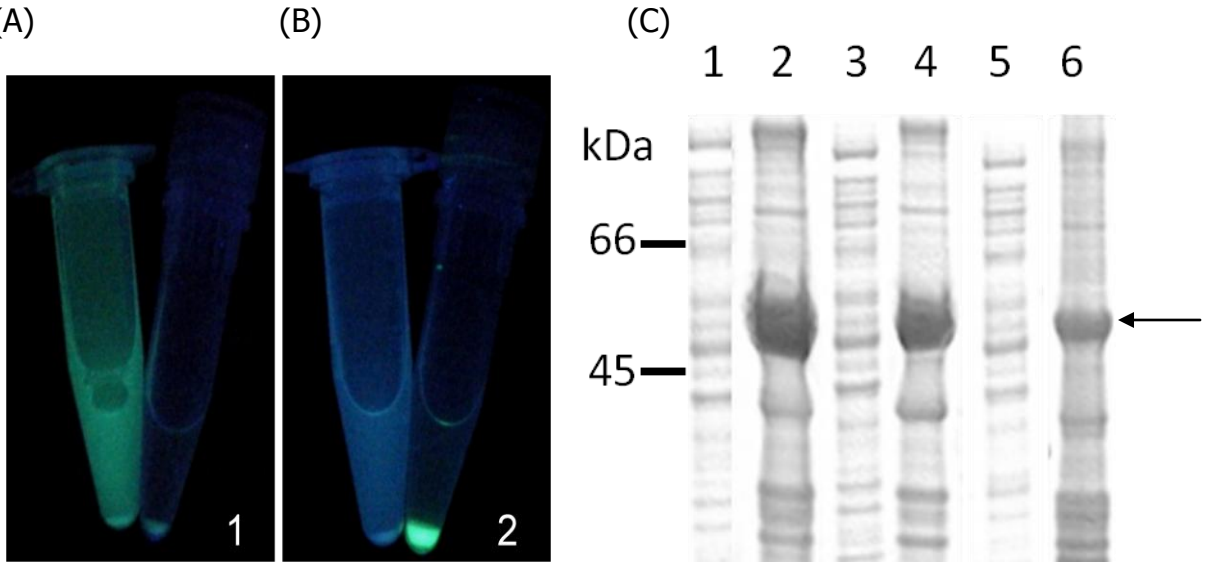


Figure 5.19: Solubility of PcCir2Dom-GFP. (A) The tubes represent the vector alone where the GFP is soluble (B) the tubes are from cells expressing pEHISGFPTEV-*pccir2dom*. The tube on the left hand side of each image is the soluble fraction, and the right hand side tubes are the insoluble fractions. (C) a CBB stained Nu-PAGE gel containing the soluble (lanes 1, 3 and 5) and insoluble (lanes 2, 4 and 6) fractions of cells expressing pEHISGFPTEV-*pccir2dom*, grown in LB media and induced with 0.5 mM IPTG. Cells were induced for 16 h (lanes 1 and 2) and 8 h (lanes 3 and 4) at 22 ^{o}C and for 8 h at 37 ^{o}C (lanes 5 and 6). The arrow shows the PcCir2Dom-GFP fusion protein. Equal volumes of the soluble and insoluble fractions were applied to the gel. Molecular weight markers are indicated on the left hand side.

Arginine and urea have been used to aid in the solubilisation and re-folding of proteins from IBs. As it would be impossible to ensure the PcCir2 domain would be in the correct conformation, it was decided against attempting to re-fold it, however, in the future it might be useful to attempt this in the efforts to obtain a pure protein sample. Arginine was added to M9 minimal media at a concentration of 1 M to analyse if it might improve the solubility of protein produced from either pEHISGFPTEV-*pccir2dom* or pEHISTEV-*pccir2dom*. Again no soluble protein was detected after separation of soluble and insoluble proteins on a NuPAGE gel (Figure 5.20 lanes 5 – 8).

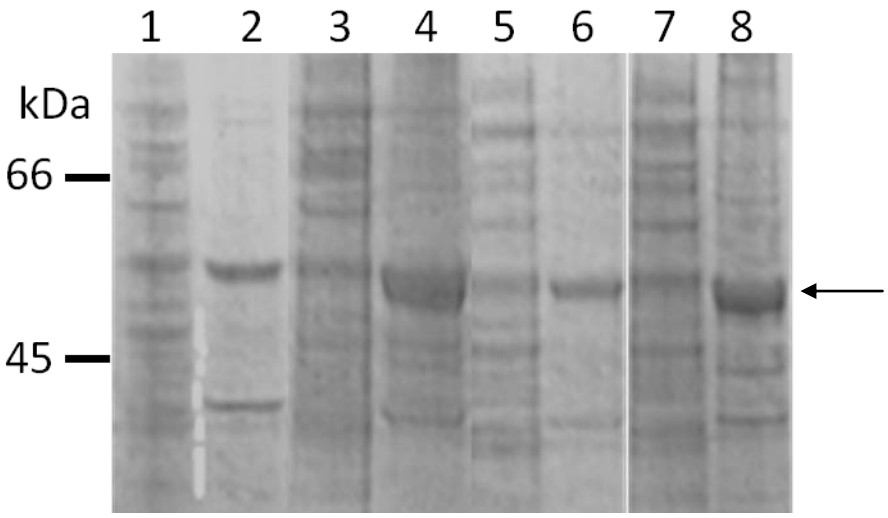


Figure 5.20: Solubility of PcCir2Dom-GFP after addition of arginine to the growth medium. CBB stained NuPAGE showing the soluble (lanes 1, 3, 5 and 7) and insoluble (lanes 2, 4, 6 and 8) fractions of cells expressing pEHISGFPTEV-*pccir2dom* and grown in M9 minimal salts media (all lanes) with supplemented 1 M arginine (lanes 5 – 8). Cells were induced for 6 h (lanes 1, 2, 5 and 6) and 16 h (lanes 3, 4, 6 and 7) at 22 ^{o}C. The arrow shows the PcCir2Dom protein. Molecular weight markers are indicated on the left hand side.

The same expression vector lacking the GFP gene, pEHISTEV-*pccir2dom*, was also used as it was thought that the N-terminal GFP extension could have been causing problems with the solubility of the protein. The pEHISTEV-*pccir2dom* vector was transformed into BL21 (DE3) cells. Protein expression was induced using 0.1 and 0.05 mM IPTG. Cells were grown for a further 3, 6 and 8 h at either 22 ^{o}C or 37 ^{o}C. Cell lysate from each of the growth conditions was applied to a NuPAGE gel, and protein separated by electrophoresis. In each of the conditions tested protein corresponding to the correct molecular weight (24.6 kDa) was observed (Figure 5.21, A). Total

soluble and insoluble protein was extracted from each of the conditions tested and was separated on a NuPAGE gel and stained with CBB. The image in Figure 5.21, B demonstrates that PcCir2Dom is observed in the insoluble fractions (lanes 2, 4, and 6). In Figure 5.21 B, there is a well expressed band towards the top of the gel in the insoluble fractions, the identity of this protein was never analysed as it was thought to be an *E. coli* protein or an aggregate of the PcCir2 protein.

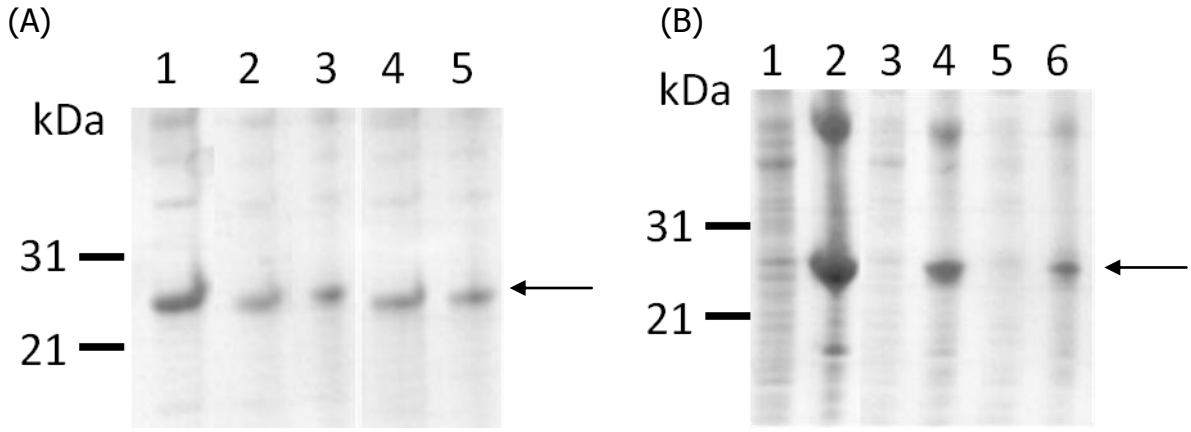


Figure 5.21: Expression and solubility of PcCir2Dom in pEHISTEV. (A) Cell lysate from BL21 (DE3) cells applied to a NuPAGE gel and stained with CBB. Cells in lane 1 and 5 were grown at 22 ^{o}C, cells in lanes 2 – 4 were grown at 37 ^{o}C after induction with 0.1 mM (lanes 1 and 4) and 0.05 mM (lanes 2, 3 and 5) IPTG. Cells were grown for 16 h (lane 1), 3 h (lane 2), 6 h (lane 3) and 8 h (lanes 4 and 5) after induction. (B) Soluble (lanes 1, 3 and 5) and insoluble (lanes 2, 4 and 6) fractions separated from cells expressing the PcCir2Dom protein. Lanes 1 and 2 were grown for 16 h at 22 ^{o}C after induction with 0.1 mM IPTG. Lanes 3 – 6 were grown for 8 h at 37 ^{o}C after induction with 0.05 mM (lanes 3 and 4) and 0.1 mM (lanes 5 and 6) IPTG. The arrows indicate the PcCir2Dom protein. Molecular weight markers are indicated on the left hand side.

5.4.2 Optimisation of PcCir2 Domain expression in pGFPe

DNA encoding the putative soluble domain, *pccir2domain* (*pccir2dom*) was amplified from the *pccir* gene using the primers PcCir GFPe F and PcCir2 210 R (Appendix 2). The 630 bp domain was cloned into the *Kpn*I and *Xho*I restriction endonuclease sites of the plasmid vector pGFPe. This gave rise to a third domain expression construct, pGFPe_ *pccir2dom*. The pGFPe plasmid contains a *gfp* gene situated C-terminally to the cloned gene and a TEV protease site to allow cleavage of the GFP fusion protein. In addition, a His-tag is encoded by the plasmid vector thus the expressed protein will possess a C-terminal 6xHis-tag. Expression of recombinant protein is driven by the

IPTG-inducible T7 promoter. As in Sections 5.3.3 and 5.4.1 above, GFP fluorescence was used as an indicator of correct folding.

Small scale trials were carried out using LB broth as the expression medium at 22 °C. Cultures were induced with 0.05 mM IPTG as has been described previously and 0.4 mM IPTG as described by Rapp *et al* (2004). Samples were taken at various time points (3, 6, and 8 h) and pelleted to assess the samples for GFP fluorescence *a proxy* for protein expression. Figure 5.22, A displays a CBB stained gel containing lysate from cells after induction of protein expression with 0.05 mM IPTG for 3, 6 and 8 h. The gel shows expression of the 54.6 kDa fusion protein, as well as two smaller proteins around 40 kDa. Not much difference is observed in the levels of expression over the different incubation times. Cells from the 3 h and 8 h cultures were separated into the soluble and insoluble fractions applied to a NuPAGE gel. The overexpressed PcCir2Dom was present in the insoluble fraction only (Figure 5.22, B) with no visible proteins in the soluble fraction.

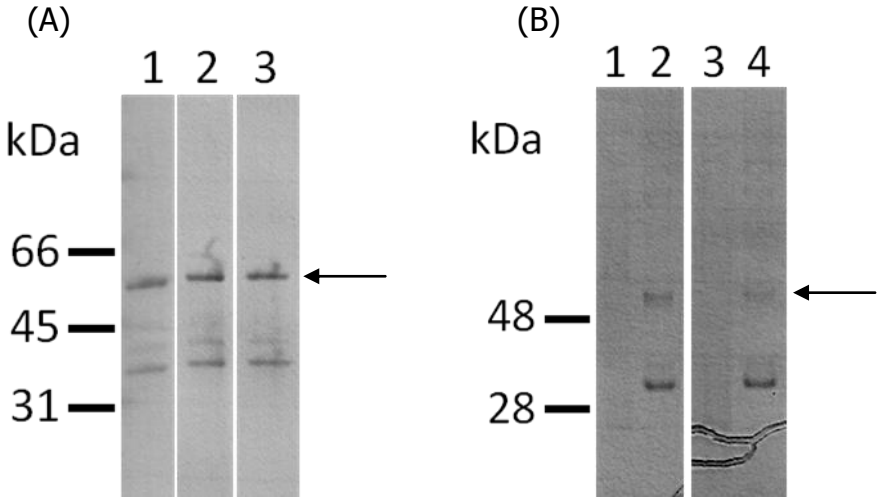


Figure 5.22: Expression and solubility of PcCir2Dom in pGFPe. (A) CBB stained gel of PcCir2Dom-GFPe fusion protein. Cells were grown at 22 °C in LB broth and induced with 0.05 mM IPTG. Samples were taken at 3 h (lane 1), 6 h (lane 2) and 8 h (lane 3). (B) Soluble (lanes 1 and 3) and insoluble (lanes 2 and 4) protein fractions separated on a NuPAGE gel. Cells were grown as before, lanes 1 and 2 for 3 h after induction and lanes 3 and 4 for 8 h. The arrows indicate the PcCir2Dom protein. Molecular weight markers are indicated on the left hand side.

A potential reason for PcCir2Dom being insoluble would be the presence of additional stretches of hydrophobic amino-acids, for example as in a second trans-membrane domain within PcCir2. However, hydropathy plots failed to identify a second trans-membrane domain in PcCir2 hence the reason for PCir2Dom's insolubility remains uncertain. Also, the RIFIN proteins, related to CIR proteins, are thought to possess two TMDs. Lowering the default cut-off in TMHMM revealed a second potential TMD between residues 50 – 70. Further scrutiny of the hydropathy plot shows a second peak at the N-terminus of the amino acid sequence (circled in Figure 5.23, A), the peak does not reach the lower cut-off value and covers amino acids 50 – 70. It is possible that this hydrophobic region could be an addition TMD or may interact with the membrane.

5.4.3 Optimisation of PcCir2Dom2 expression in pEHISGFPTEV

To examine the possibility of a second hydrophobic domain, a second nucleotide domain encoding residues 73 – 210 was amplified using primers PcCir2Dom2 F and PcCir2Dom R from the *pccir*2 gene sequence and termed PcCir2Dom2 (Appendix 2). A schematic shows the position of PcCir2Dom2 compared to the full-length protein and original PcCir2 domain (Figure 5.23, B).

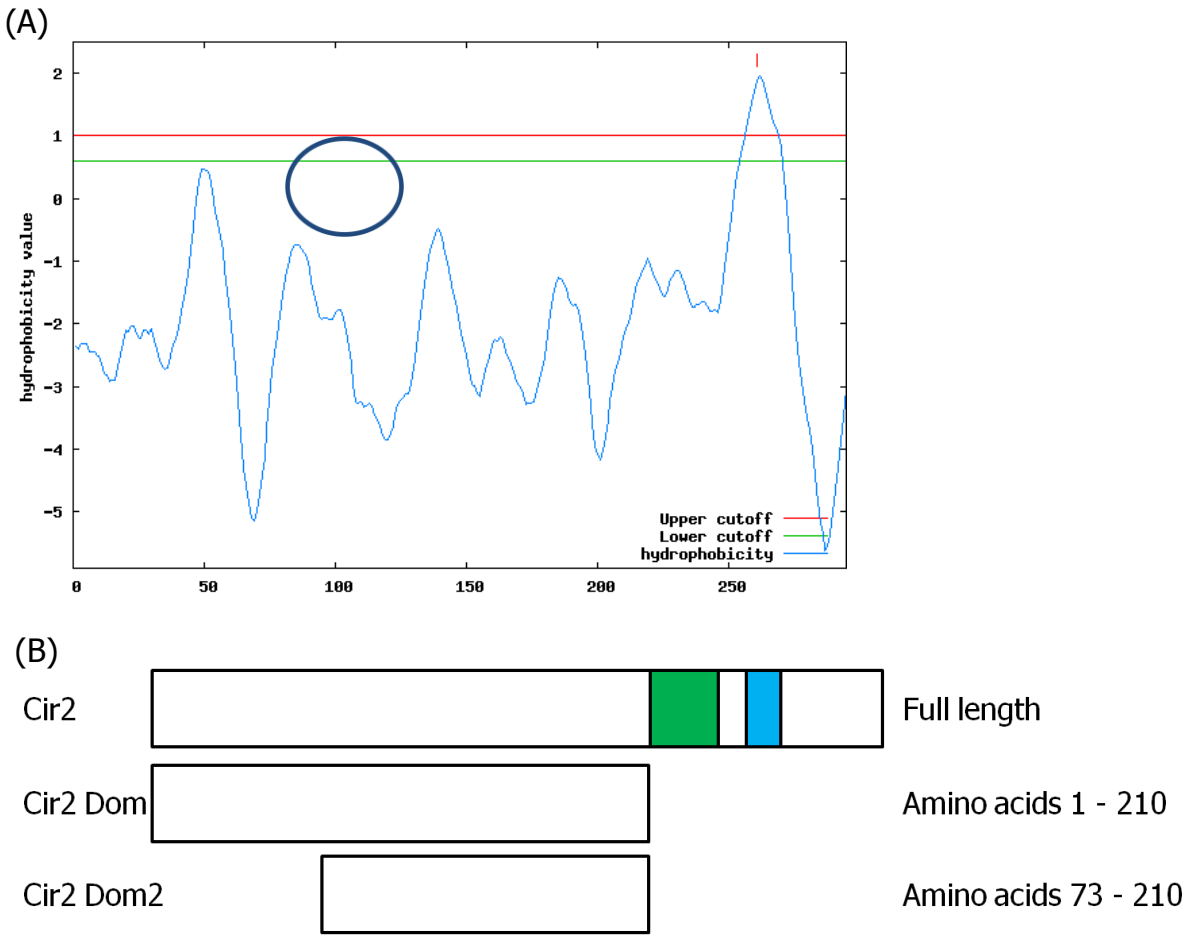


Figure 5.23: PcCir2 protein domain 2. (A) showing the hydropathy plot from Figure 5.1 with hydrophobic domain circled. (B) gives a schematic of the protein and protein domains, full-length PcCir2 (Cir2), PcCir2Dom (Cir2 Dom) and PcCir2Dom2 (Cir2 Dom2). Indicated by the blue block is the TMD, the green block shows the position of the HVR. Diagrams are not to scale.

This putative new domain was amplified from *pccir2*. The sequence encoding PcCir2Dom2 was inserted into pEHISGFPTEV yielding pEHISGFPTEV-*pccir2dom2*. Expression trials were carried out as before using the GFP fluorescence as an indicator of protein expression. Strong fluorescence was observed when the protein was expressed in BL21 (DE3) cells and induced with 0.05 mM IPTG, and grown for different times at 37 °C and at 22 °C. Cell lysate was separated by a NuPage gel, to check that the correctly sized protein was expressed (Figure 5.24, A). The size of the PcCir2Dom2-GFP fusion protein was predicted to be 47.4 kDa. From this it can be seen that PcCir2Dom2-GFP was expressed well, however when cell lysate was separated into the soluble and insoluble fractions, the majority of expressed protein was still present in the insoluble fraction (Figure 5.24, B).

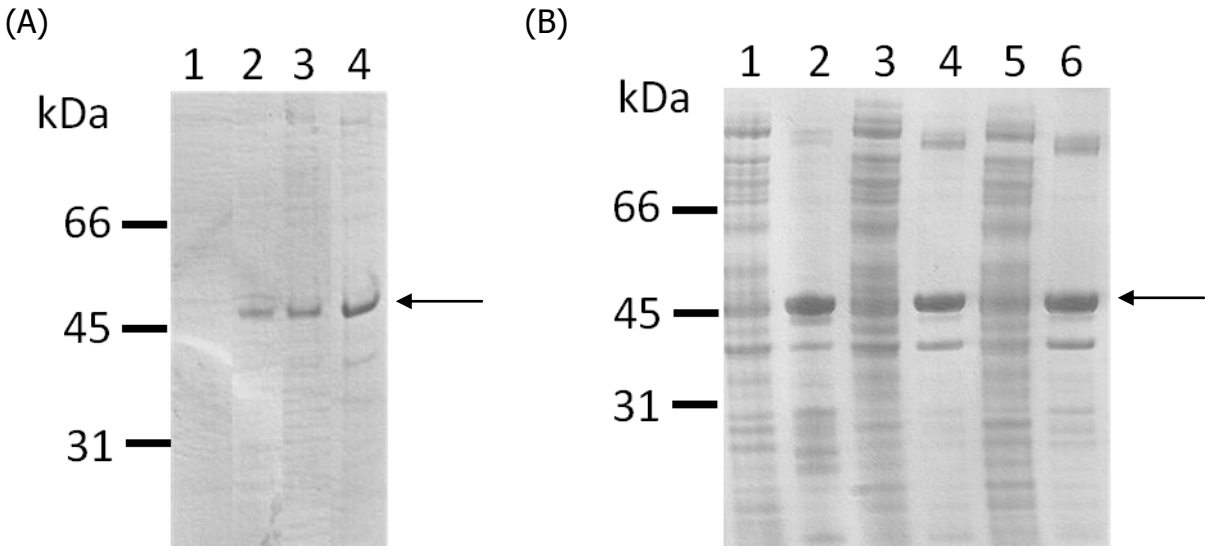


Figure 5.24: Expression of PcCir2Dom2 in BL21 (DE3) cells. (A) CBB stained NuPAGE gel showing the expression of PcCir2Dom2 in BL21 (DE3) cells in cell lysate after induction with 0.05 mM IPTG. Lane 1 shows protein from an uninduced culture. Lane 2 contains protein from a culture induced for 3 h at 37 ^{o}C. Lanes 3 and 4 were induced at 22 ^{o}C for 6 and 8 h, respectively. (B) Cells expressing the PcCir2Dom2 protein, using the same conditions as (A), were separated into the soluble (lanes 1, 3, and 5) and insoluble (lanes 2, 4, and 6) fractions. Lanes 1 and 2 are from a culture h at 37 ^{o}C. Lanes 3 and 4 were induced at 22 ^{o}C for 6, and, lanes 5 and 6 induced at 22 ^{o}C for 8 h. The arrows show the PcCir2Dom2 protein. Molecular weight markers are indicated on the left hand side.

The addition of arginine as described in Section 5.4.1 was tested. Cells were grown in M9 minimal salts media for 16 and 20 h with the addition of arginine. A bacterial expression strain known as Arctic Express™ was also tested as this strain can be grown at temperatures as low as 10 ^{o}C. Arctic Express™ cells co-express two cold-adapted chaperonins from the extremophillic bacterium *Oleispira antarctica* that can process proteins at lower temperature. Cells were thus grown in LB medium at 10 ^{o}C for 20 h as stated in the Arctic Express™ manual and induced with either 0.1 mM or 0.05 mM IPTG. Arctic Express™ expression trials were carried out in LB medium. The PcCir2Dom2-GFP fusion protein was expressed well but remained, for the most part, insoluble. Figure 5.25 shows the results for BL21 (DE3) cells (A) and Arctic Express™ (B) grown with and without arginine, separated into the insoluble and soluble fractions. Two additional well expressed bands can be seen in the soluble fractions, these are the chaperone proteins present in the Arctic Express™ cells, Cpn60 is 57 kDa, and Cpn10 is 10 kDa.

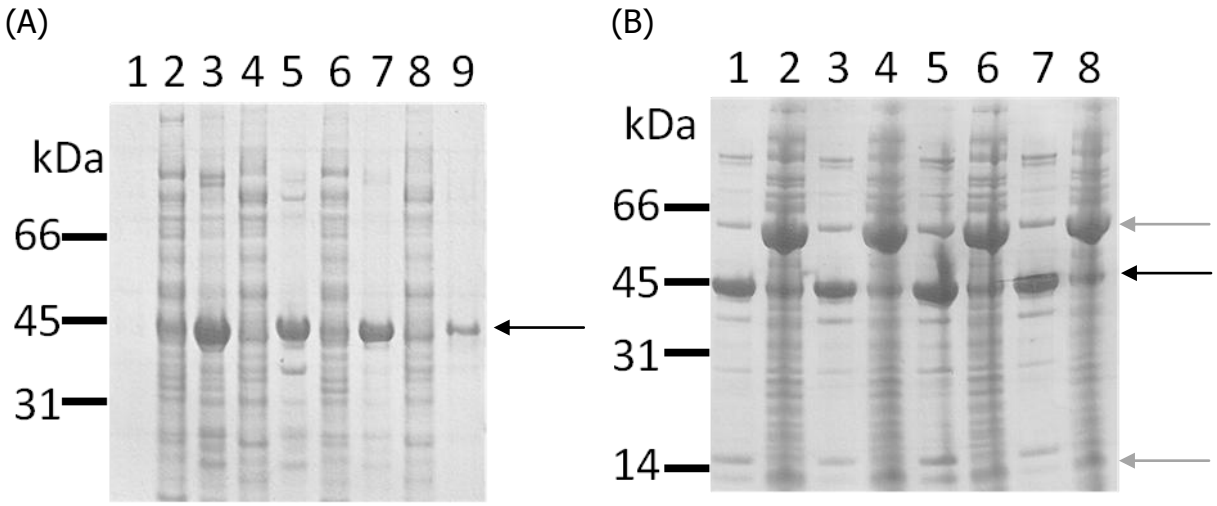


Figure 5.25: Solubility of PcCir2Dom2 with supplemented arginine. (A) CBB stained NuPAGE gel showing the expression of PcCir2Dom2 in BL21 (DE3) cells induced with 0.05 mM IPTG. Cells have been separated into soluble (lanes 2, 4, 6, and 8) and insoluble fractions (lanes 3, 5, 7 and 9), and supplemented with (lanes 2, 3, 6, and 7) and without (lanes 4, 5, 8, and 9) arginine. Lane 1 contains an uninduced control. Lanes 2 – 5 have been grown for 20 h after induction and lanes 6 – 9 grown for 16 h after induction. (B) CBB stained NuPAGE gel showing the expression of PcCir2Dom2 in Arctic Express™ cells. The cells have been separated into the soluble (lanes 2, 4, 6, and 8) and insoluble (lanes 1, 3, 5 and 7) fractions and have been treated with (lanes 1, 2, 5, and 6) and without (lanes 3, 4, 7, and 8) arginine. Lane 1 – 4 show protein induced with 0.1 mM IPTG and lanes 5 – 8 induced with 0.05 mM IPTG. The black arrows show the PcCir2Dom2 protein. The upper and lower bands (grey arrows) in the soluble fractions indicate the Cpn60 and Cpn10 proteins which are co-expressed in the Arctic Express™ cells. Molecular weight markers are indicated on the left hand side.

Possibly the protein domain is not a particularly soluble protein. As the protein was missing the N-terminus it could be that it was unable to fold correctly, and requires the N-terminus as a guide to its native state. If this is correct, the protein would not have been produced in the proper state and as a result may have been deposited in IBs. There is nothing known about the protein and its possible function, thus purification from IBs was not of practical value as there is no way on ensuring the protein would be re-folded in the correct orientation. Therefore it was decided not to pursue the strategy of producing PcCir2 protein domains for structural studies including crystallography.

5.4.4 Data collected by collaborators

The following work was carried by Prof. A. Hawkins and collaborators at Newcastle University. The results are presented here to complete the data set on the analysis of the expression of the PcCir2 domain.

The *pccir2dom* DNA sequence was cloned onto the plasmid pMUT279, and the protein expressed in *E. coli* BL21 AI cells. The cells were grown at 25 ^{o}C for 9 ½ h, the temperature reduced to 15 ^{o}C. The culture was induced with 0.2 % arabinose and the cells incubated for a further 17 h. Cells were harvested (50 g wet weight) and the protein purified using a Proband IMAC column using imidazole. The eluted protein was dialysed using 50 mM potassium phosphate (pH 7.2), 1 mM dithiothreitol (DTT) and further purified by MONO Q FPLC ion exchange chromatography using a sodium chloride gradient. Fractions from the ion exchange chromatography were pooled and a sample applied to a 12 % SDS-PAGE gel to analyse purity. The sample was concentrated to 0.27 mg/ml in 0.5 ml using a 10 kDa MWCO Vivaspin centrifugation device.

Dynamic light scattering (DLS) was carried out using the concentrated protein solution. DLS showed the protein preparation was monodisperse. Further to this, the secondary structure was analysed by far ultra violet circular dichroism (UV-CD). The conclusions are that the protein has a mixed structure of α-helix and ß-sheet components, and the overall spectrum is consistent with that of a folder protein. A more detailed analysis of secondary structure could not be carried out due to salts in the buffer.

This work shows the small percentage of soluble PcCir2Dom protein present could be extracted and used for structural analysis. However, the concentration of protein produced was too low for further investigation.

5.5 PcCir2 Membrane Topology

In the investigation of membrane protein topology it is insufficient to employ *in silico* prediction methods alone, other approaches including biochemical, should also be taken. The results presented previously, suggested that the first PcCir2 domain could be associated with the membrane. This raises the possibility of a membrane

associated domain or second TMD present in the PcCir2 protein. This section reports the biochemical characterisation of the secondary structure of PcCir2.

The membrane topology of PcCir2 is important as it contains a HVR which has been proposed as the domain exposed to the immune system and hence the hypervariability means expression of different variants allows immune avoidance. An understanding of PcCir2 membrane topology, especially the positioning of HVRs, is important to the elucidation and understanding of the function/s of this gene family.

5.5.1 PhoA and GFP protein fusions as a basis for topology modelling

PcCir2 fusions to GFP or PhoA were made at various points along the amino acid chain of PcCir2, using standard vectors which fuse the reporter gene to the C-terminus of the protein domain under study, according to the method described by von Heijne and colleagues (Rapp *et al*, 2004, Daley *et al*, 2005). The use of C-terminal fusions has proven more accurate in the determination of protein topology using *E. coli* (Rapp *et al*, 2004). Six gene truncations were prepared for the membrane topology analysis, shown schematically in Figure 5.26, A. The full-length PcCir2 protein schematic indicates the locations of the TMD, hydrophobic region and the position of the HVR. Each of these truncations was cloned into the pHA-1 vector encoding a C-terminal *phoa* gene, and a pGFPe vector encoding a C-terminal *gfp* gene. An example plasmid from each is shown in Figure 5.26, B and C containing the full-length *pccir*2 gene. Figure 5.26, D shows the cut vectors and five of the gene fragments (the full-length *pccir*2^{312} has been omitted as it is shown in Figure 5.9).

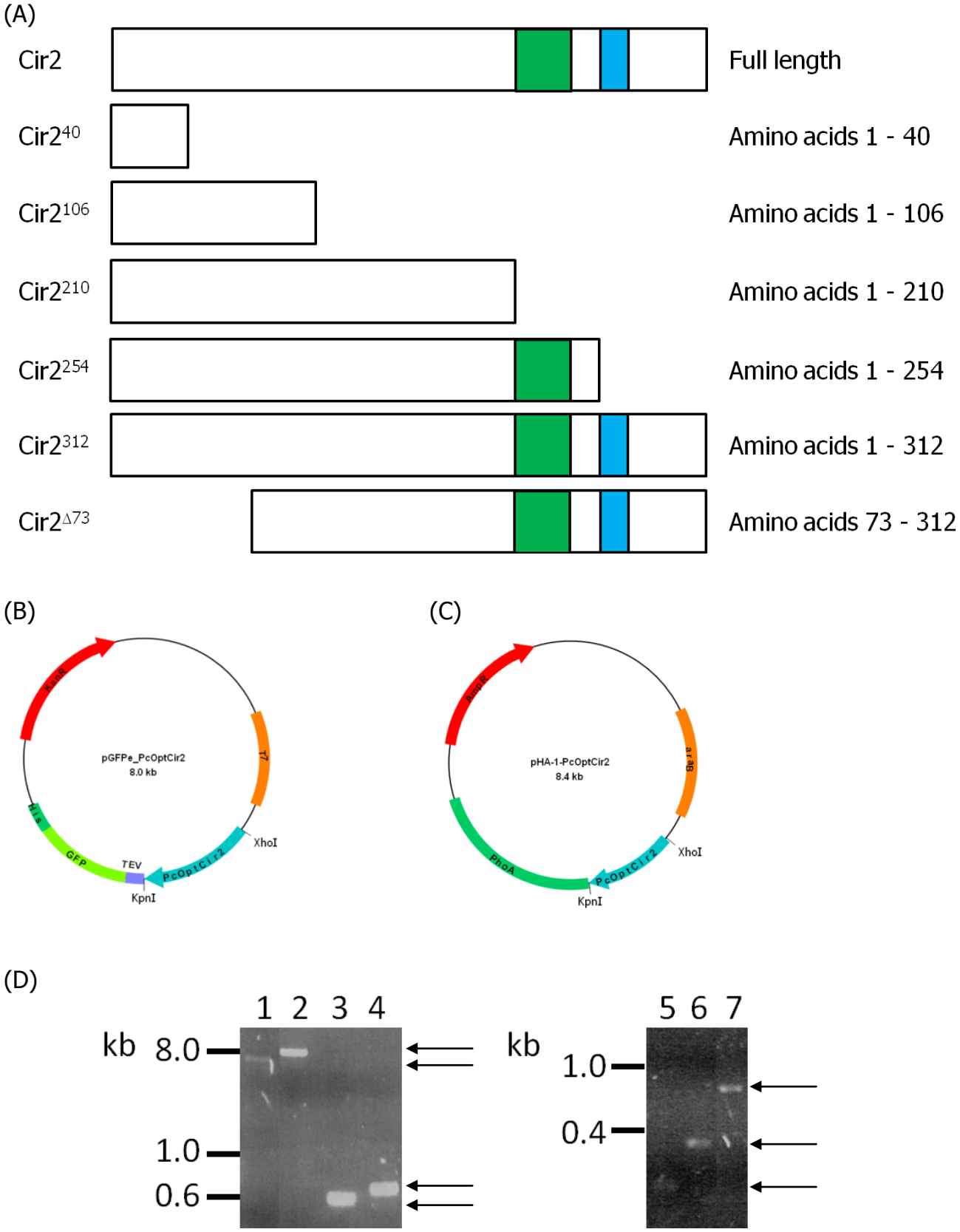


Figure 5.26: Cloning of truncated proteins. (A) Schematic of the truncated proteins, PcCir2, full-length PcCir2 with blue and green boxes denoting the TMD and HVR, respectively. Fusions of GFP (green

fluorescent protein) or PhoA (alkaline phosphatase) were made to the C-terminal of each truncation. This produced one set of proteins with a GFP fusion, and one set with a PhoA fusion protein. PcCir2^{40} 4.1 kDa, PcCir2^{106} 12.2 kDa, PcCir2^{210} 24.6 kDa, PcCir2^{254} 59.3 kDa, PcCir2^{312} 66.3 kDa, PcCir2$^{\Delta 73}$ 58.2 kDa. (B) Example plasmid showing the *cir*2 gene upstream of the PhoA reporter. (C) Example plasmid showing the *cir*2 gene upstream of the *gfp* reporter gene, TEV protease site and 8xHis tag. Diagrams are not to scale. (D) The 630 bp (*pccir*2^{210} lane 3), 708 bp (*pccir*2 $^{\Delta 73}$ lane 4), 120 bp (*pccir*2^{40} lane 5), 318 bp (*pccir*2^{106} lane 6) and 762 bp (*pccir*2^{254} lane 7) *pccir*2 gene truncations were cloned into the 7000 bp GFPe (lane 1) and pHA-1 (lane 2) vectors using the restriction sites *Xho*I and *Kpn*I. Arrows point to the plasmids and inserts.

The strategy behind PhoA/GFP method is the comparison between the activities of two reporter proteins. Each of the fusion proteins is active on a different side of the *E. coli* inner membrane: GFP in the cytoplasm and PhoA in the periplasm. In this regard, the GFP protein is not exported to the periplasm correctly and thus little to no fluorescence should be observed if the fusion is located on the periplasmic side. If the C-terminus of the hybrid protein locates on the cytoplasmic side, high levels of fluorescence should be observed. Conversely, PhoA is not active on the cytoplasmic surface but is activated in the periplasm. PhoA is active only as a dimer. The reducing environment of the *E. coli* cytosol prevents this dimerisation and thus the activity of the protein. Consequently, if high PhoA activity is observed the C-terminus of the hybrid protein is located on the periplasmic surface. This is demonstrated schematically in Figure 5.27.

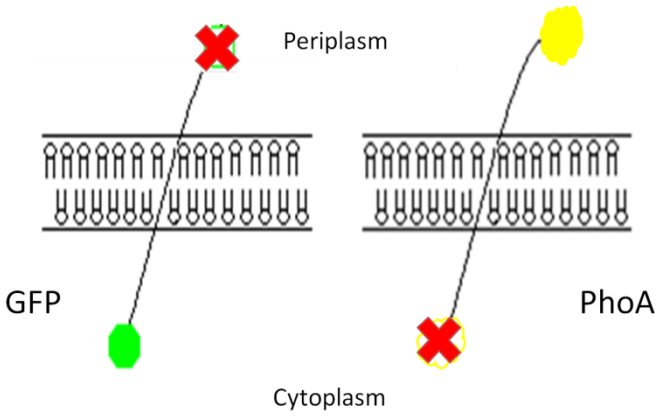


Figure 5.27: Location of fusion protein fluorescence or activity. Schematic of a membrane showing the areas of activity for GFP and PhoA. GFP fluorescence is observed only if the C-terminus of the protein is found in the cytoplasm as the protein is not exported in the correct orientation. PhoA activity is only detected in the periplasm, and thus the C-terminus of the protein must be located in the periplasm.

It is essential to have similar expression levels of fused proteins under examination to obtain meaningful results, since the assay depends on quantifying differing levels of GFP fluorescence and PhoA activity. Thus, expression trials were carried out for each of the six fusions to ensure that (i) each bacterial colony expresses the same fusion protein and at a similar level, and (ii) expression levels between fusions are comparable.

5.5.1.1 Expression of GFP constructs

The full-length gene and sequences encoding the six protein fragments were inserted into the pGFPe plasmid upstream from a *gfp* gene creating six constructs with a C-terminal GFP fusion Figure 5.26, B. Experimental trials were carried out using *E. coli* BL21 (DE3) cells and induced with IPTG as the gene is under the control of the T7 promoter. IPTG concentrations of 0.1 and 0.05 mM were tested, as well as two temperature conditions (22 ^{o}C or 37 ^{o}C) and incubation time after IPTG induction (3 h, 6 h, 8 h and 16 h). Crude membranes were prepared from each culture and 1 µg protein ran on a NuPage gel, and western blotted using an HRP conjugated anti-GFP antibody. Figure 5.28, A shows the resulting western blot containing all 6 fusion proteins. Lane 1 contains PcCir2^{40}-GFP which has an estimated molecular weight of 34.1 kDa, lane 2 PcCir2^{106}-GFP with estimated MW 42.2 kDa, lane 3 PcCir2^{210}-GFP with estimated MW 54.6 kDa, lane 4 PcCir2^{254}-GFP with estimated MW 59.3 kDa, lane 5 PcCir2^{312}-GFP with estimated MW 66.3 kDa and lane 6 PcCir2$^{\Delta 73}$-GFP with estimated MW 58.2 kDa. The protein domains not containing the TMD, PcCir2^{40}-GFP, PcCir2^{106}-GFP and PcCir2^{210}-GFP, all show a protein band size corresponding to the expected molecular weight. The two domains containing the TMD, PcCir2^{312}-GFP and PcCir2$^{\Delta 73}$-GFP and fusion PcCir2^{254}-GFP not containing the TMD ran further on the gel than expected, however this observation is sometimes noted with membrane proteins. Two gene fusion constructs PcCir2^{312}-GFP and PcCir2^{254}-GFP gave a multiple banding pattern and protein sequencing confirmed that the lower band of PcCir2^{312}-GFP was the correct protein. It should be noted that the amount of protein detected (by way of band size) in sample PcCir2^{254}-GFP was slightly less than the five others. Figure 5.28 B shows streaked out colonies from a single colony for each fusion. The *gfp* gene product fluoresces green under UV light. After storage at 4 ^{o}C for approximately one week the bacteria cells exhibited a bright green fluorescence for construct PcCir2^{210}-GFP with a paler green for constructs PcCir2^{40}-GFP and PcCir2^{254}-GFP. The paler green

observed for the PcCir2^{254}-GFP colony could be due to the lack of expression as observed in Figure 5.28, however, the sample still shows green fluorescence and was not further optimised. PcCir2^{106}-GFP, PcCir2^{312}-GFP and PcCir2$^{\Delta 73}$-GFP constructs remained non-fluorescent. This result implies that whereas PcCir2^{312}-GFP and PcCir2$^{\Delta 73}$-GFP fusions are periplasmic due to the lack of GFP fluorescence, the C-termini of fusions PcCir2^{40}-GFP, PcCir2^{210}-GFP, and, PcCir2^{254}-GFP are cytoplasmic.

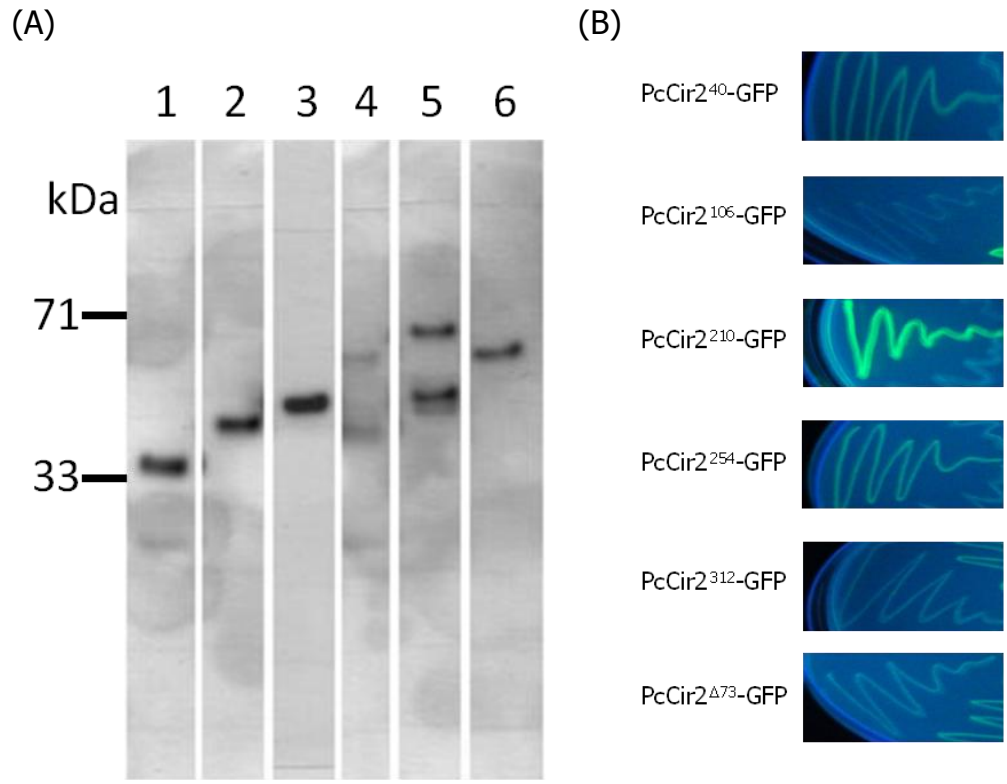


Figure 5.28: Expression and analysis of PcCir2 truncated proteins fused to GFP. (A) Bacterial cells were grown under assay conditions and induced with IPTG. Cell lysate was prepared and applied to a gel and western blotted using an anti-GFP antibody. Lane 1 PcCir2^{40}, lane 2 PcCir2^{106}, lane 3 PcCir2^{210}, lane 4 PcCir2^{254}, lane 5 PcCir2^{312}, and lane 6 PcCir2$^{\Delta 73}$. (B) Furthermore, live bacteria were analysed for reporter gene fluorescence. LB agar was supplemented with IPTG and exposed to UV trans-illumination to reveal GFP fluorescence.

The optimal protein expression conditions selected were found to be 0.05 mM IPTG inducer coupled with bacterial incubation for 3 h at 22 ^{o}C with agitation. Darker bands were observed on the gels for longer induction periods such as 8 and 16 h suggesting a greater concentration of protein had been produced, however variation in expression levels were observed. In this respect, bacterial cells did not display bright GFP

fluorescence when grown at 37 °C, but the fusion protein was expressed well when analysed using PAGE. Despite that the conditions employed did not, *per se*, give the highest concentration of protein expression, GFP production was sufficient to carry out the GFP assays.

5.5.1.2 Expression of PhoA constructs

To optimise the production of PhoA containing chimeric constructs expression trials were carried out. The *phoa* gene fusion constructs are driven by the *araB* promoter and thus the inducer L-arabinose was used to induce expression of the fusion proteins. The bacterial strain used for the PhoA assay, and expression trials are *E. coli* CC118, which lacked the endogenous *phoa* gene. To ensure that cells were producing the PhoA fusion gene, bacteria were plated on LB-agar containing 1 % L-arabinose and 40 µg/ml 5-bromo-4-chloro-3-indolyl phosphate (BCIP). Figure 5.29, A shows a western blot containing the 6 fusion proteins. Lane 1 contains PcCir2^{40}-PhoA which has an estimated molecular weight of 53.2 kDa, lane 2 PcCir2^{106}-PhoA with estimated MW 60.8 kDa, lane 3 PcCir2^{210}-PhoA with estimated MW 73.3 kDa, lane 4 PcCir2^{254}-PhoA with estimated MW 77.9 kDa, lane 5 PcCir2^{312}-PhoA with estimated MW 84.9 kDa and lane 6 PcCir2$^{\Delta 73}$-PhoA with estimated MW 76.9 kDa. The anti-PhoA antibody was tested at several dilutions and the six fusions gave a rapid reaction in the presence of the detection reagents and produced unspecific bands on the western blot. Aliquots were unable to be sent for MS protein sequencing as after staining with CBB the bands were not sufficiently distinguishable to cut from the gel. Samples containing PcCir^{40}-PhoA and PcCir^{210}-PhoA showed very faint bands.

Figure 5.29, B contains streaked-out colonies originating from a single colony for each fusion. The *phoa* gene product, alkaline phosphatase converts BCIP to a compound which has a dark blue colour. After storage at 4 °C for approximately one week, bacterial cells demonstrated a dark blue colour change for constructs PcCir2^{312}-PhoA and PcCir2$^{\Delta 73}$-PhoA. The PcCir2^{106}-PhoA construct indicates a very slight change but remained pale blue. Constructs PcCir2^{40}-PhoA, PcCir2^{210}-PhoA and PcCir2^{254}-PhoA did not give rise to a noticeable colour change. This result suggests that the PcCir2^{312}-PhoA, PcCir2$^{\Delta 73}$-PhoA and possibly PcCir2^{106}-PhoA fusions are periplasmic due to the phosphatase activity and that PcCir2^{40}-PhoA, PcCir2^{210}-PhoA and PcCir2^{254}-PhoA are cytoplasmic.

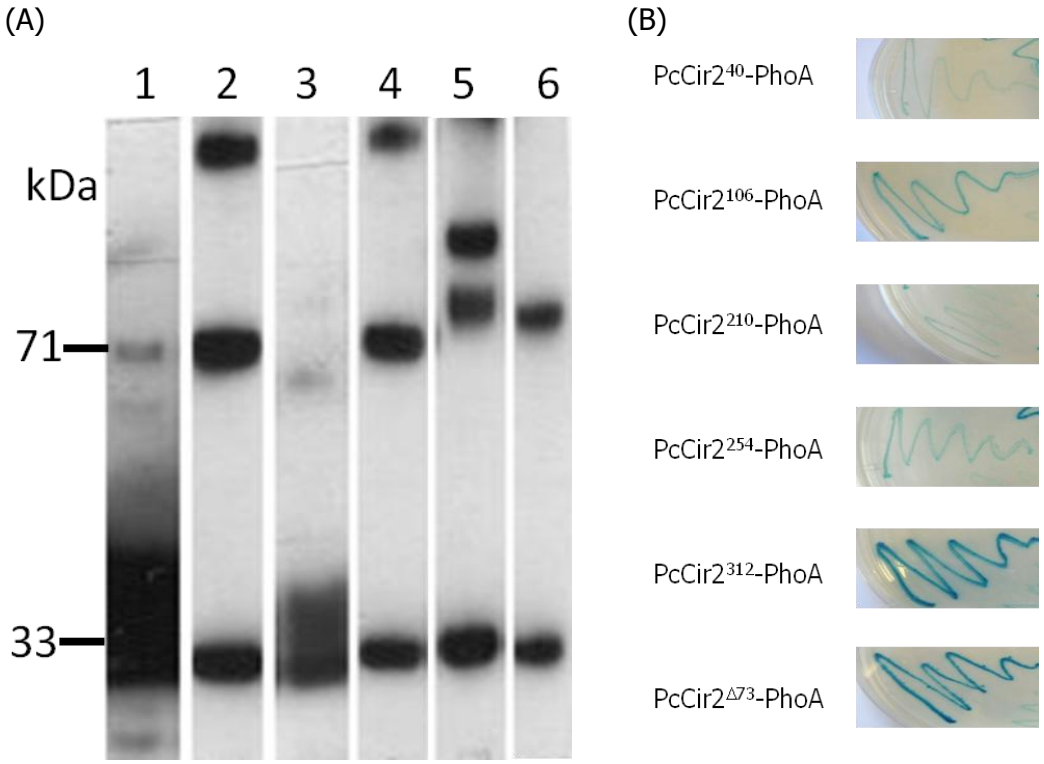


Figure 5.29: Expression and analysis of PcCir2 proteins fused to PhoA. (A) Bacterial cells were grown under assay conditions and induced with L-arabinose. Cell lysate was prepared and applied to a gel and western blotted using an anti-PhoA antibody. Lane 1 PcCir2^{40}, lane 2 PcCir2^{106}, lane 3 PcCir2^{210}, lane 4 PcCir2^{254}, lane 5 PcCir2^{312}, and lane 6 PcCir2$^{\Delta 73}$. (B) Furthermore, live bacteria were analysed for reporter gene activity. Bacterial cells growing on LB agar plate supplemented with L-arabinose and BCIP to monitor PhoA production. The BCIP turns blue due to PhoA activity.

CC118 cells were grown at 37 ^{o}C until the exponential stage of growth, which took approximately 2.5 – 3 h. Following this growth phase, the cells were induced with L-arabinose to final concentrations of 0.1 % and 0.2 %. Cells transformed harbouring the PcCir2^{210} fusion were allowed to incubate for a further 2 h and 4 h before harvesting. Protein from induced cultures was electrophoresed on a NuPAGE gel. The gel containing the separated proteins was western blotted using an anti-PhoA antibody directed against bacterial PhoA (Abcam, Cambridge, UK). Very little or no protein was observed in the 2 h culture time, yet the 4 h samples showed good expression of the fusion protein. There was no observable difference on protein expression when comparing the 0.1 % and 0.2 % the concentrations of L-arabinose inducer for the PcCir2^{210}-PhoA fusion protein. Consequently, bacterial cells harbouring each of the six fusion constructs were grown to log phase, and incubated for 4 h after induction with 0.1 % L-arabinose. From the data presented in Figure 5.29, it can be seen that four of

the six fusions proteins were produced at similar levels, while the PcCir2^{40} and PcCir2^{210} fusions which are less well expressed. The conditions for the assay were selected as a 4 h growth period after induction with 0.1 % L-arabinose. Cells were grown throughout at 37 ^{o}C.

5.5.2 PhoA activity and GFP fluorescence assays

5.5.2.1 GFP fluorescence assay

E. coli cells containing plasmids with each of the GFP fusions were grown in triplicate to the exponential phase which took approximately 90 mins. The hybrid genes were induced with 0.05 mM IPTG, and incubated with agitation (180 rpm) for a further 3 h. After 3 h incubation, the OD_{600} were taken, and 1 ml culture was pelleted for GFP assays. Fluorescence of each cell suspension was measured in terms of fluorescence units (FU) and normalised to the OD_{600} of the culture. The normalised value for each sample in the assay was divided by the total mean. The resulting figure was combined with the PhoA data presented in the next section, and entered into the chart below. An average value was taken for each of the three repeats of the five experiments, and entered in Table 5.4. The average value for each of these is also tabulated.

	Experimental Repeats					
Construct	1	2	3	4	5	Average
PcCir2^{40}-GFP	1.62	2.02	1.63	1.56	1.62	1.69
PcCir2^{106}-GFP	0.75	0.56	0.48	0.54	0.53	0.57
PcCir2^{210}-GFP	1.31	1.14	1.23	1.17	1.01	1.17
PcCir2^{254}-GFP	1.48	1.92	1.76	1.85	1.91	1.78
PcCir2^{312}-GFP	0.46	0.42	0.38	0.33	0.35	0.39
PcCir2$^{\Delta 73}$-GFP	0.48	0.43	0.40	0.29	0.39	0.40

Table 5.4: The normalised and mean divided values for each of the 5 replicates of the GFP fusion proteins. The final columns show the average for each sample. 1 – 5 are the repeats for each experiment, with the average value from the 3 expreiments. Raw data can be found in Appendix 7.3.

5.5.2.2 PhoA activity assay

Bacterial cells containing one of each of the plasmids from PhoA fusions were grown, in triplicate, until the exponential phase. Cells were induced with 0.1 % L-arabinose, and incubated for a further 4 h, before the assay was carried out. The calculation given in Section 2.4.3 was used to determine the PhoA activity for each of the samples (15 in total). The value for each of the 15 samples was divided by the total mean to

give the 'normalised data', this was entered into the chart below. An average value was taken for each of the three repeats in the five experiments, and is entered in table 5.5, the average for each of these is also given.

	Experimental Repeats					
Construct	1	2	3	4	5	Average
PcCir2^{40}-PhoA	0.45	0.54	0.59	0.45	0.36	0.48
PcCir2^{106}-PhoA	0.60	0.24	0.55	0.60	0.82	0.56
PcCir2^{210}-PhoA	0.70	0.41	0.68	0.34	0.42	0.51
PcCir2^{254}-PhoA	0.53	0.32	0.52	0.58	0.70	0.53
PcCir2^{312}-PhoA	2.36	2.64	2.71	2.65	2.77	2.63
PcCir2$^{\Delta 73}$-PhoA	1.13	0.94	1.37	1.31	1.94	1.34

Table 5.5: The normalised and mean divided values for each of the 5 replicates of the PhoA fusion proteins. The final columns show the average for each sample. 1 – 5 are the repeats for each experiment, with the average value from the 3 expreiments. Raw data can be found in Appendix 7.4.

The average PhoA activity in cells with the PcCir2^{312} construct, 2.63 units, is higher than that of the PcCir2$^{\Delta 73}$ construct, 1.34 units. The values should be fairly similar as the fusion is in the same position. Possibly the truncation of the protein has affected the folding and insertion of the PcCir2$^{\Delta 73}$ fusion protein.

5.5.3 Combining the data

The data were normalised by taking the mean of the total data set for GFP or PhoA and dividing each data point by the overall mean. This provided values for each set of experiments that were easily comparable. The values for each of the five replicates of the GFP and PhoA assays were plotted against each other and the chart is presented in Figure 5.30. Where the values for each truncated protein cluster towards the y-axis, the samples had a high GFP fluorescence and low PhoA activity, this suggests the C-terminus of the truncated protein is cytoplasmic. Where the values cluster towards the x-axis, the samples had a low GFP fluorescence and high PhoA activity, this suggests the C-terminus of the truncated protein is periplasmic.

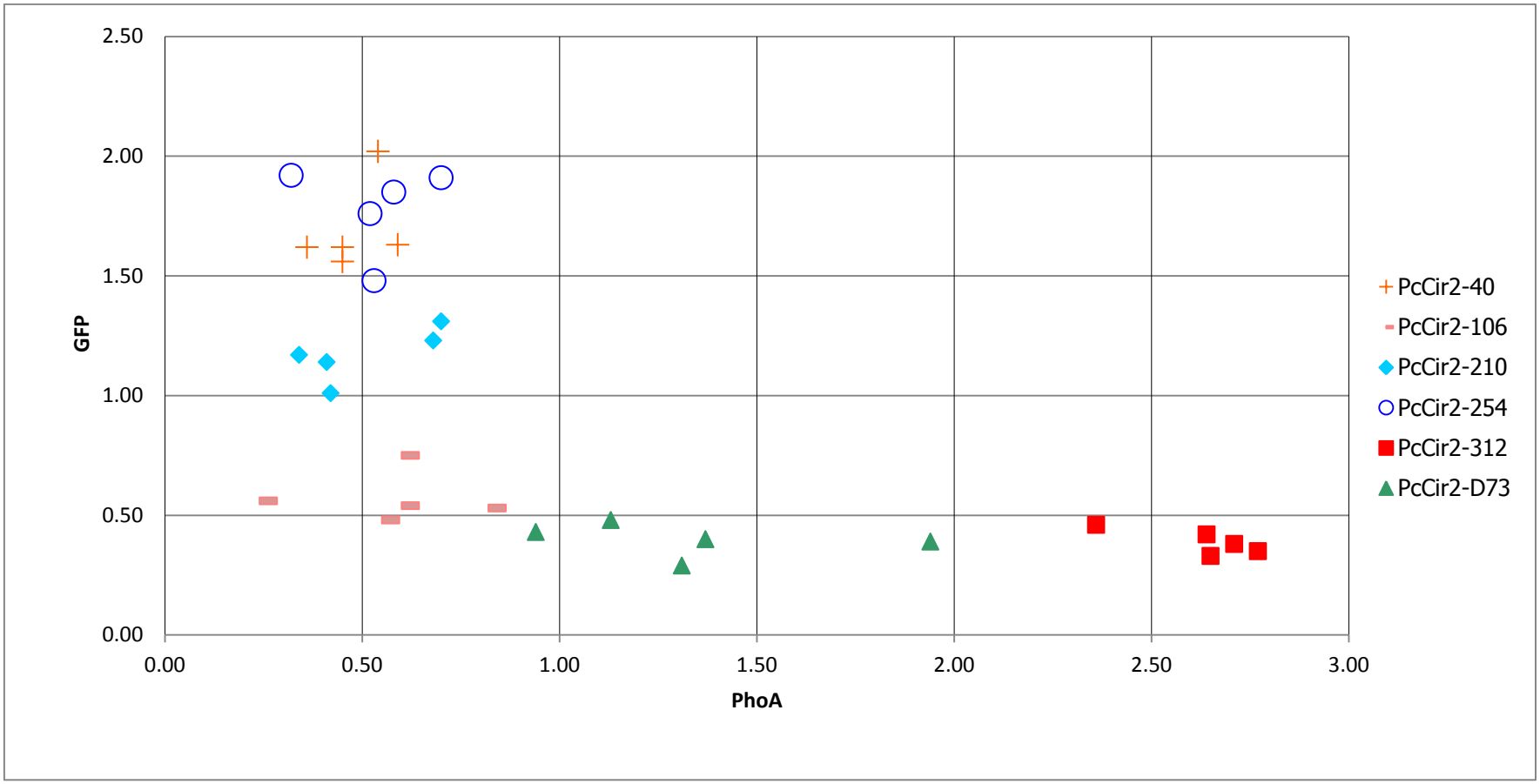


Figure 5.30: PhoA activity vs. GFP fluorescence in truncated fusion proteins. The GFP fluorescent units (calculated from the means) have been plotted against the normalised PhoA activity units (calculated from the means) and entered into the chart. The sway to the left hand side illustrates high PhoA activity and low GFP activity and implying a cytoplasmic C-terminal. The sway to the right hand side shows high GFP activity and low PhoA activity and suggests a periplasmic C-terminal. The PcCir2^{40}, PcCir2^{210} and PcCir2^{254} domains are present on the left hand side. This suggests their C-termini are cytoplasmic. Raw data can be found in Appendix 7.3 and 7.4.

From Figure 5.30, the PcCir2^{210} data is present on the left hand side of the chart demonstrating low PhoA activity and high GFP activity. Conversely, PcCir2^{312} and PcCir2$^{\Delta 73}$ display high-moderate PhoA activity and low GFP activity. The results of PcCir2^{106} do not show a sway to any particular side of the chart. Given that the PcCir2^{106} constructs gave low activities for both GFP and PhoA but was protein detectable by western analysis, subcellular localisation was investigated. Inclusion bodies were therefore isolated and washed to remove any residual membranes. The resultant pellets from the PhoA and GFP constructs were resuspended in buffer and applied to a NuPAGE gel and western blotted (Figure 5.31). The PhoA and GFP constructs from PcCir2^{312} were included as controls as they are present in the membrane, this was to ensure that the membranes had been thoroughly separated from the IBs.

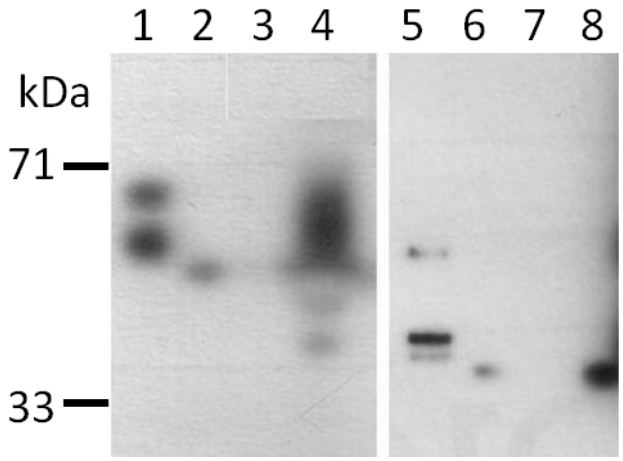


Figure 5.31: Inclusion body separation. Western blot to confirming the presence of the PcCir2^{106} fusion proteins in the IBs, and the absence of PcCir2^{312} from the IBs. Lanes 1, 3, 5 and 7 are the PcCir2^{312} constructs and lanes 3, 4, 6 and 8 are the PcCir2^{106} constructs. Lanes 1,2, 5 and 6 are cell lysate, and, lanes 3, 4, 7 and 8 are IB preparations. In lanes 1 – 4 the blot has been carried out with the anti-PhoA antibody and in lanes 5 – 8 the anti-GFP antibody.

The combined results of the GFP and PhoA fusion protein assays provide a novel model for the topology of PcCir2, one that places the HVR and the C-terminus outside the cell, with the N-terminus on the inside. The *in silico* prediction methods have flaws in predicting the orientation of single TMD proteins (Dr. D. Daley, personal communication). Thus, TMHMM fix[34] (Melén *et al*, 2003) was used to constrict the C-terminus to the outside, and residue 210 (where the fusion was made) to the inside.

[34] http://www.sbc.su.se/~melen/TMHMMfix/

This produced a model of the protein that fitted to the experimental data and gave a good Reliability score of 0.72 (a score of 0.00 suggests there may be other topologies as likely as the suggested model[35] (Melén *et al*, 2003)) and an expected accuracy score of 80 %. The Expected Accuracy Score is proportional to the Reliability score and estimates how probable it is that the suggested topology is correct (taken from[35]).

The experimentally determined topology reverses the predicted topology. The N-terminus, and therefore the hypervariable region, is intracellular. There are many possible reasons for this. One is the heterologous expression system. Another may be the trafficking of the protein, perhaps the *E. coli* membrane is treated as a transport vesicle by the protein or the ER. It could also be that this particular protein has an intracellular location. These possibilities will be discussed fully in Chapter 6.

[35] http://www.sbc.su.se/~melen/TMHMMfix/

5.6 Summary

- The PcCir2 protein belongs to the PIR protein family, possibly also including the RIFIN and STEVOR proteins of *P. falciparum*. PcCir2 contains two conserved sequence motifs shared by the PIR family. PcCir2 is predicted to contain a single transmembrane domain towards the C-terminus of the protein. The N-terminus of the protein is predicted to be extracellular. The protein family are thought to be variant surface antigens. PcCir2 contains an area thought to be hypervariable.

- The *pccir2* coding sequence was redesigned to an *E. coli* codon bias to facilitate expression of the protein in the bacteria. The synthetic codon-optimised gene was cloned into four vectors for protein expression trials, pTTQ18, pTTQ18-TEV, pTTQ10H and pGFPe. In each of the vectors and under each of the conditions tested good protein expression was observed on Coomassie Brilliant Blue stained NuPAGE gels. All vectors encode a C-terminal His-tag, and, pTTQ18-TEV and pGFPe also encode TEV protease sites.

- The PcCir2 protein was only solubilised using the harsh zwitterionic detergents Fos Choline-12 and Fos Choline-14. From expression using the plasmid pTTQ18 all the PcCir2 protein produced was solubilised, for expression from pTTQ18-TEV and pGFPe approximately 50 % of the protein was solubilised. Protein solubilised by Fos Choline-12 was taken forwards for purification. Purification was carried out using a cobalt metal affinity column to bind the protein via the His-tag. When the purified product was tested on a NuPAGE gel various contaminants were observed.

- A putative ordered domain of PcCir2 was found. This domain, PcCir2Dom was expressed from three vectors, pEHISGFPTEV, pEHISTEV and pGFPe. Two contain N-terminal His-tags (pEHISGFPTEV and pEHISTEV) and pGFPe a C-terminal His-tag. All vectors encode TEV protease sites. In each of the vectors and under each of the conditions tested good protein expression was observed on Coomassie Brilliant Blue stained NuPAGE gels. However, when separated into soluble and insoluble fractions, most of the PcCir2Dom protein was present in the insoluble fractions. A second domain missing the first 72 N-terminal amino acids was then created. This second domain, PcCir2Dom2, is lacking a hydrophobic area that may have prevented the

solubilisation of the overexpressed PcCir2Dom protein. This new domain was expressed from the pEHISGFPTEV vector only. The PcCir2Dom protein was present in the insoluble *E. coli* protein fractions.

- The membrane topology of the PcCir2 protein was assessed in *E. coli* using C-terminal GFP and PhoA fusions. The full-length protein was split into six truncations, and GFP or PhoA fused to each on the C-termini. An assay system was used to assess the GFP fluorescence and PhoA activity from each of the truncations. The results suggested a topology where the N-terminus of PcCir2 is intracellular and the C-terminus extracellular.

6 Discussion

The overall aim of the research presented in this thesis was to functionally, biochemically and structurally characterise two *Plasmodium* membrane proteins to gain insights to the mechanisms involved and in so doing develop armaments against malaria. The discussion is split conveniently into two parts, (i) to describe PfNitA (Section 6.1) and (ii) to cover PcCir2 (Section 6.2).

6.1 PfNitA

All *Plasmodium* parasites for which the genome has been sequenced encode a putative formate-nitrite transporter gene, that we have termed *pfnita*. Orthologues are found in other apicomplexan parasites, prokaryotes, fungi and algae (Jia *et al*, 2009, Wang *et al*, 2008, and, Fernandez and Galvan, 2007). In contrast, there are no orthologues in higher organisms such as plants and mammals. Therefore this transporter should *a priori* make a selective drug target.

Transport proteins serve many functions including the facilitation of signals and the passage of nutrients across biological membranes. Transport may be between cellular organelles or from the extracellular medium into cells and *vice versa*. Negatively charged ions are unable to cross the phospholipid bilayer. However, even certain molecules such as water and urea, that are able to diffuse across the bilayer, require transport proteins to assist the process (Darnell *et al*, 2000). Transport proteins make attractive drug targets and there are many currently used chemotherapies that interfere with transporter mediated mechanisms of uptake (Kell, 2012). It was our intention to characterise the PfNitA transporter and thereby lay down the foundations for future research work to be carried out on this potential drug target.

Therefore the questions raised in this thesis were:

(i) What is the function of the formate-nitrite transport protein in *P. falciparum* and what is its subcellular localisation?

(ii) Is the *pfnita* gene and its protein product essential to IDC parasite survival?

6.1.1 Substrate specificity of PfNitA.

The NitA protein of *Plasmodium falciparum* belongs to the formate-nitrite transporter protein family. The family name suggests that the substrates for members of this family are formate, nitrite or both. However, various publications show that FNT proteins transport, in addition to formate and/or nitrite, the hydrosulphide ion, bicarbonate and lactate may also be substrates of the transporter family (Suppmann and Sawers, 1994, Jia *et al*, 2009, Czyzewski and Wang, 2012, Fernandez and Galvan, 2007, and Rycovska *et al*, 2012).

The results presented in Chapter 3 of this thesis show that, when a synthetic codon-optimised *pfnita* gene (*pfnita_opt*) for expression in the lower eukaryote *Aspergillus nidulans* was transformed into this fungus no expression of the *P. falciparum* protein occurred (Figure 3.10). The original rationale was that the *pfnita* gene could be expressed in *A. nidulans* for the purposes of the complementation of a deletion mutant strain lacking the orthologous nitrite transporter encoding gene. Further characterisation of the *P. falciparum* transport protein could be conveniently studied in this experimentally amenable model fungal organism. It was noted that *A. nidulans* does not possess a formate transporter gene from *in silico* analyses and therefore complementation for this function could not be assessed.

In addition to *pfnita_opt*, a further synthetic *pfnita* gene (*pfnita_optEc*) was designed and created for expression in the model prokaryote *Escherichia coli* and functional complementation attempts of null mutants lacking the orthologous proteins, either EcNirC or EcFocA. The two synthetic codon-optimised genes, *pfnita_opt* or *pfnita_optEc*, were transformed into loss-of-function nitrite or formate transporter mutant *E. coli* strains and expressed under anaerobic conditions using specialised medium for later biochemical assays (Figure 3.12). Although the genes were expressed under such optimal conditions the expression levels were lower than the re-expressers (Figures 3.12 and 3.18). In one experiment, however, the expression of *pfnita_opt* was unable to be verified due to technical problems associated with the western blots.

For the functional complementation of an *ecnirc* nitrite transporter null mutant both synthetic codon-optimised genes were expressed followed by protein characterisation. The uptake of nitrite into cells was studied by investigating the depletion of nitrite from media containing the bacteria. The depletion of nitrite from the medium equates to the uptake of nitrite by the bacterial cells. In two positive control strains, a WT and the EcNirC re-expresser, the concentration of nitrite diminished from the medium over a two hour time period (Figure 3.15). In three negative control samples, null mutant, heat killed WT cells, and a blank control omitting cells (i.e. No Cells), the levels of nitrite did not differ significantly over the two hours of incubation (Figure 3.15). From concomitant experiments it was clear that the strains expressing the synthetic codon-optimised *P. falciparum* genes indicated that the PfNitA_optEc protein was not involved in the uptake of nitrite from the medium as the strain performed similarly to the negative controls (Figure 3.15). Moreover statistical analysis using a one-way ANOVA confirmed there were no significant differences between the *pfnita* expressing cells and the three negative controls (Table 3.5). In summary, the results presented in this work shows that the PfNitA permease does not transport nitrite when expressed in *E. coli* under the assay conditions presented in this thesis.

In addition to a nitrite transporter, *E. coli* also possesses a formate transporter, *ecfoca*. The null mutant *ecfoca* strain was used for the functional complementation of this gene, coupled with the formate transport assays. Formate is a toxic by-product of *E. coli* anaerobic glucose metabolism and is therefore excreted from the cell to prevent the acidification of the cell cytoplasm. However, when the pH of the medium drops to pH < 6.8 the function of the permease is altered and formate is imported into the cytoplasm. The transport of formate into the cells was assessed using an indirect assay with the *lacZ* gene fused to the *fdhF* promoter (Falke *et al*, 2009). This promoter is only active upon formate import and the formate transport capability assay is based on the resulting β-galactosidase activity. In short, a high β-galactosidase activity equates to formate influx. In the positive control strain, the EcFocA re-expresser, β-galactosidase activity was found to be markedly increased compared to the null mutant (Figures 3.17 and 3.19). In two negative control strains, null mutant and *pccir2* expressing cells, the levels of β-galactosidase activity did not differ greatly over the four hours of experimentation (Figure 3.19). Unfortunately, the *A. nidulans* codon-optimised gene (*pfnita_opt*) was not assayed as technical problems with the

western blot meant that expression of the protein could not be confirmed. Expression of the *E. coli* codon-optimised protein (PfNitA_optEc) was poor compared to the EcFocA re-expresser (Figure 3.18). The tests strains expressing the synthetic codon-optimised *P. falciparum* gene did not function with regard to the uptake of formate as the β-galactosidase activity was similar to the negative controls (Figure 3.19). Statistical analysis of these strains did not reveal significant differences between the test strains (Table 3.12). This is potentially due to the large standard deviations observed in the samples. To decrease the standard deviations, more assays would have to be carried out. However, the β-galactosidase activity was slightly higher than the negative controls, and given the poor expression of the protein, we speculate that PfNitA can function in the transport of formate. To assess this theory, further research would need to be carried out.

S. cerevisiae is another potential organism that could be used for complementation assays, especially as a knock-out library is available[36]. As this yeast is a eukaryotic organism it may carry out PTM's that are required for the activity of the PfNitA transporter. However, although found in the FNT family (due to a FNT domain[37]) and annotated as a formate-nitrite transporter on GeneDB[38] the protein does not appear to be involved in monocarboxylate transport (Jennings and Cui, 2008 and [39]). It is possible that the protein plays a role in the transport of the chloride ion (Jennings and Cui, 2008). Given the uncertainty of the transporters substrate, and the fact that to examine the substrate/s in *E. coli* certain metabolic pathway had to be induced, it was concluded that it would be too difficult to design an assay for complementation in *S. cerevisiae*.

6.1.2 Subcellular localisation of PfNitA

The subcellular localisation for PfNitA was not predicted from *in silico* analyses. For instance, a mitochondrial or apicoplast location was not indicated by the prediction programs. That no signal peptide was observed was further circumstantial evidence.

[36] http://www-sequence.stanford.edu/group/yeast_deletion_project/deletions3.html. Accessed on 24 January 2013.
[37] http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?INPUT_TYPE=precalc&SEQUENCE=45270094. Accessed on 24 January 2013.
[38] http://old.genedb.org/genedb/Search?name=YHL008C&organism=cerevisiae. Accessed on 24 January 2013.
[39] http://www.yeastgenome.org/cgi-bin/locus.fpl?locus=YHL008c. Accessed on 24 January 2013.

Neither does the amino acid sequence contain a PEXEL motif (Hiller *et al*, 2004, and Marti *et al*, 2004) to direct the protein outside of the parasite cell. Thus, a biochemical investigation into the subcellular localisation of PfNitA was initiated. Parasites were transfected with plasmids containing a GFP tagged or an HA tagged *pfnita*_opt gene (Figure 4.1). Real time live fluorescence microscopy of PfNitA-GFP and immunofluorescence analysis (IFA) of PfNitA-HA showed that the PfNitA transporter was located to the parasite plasma membrane (PPM) and in some instances the digestive vacuole (DV) (Figures 4.4, 4.5, and 4.6). The parasite plasma membrane location of the transporter suggests a role in efflux or influx of metabolites or compounds from the parasites. IFA co-localisation studies presented in this thesis supported the PPM location but not the DV location (Figure 4.7). This result, however, does not completely rule out the DV as a location for the permease. If the over-expressed PfNitA_opt protein is detrimental to the parasite, the protein could be transported to the digestive vacuole for degradation, although a proteosome would be the expected subcellular organelle for protein breakdown. This is, of course, only speculative and would require further analysis. In the yeast *Saccharomyces cerevisiae*, the subcellular localisation has likewise not been determined accurately for the protein orthologue. Nonetheless, studies using GFP tagged proteins suggest the *S. cerevisiae* orthologue is associated with the vacuole in yeast (Paulsen *et al,* 1998, in Jennings and Cui, 2008). No functional role has been reported for the *S. cerevisiae* transporter but it is conceivable that the proteins may have similar vacuolar functions in both *S. cerevisiae* and *P. falciparum*. Overall, the research suggests the PfNitA protein is localised to the PPM and DVM in intraerythrocytic parasites.

It is plausible that the PfNitA protein acts as a monocarboxylate transporter and is responsible for the passage of compounds into or out of the parasite itself. There is indeed a putative monocarboxylate transporter gene present in the *P. falciparum* genome, designated PF3D7_0210300[40]. To the best of my knowledge, no work has been carried out on this gene and/or protein or orthologues in other *Plasmodium* species. Many of the studies performed on *Plasmodium* transporters have been with proteins that are responsible for the passage of nutrients into the parasite from the surrounding medium (e.g. Slavic *et al*, 2010, and el Bissati *et al*, 2008). However, very

[40] http://plasmodb.org/plasmo/. Accessed on 4 September 2012.

few studies have identified proteins that play a role in the efflux of a substrate. An exception, namely the PfCRT protein, is thought to act on the efflux of chloroquine from the parasites DV, although this is not the sole function of the PfCRT protein (Martin *et al*, 2009a).

Nitrites are found in the DV and may be involved in the detoxification of haem (Ostera *et al*, 2008 and Ostera *et al*, 2011). It is possible that because the PfNitA transporter is present in the DVM, it functions in the transport of nitrite and/or nitrite into the DV and/or possibly transports haemoglobin breakdown end-products back into the parasite cytoplasm for use in cellular metabolism or further export between the parasite and/or erythrocyte.

Dual localisation of proteins in *P. falciparum* is not unknown. For instance the protease falcilysin is dually located in the digestive vacuole where it has a role in Hb degradation and is also found in the apicoplast where it in the degrades transit peptides leading to production of the full-length functional protein (Ponpuak *et al*, 2007). The falcilysin protein is also trafficked to the mitochondrion (Ralph, 2007). The lipoate protein ligase A2 (LplA2) from *P. falciparum* has been shown to be present in the mitochondrion and apicoplast (Günther, 2007). This particular protein is involved in the metabolism of lipoic acid that is found in the mitochondrion and apicoplast and LplA2 is essential for parasite sexual development in the mosquito (Günther, 2007). The vacuolar H^+-ATPase (V-ATPase) is observed in four organelles of the parasite namely the parasite plasma membrane, parasitophorus vacuole membrane, the digestive vacuole and in cytoplasmic vesicles (Hayashi *et al*, 2000). V-ATPase is known to play a role in the extrusion of protons from the parasite in the maintenance of cytoplasmic pH. However, it should be noted that in the DV, the V-ATPase actively imports protons to reduce the pH of the DV (Hayashi *et al*, 2000). In light of these examples, it is therefore possible that the *Plasmodium* protein is located to the two organellar membranes. Given the range of anionic substrates demonstrated for orthologous proteins, it is also possible different substrates are transported in the different locations.

6.1.3 Knock-out of PfNitA

Orthologues of PfNitA in *E. coli* and *A. nidulans* have previously been knocked-out demonstrating these orthologues are not essential to the growth of the organism. For example, in *A. nidulans*, deletion of the *annita* gene does not affect growth (on complex medium) due to the presence of two nitrite/nitrate transporters (Unkles *et al*, 2011). In *P. falciparum* there are few transport proteins compared to other eukaryotes, and many of these could not be knocked-out successfully (e.g. Slavic *et al*, 2010, el Bissati *et al*, 2008, and, Waller *et al*, 2003). The paucity of transport proteins in *Plasmodium* species suggests such proteins may be essential to parasite survival. To analyse the essentiality of *pfnita* to intraerythrocytic parasites, gene knock-out studies were carried out in blood stage *P. falciparum*.

The *pfnita* gene could not be disrupted in two different strains of *P. falciparum* by double cross-over recombination (Figure 4.10). We attempted to knock-out the gene in the presence of an additional copy of *pfnita* by expressing the synthetic codon-optimised gene from an episomal plasmid (Figure 4.12). If *pfnita* is essential for intraerythrocytic stage parasites, a knock-out of the gene should be possible as a second copy is present and the protein it encodes is expressed. This method employed was shown to be successful to yield a knock-out strain of *P. falciparum* for thioredoxin reductase and for the hexose transporter (Krnajski *et al*, 2002, and Slavic *et al*, 2010). The synthetic codon-optimised gene was chosen in an attempt to prevent recombination with the natural *P. falciparum* gene locus. Despite the additional synthetic codon-optimised gene, the *pfnita* locus was not targeted as judged by the failure to generate a viable strain (Figure 4.12 and 4.13). In fact, from the hybridisation pattern obtained after Southern analysis, it is possible instead that recombination between the two transfected plasmids occurred while integration at the intended locus did not take place. The recombination of plasmids may be a technical problem and consequently would need to be rectified before future experiments could take place.

The *P. falciparum* gene locus of *pfnita* was, however, targeted showing that this gene is not *per se* refractory to recombination (Figure 4.18). The tagging of the endogenous gene was achieved using a sequence homologous only to the 3' end of the *pfnita* gene rather than the full ORF as this tactic made certain that only one

functional copy of the gene was present at the recombined locus. This approach showed that the changes to the gene locus through the integration of the pCHD-Hsp86-*ctmI*-$(HA)_3$ plasmid was not detrimental *per se* to parasite survival. Instead, there did appear to be a random integration event (or events), occurring at another gene locus, of the plasmid. Random integration of plasmids can occur, and appear to be selected for when integration into the target locus leads to a growth defect (Günther *et al*, 2009, and Patzewitz, 2009). Only a small percentage of parasites underwent recombination with the pCHD-Hsp86-*ctmI*-$(HA)_3$ plasmid suggesting there may be fitness costs involved. The results obtained by Southern blot analyses for pCHD-Hsp86-*ctmI*-$(HA)_3$ recombination were somewhat ambiguous. The data obtained suggested non-specific integration of the knock-out construct into the *P. falciparum* genome as well as integration at the correct locus (Figure 4.18). The possible non-specific integration requires further investigation including pulse field gel electrophoresis to investigate other loci involved. In addition, integration at the correct locus should be examined by the cloning of the appropriate cell line by limiting dilution. As the locus could not be targeted by the pCHD-Hsp86-*ctmI*-GFP plasmid (Figure 4.16) the presence of the longer C-terminal tag was possibly deleterious to the organism or interfered with the single cross-over required for integration, potentially due to the larger size of the plasmid.

A second strategy pursued (but not completed) was to target the *pfnita* locus with a plasmid containing a partial gene but without any N- or C-terminal tags (Figure 4.19). The pHH1 based plasmid constructs were created for the purpose of a knock-out as well as a knock-out control. Such plasmids recombine by single cross-over recombination events, for the knock-out the native locus is interrupted and two non-functional copies of the gene inserted. For the knock-out control, integration of the plasmid should result in a single functional copy of the gene and a second truncated copy of the gene. In short, the knock-out control should show whether the gene locus is targetable in principle.

6.1.4 Conclusions

The results obtained from knock-out studies of *pfnita* in *P. falciparum* suggest that the *pfnita* gene is essential for the intraerythrocytic stages, as a strain harbouring such a disrupted gene appears to be unobtainable. This would fit with available expression

data from the PlasmoDB database that implies *pfnita* is present throughout the intraerythrocytic development stages.

The substrate for PfNitA remains elusive. However, given the functions and reported substrates in other organisms as well as the likely essentiality of the gene and the protein locations from the results presented in this thesis, it may be speculated the permease functions in the removal of waste products from the growing parasite. Waste products may be exported from the parasite into the parasitophorus vacuole. The parasitophorus vacuole membrane is porous and in turn may allow the waste metabolites to pass into the RBC lumen from where these molecules are excreted into the hosts' bloodstream, possibly through the new permeability pathways created by the parasite. In these respects lactate is a possible waste metabolite as it is the main by-product from glucose metabolism by the parasite. It is noteworthy that there is no annotated lactate transporter in the parasite genome on PlasmoDB[41].

6.1.5 Future work

Formate uptake analysis

An aspect that is of most importance is to confirm whether formate is transported by the PfNitA transporter as suggested by the research results presented in this thesis. Investigating the presence or concentration of formate in spent RPMI medium is complicated. Available methods to analyse formate are not sensitive enough to detect relatively low medium concentrations. The limits of the published methods are concentrations of 7 mg/L (Blomme *et al*, 2001). Also, mass spectrometry-based metabolomics cannot detect molecules as small as formate (S. Sethia, personal communication). If formate is transported into the parasite cell, heavy labelled formate could be used to follow formate metabolism in the cell as well as where its by-products are found. However, we hypothesise that formate would be produced as a toxic end-product of metabolism that would require removal from the parasite cell.

To analyse the uptake or efflux of formate and other metabolites, *pfnita* could be expressed in *Xenopus laevis* oocytes which has been used successfully before for investigation of permeases. Secondly characterisation of the uptake of a heavy

[41] http://plasmodb.org/plasmo/. Accessed on 4 September 2012.

labelled/tracer formate could be carried out in laboratory grown *Plasmodium* cultures. As well as investigating formate, the transport of other metabolic end-products such as acetate and lactate may be studied.

Another possible line of enquiry to investigate the transportation of formate is to use the 'Conditional Knock-out' method in which a gene knock-out strain is obtained with an additional copy of the gene being present on an episomal plasmid. Instead of using the same gene that is being targeted, as attempted in this study with the codon-optimised copy, a specific lactate or general monocarboxylate transporter could be used as the complementary gene copy. If the process was successful the results would indicate if the lactate transporter carries out the same function as the PfNitA permease. Although lactate is an end-product of glucose metabolism in *Plasmodium* and lactic acid in the human bloodstream leads to lactic acidosis in patients, a lactate transporter has not been annotated in the *Plasmodium* genome[42] as stated above.

Indeed, work in the laboratory of Prof. Kiaran Kirk at the Australian National University (Marchetti, Martin and Kirk, personal communication and Marchetti, 2012) it has been shown that the PfNitA transporter is capable of transporting formate and to a lesser degree nitrite. In addition to these compounds their data has also shown that lactate can be transported by the PfNitA transporter. Moreover this research group has defined the PfNitA permease as a moncarboxylate transporter (Marchetti, 2012).

Inhibitor studies

If the functional complementation assays described in Chapter 3 are optimised, the strains could be used to identify potential inhibitors of the *Plasmodium* protein. The Malaria Box[43, 44] is a collection of diverse compounds that have shown to be active against the intraerythrocytic stages of *P. falciparum* but were not toxic to mammalian cells (Figure 6.1). The Malaria Box contains 400 drug-like and probe-like compounds.

[42] http://plasmodb.org/plasmo/. Accessed on 4 September 2012.
[43] http://www.mmv.org/malariabox. Accessed on 20 August 2012.
[44] https://www.ebi.ac.uk/chemblntd. Accessed on 20 August 2012.

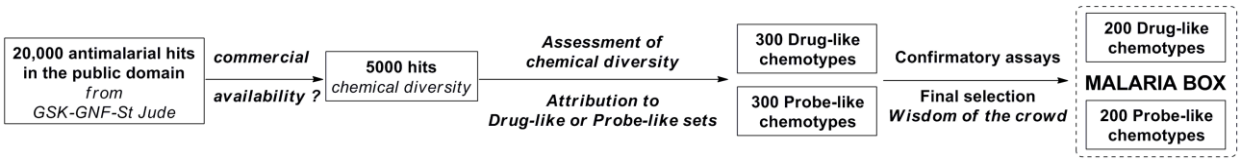


Figure 6.1: Malaria Box selection process. Image is taken from the 'Medicines for Malaria Venture' website[45].

Assays in bacterial strains expressing the PfNitA protein would be carried out in the presence and absence of the Malaria Box inhibitors. The prevention of substrate uptake would indicate the compound used inhibited the PfNitA permease. The blocking of the transporter would most likely not kill the bacteria as they can survive without the substrate. Therefore concentrations of compounds that are inhibitory but that do not kill the cells may be identified. Potentially, the 3D structural model of PfNitA generated in this thesis and computer programs to map inhibitor compounds to this model could be used to narrow down the inhibitors investigated using the assay system.

Knock-down using mutagenesis

In *A. nidulans*, the orthologous *annita* gene has been subjected to a wide range of *in vitro*-generated mutations resulting in alteration of the amino acid sequence. Following transformation of mutant constructs into *A. nidulans* strains have been assessed for their ability to (i) grow on nitrite as the sole nitrogen source, (ii) express the mutant protein to WT levels and (iii) transport nitrite (Symington, 2009 and Unkles *et al*, 2011). These studies have shown that, following mutation of four highly conserved asparagines residues located in the proteins transmembrane domains (Section 1.5.3.1, *Aspergillus*), one mutant in particular showed good protein expression and the mutation made only a slight difference to the growth of the strain on nitrite. This mutation, N214D, added a negative charge to the protein, but had little effect on the size of the amino acid side chain, and is therefore unlikely to affect the protein folding (Symington, 2009). In radioactive tracer experiments, the N214D *A. nidulans* mutant strain showed reduced nitrite influx compared to that of the strain with the WT protein (Dr. S.E. Unkles, unpublished work).

[45] http://www.mmv.org/sites/default/files/uploads/images/RandD/Selection_Process_Malaria_Box.png. Accessed on 20 August 2012.

It may be possible to exploit the data from *A. nidulans*, and the functional complementation assays described in Chapter 3 to attempt a 'knock-down' of the *P. falciparum* NitA transporter by creating an equivalent mutation that may reduce the activity of the transporter. The *E. coli* complementation assays could be used to identify any residues that lower the level of substrate transported by the protein. Mutants would be generated from *P. falciparum* gDNA and transfected into the parasites. This experimental approach relies on the effectiveness of the functional complementation assays and a single cross-over taking place at the correct locus. An allelic exchange experiment would introduce the respective mutations into the *P. falciparum* gene and may lead to the generation of a mutant *pfnita* gene encoding a transporter with reduced transport activity. Given the correct recombination genotype, the phenotype such as growth patterns could be analysed in comparison to WT and transfected non-mutant parasites (for instance, those generated for a knock-out control).

6.2 PcCir2

The Cir2 protein of *P. chabaudi* belongs to the PIR protein family. Members of this family are found in all sequenced malarias although the link between the PIR's and *P. falciparum* orthologues is questionable. Members of the PIR family are thought to play a role in immune evasion as the amino acid sequences contain a so called hypervariable region (HVR) as discussed in Chapter 1 (Figures 1.11 and 5.2).

The DNA and protein sequences of *pccir*2 do not resemble other non-redundant sequences from other species, and therefore no suggestions about the function of the protein family can be made. It was our intention to obtain structural information that might provide clues to the function of the protein.

The questions we wished to answer in this thesis were:

(i) Can we produce sufficient purified protein for crystallography trials?

(ii) What is the 2D or 3D structure of PcCir2 and can we hypothesise a function from this information?

6.2.1 Recombinant protein expression and purification

To analyse the structure of PcCir2, constructs were designed with the objective to overexpress the protein in *E. coli*. The *pccir2* gene was cloned into 3 expression plasmids each with a C-terminal His-tag for production of the full-length protein (Figures 5.3 and 5.9). As the protein appeared to be insoluble, two further constructs that would potentially produce soluble proteins were created and designated PcCir2Dom and PcCir2Dom2 (Figure 5.23). One construct was truncated at the C-terminus before the TMD. The second domain was truncated at the N-terminus after a hydrophobic stretch of amino acids and at the C-terminus before the TMD.

Purification of full-length PcCir2 using cobalt affinity chromatography proved difficult, as the recombinant protein did not appear to be particularly stable. The protein appeared to degrade during expression (Figure 5.4, 5.10 and 5.13) and aggregated during purification (Figures 5.8, 5.12, and 5.15). Moreover, the recombinant protein purified with contaminating bacterial proteins (Figures 5.8, 5.12, and 5.15). Further purification and protein concentration methods were not investigated.

The two truncated constructs were not expressed as soluble proteins, despite many attemps to do so by altering expression temperatures and times, media and bacterial strains (Figures 5.19, 5.20, 5.21, 5.22, 5.24 and 5.25). Therefore we were unable to produce protein that could be taken forwards for structural analysis and crystallography trials. Soluble protein was obtained by collaborators, but the quantities were not sufficient for structural analyses.

The PcCir2 amino acid sequence contains a single putative transmembrane domain (Figure 5.1). The N-terminal of the protein is moderately hydrophobic, although it is not thought to be a TMD. As with the N-termini of the RIFIN proteins (Kyes *et al*, 1999), the weak hydrophobic N-terminus of PcCir2 may represent a cleavable signal peptide is present. When the PcCir2 protein domains (lacking the TMD) were expressed in *E. coli* they were poorly soluble. It is possible that this weak hydrophobic stretch of amino acids was responsible for the poor solubility of the protein domain. Potentially this hydrophobic stretch of amino acids encodes a cleavable signal peptide that would remain uncleaved when expressed in *E. coli*, and this prevented the production of soluble protein.

6.2.2 Membrane topology

To experimentally determine the topology of the PcCir2 protein, a set of C-terminal fusion proteins were made for use in GFP fluorescence and PhoA activity assays as described by Rapp *et al* (2004) and Daley *et al* (2005). The *pccir2* sequence was truncated at five locations in the amino acid sequence to generate five truncated proteins as well as the single full-length protein (Figure 5.26).

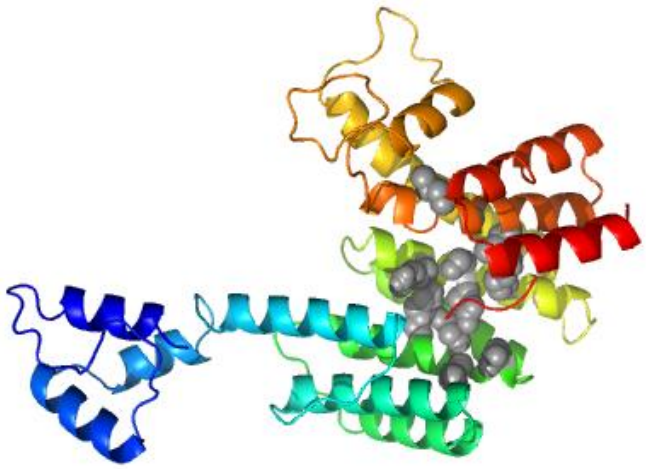


Figure 6.2: A putative PcCir2 structural model with putative binding sites. The structural model was downloaded from the I-TASSER server and edited in PyMol. The putative binding residues as suggested by I-TASSER are shown in grey, and the N-terminus is blue and C-terminus in red. A clear TMD is absent from this model.

Of these six proteins, one was found to be packed into *E. coli* inclusion bodies and was excluded from our studies (Figure 5.31). The data from each of the assays was combined (Figure 5.30) and analysed to provide a 2D topology model. The topology of the PcCir2 protein as determined experimentally has the N-terminus of the protein locatedintracellularly and the C-terminus on the outside. By extention this suggests the hypervariable region is internal.

The PcCir2 amino acid sequence was subject to analyses using the I-TASSER[46] (Zhang, 2008) server to predict the structure and possible function/s of the PcCir2 protein. The I-TASSER server selects known crystal structures as templates for protein structure predictions given an amino acid sequence. When the PcCir2 amino acid sequence was uploaded to the server, I-TASSER predicted five different structural models. The

[46] http://zhanglab.ccmb.med.umich.edu/I-TASSER/. Accessed on 4 September 2012.

templates for these models were binding proteins and receptors. This may imply that PcCir2 is a binding protein or a receptor, though there is perhaps not enough similarity between these proteins to be conclusive. An I-TASSER generated model is depicted in Figure 6.2 with the putative binding residues indicated in grey.

The topology of the PcCir2 protein places the hypervariable region inside the cell. Such a location could be due to a difference in the mechanisms of trafficking found in *P. falciparum* that are absent in *E. coli*. *P. falciparum* protein trafficking is not yet fully understood although many mechanisms have been suggested (Charpian and Przyborski, 2008). It is also believed that membrane proteins travel along a 'membrane system' before they are released into the RBC (Saridaki *et al*, 2008). Nevertheless, *E. coli* is a host that has been used for the expression of eukaryotic membrane proteins.

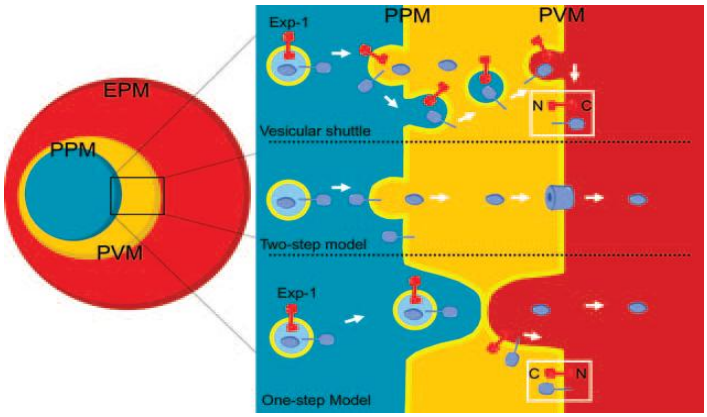


Figure 6.3: Trafficking of proteins in *Plasmodium*. A schematic diagram taken from Charpian and Przyborski (2008) to illustrate the various ways in which a protein can be trafficked outside of the parasite. Membrane proteins are associated with membrane throughout the trafficking process.

The scheme presented in Figure 6.3 illustrates the ways in which *Plasmodium* proteins are thought to be trafficked outside the RBC. The PcCir2 protein may be transported from the ER in a vesicle to the parasitophorus vacuole and beyond. In the vesicle, and each subsequent membrane, the protein is embedded in the layer of lipids that comprise the vesicle, parasitophorus or plasma membranes. Without the potential post-translational modification and signals of *P. falciparum* the nascent protein could *a priori* be inserted into the *E. coli* membrane in an orientation intended for the transport vesicle or PVM, and this membrane topology does not necessarily reflect the final

orientation of the PcCir2 protein. For instance, depending on the model presented in Figure 6.3, the final membrane topology of Exp1 is not the same as the topology in the transport vesicle (Charpian and Przyborski, 2008).

The BIR proteins of *P. berghei* were found to be associated with Maurer's Clefts as well as the PfMC2 proteins of *P. falciparum* (del Portillo *et al*, 2004 and Merino *et al*, 2006), while the STEVOR proteins, also belonging to *P. falciparum*, are trafficked though Maurer's Clefts on their way to the RBC membrane (Niang *et al*, 2009). The membrane-associated histidine-rich protein 1 (MAHRP1) of *P. falciparum* is a protein with no significant homology to other proteins and which also contains a single TMD (Spycher *et al*, 2006). The MAHRP1 protein is known to be exported to the RBC and is found in Maurer's Clefts. The C-terminus of MAHRP1 is located in the RBC cytoplasm, whilst research in this thesis suggests the N-terminus of PcCir2 is located in the RBC cytoplasm. It is tempting to speculate that the PcCir2 protein is indeed a Maurer's Cleft protein and the TMD is used to anchor the protein to the lipid rich Maurer's Cleft. Possibly the PIR family of proteins are chaperones that transport proteins lacking a PEXEL/VTS motif (Hiller *et al*, 2004, and Marti *et al*, 2004) outside the parasite and to the surface of the RBC. The PcCir2 protein was purified with a number of contaminating bacterial proteins. If the PcCir2 protein is a binding protein it may interact with various *E. coli* proteins which were consequently co-purified by the metal affinity chromatography and only visible after the denaturing conditions of the protein electrophoresis. The positioning of the PcCir2 protein at the surface of the erythrocyte or in Mauree's Clefts would make it a promising candidate for a variable antigen given its exposure to the immune system and the high level of sequence variability across the family of CIR proteins. Thus, one might speculate that the PIR protein family are chaperone proteins that reside in the Maurer's Clefts.

It has been reported that expression of eukaryotic membrane proteins with an N-terminus out topology in *E. coli* (such as G-protein receptors) can cause problems with regard to correct targeting and membrane insertion (Junge *et al*, 2008). The membrane protein topology prediction programs predicted PcCir2 to have an N-terminus out topology. It is possible that by expressing the protein in an *E. coli* heterologous system the protein is mis-targeted and inserted in the wrong orientation.

This could explain why the experimental model produces a 'C-terminus out' topology while the bulk of the protein and N-terminal are intracellular.

6.2.3 Conclusions

The production of PcCir2 domains and full-length protein was hampered by the problems that arose due to solubility and purification of the protein and its domains. As a result no pure protein was produced that could be used for further structural analysis. The membrane topology analysis of PcCir2 gave an unexpected result in that the hypervariable region was found to be intracellular. If the recombinant protein (s) were expressed and inserted in the *E. coli* inner membrane in the correct topology, we could speculate that the PcCir2 protein may function as a chaperone and is possibly present in the Maurer's Clefts found towards the periphery of the RBC. However, to fully investigate this possibility and to obtain a structural model for the PcCir2 protein, further work must be carried out.

6.2.4 Future work

Obtaining purified PcCir2

To obtain full-length PcCir2 protein that can be solubilised and purified efficiently, more expression trials require to be carried out. PcCir2 was expressed at high levels during the experimental stages of this thesis. However, problems occurred during the purification process. The use of a different tag such as a *Strep*-tag® or a glutathione-*S*-transferase (GST) tag may aid in the purification process. GST binds with high affinity to glutathione during purification, though as the GST amino acid sequence is 211 amino acids long it is quite a large fusion protein (Thermo Scientific/Pierce[47]). The purification process is carried out in near physiological conditions which may help to stabilise the PcCir2 protein and prevent the aggregation and/or degradation observed in the experimental work during this thesis. However there are problems in running GST tagged proteins on denaturing protein gels as the GST protein degrades (Thermo Scientific/Pierce) leading to additional bands on the gel. The *Strep*-tag® is highly selective for *Strep*-Tactin® and proteins are purified in a one step process under

[47] http://www.piercenet.com/browse.cfm?fldID=4A8ADF29-5056-8A76-4EC6-63375BA024E7. Accessed on 5 September 2012.

physiological conditions (IBA protocol[48]). The high selectivity of the *Strep*-tag® for *Strep*-Tactin® may eliminate many of the contaminating proteins observed when PcCir2 was purified, and the procedure occurring under physiological conditions. The different purification tag may prevent some of the aggregation and/or degradation observed. Potentially, the addition of a maltose-binding protein (MBP) fusion tag to the recombinantly expressed PcCir2 should also be assessed. Moreover, the MBP tag may assist with the solubility of the protein and prevent aggregation after purification (Nallamsetty and Waugh, 2007).

Recombinant expression of Cir homologues

To achieve soluble expression of a Cir domain numerous different Cir protein domains could be expressed. First an alignment of many different Cir amino acid sequences would be created, and the putative domains selected which correspond to the putative ordered domain found for PcCir2. Each of these putative domains may be cloned and expressed in *E. coli* with the intention of obtaining soluble and purified protein that could be used for crystallography trials. Refinements in experimental approach may include expression and purification of protein domains from BIR and YIR sequences in order to maximise the chances of obtaining soluble purified protein.

To investigate possible protein-protein interactions, a yeast two-hybrid screen could be carried out. However, this would depend on the identification of potential ligands for the PcCir2 protein. The use of Co-immunoprecipitation of the discussed full-length PcCir2 purification may provide some 'lead' ligands in the form of co-purified *E. coli* proteins. Interaction proteins could then be identified using mass spectrometry techniques.

Cir localisation

Several questions remain regarding the subcellular localisation of PcCir proteins. Although PcCir proteins have been localised to the red blood cell surface, it is not clear which termini is present on the surface (Janssen *et al*, 2004). A second question relates to where in the RBC/RBCm the proteins locate. For example, are they present in Maurer's Clefts? Is there an intracellular location of any PcCir (or even PIR)

[48] http://www.stratech.co.uk/pdf_folder/PR03_Streptag_pur_Protocol_0003.pdf. Accessed on 5 September 2012.

proteins? To attempt to answer these questions antibodies could be designed to degenerate protein sequences at either termini of the protein/s. Such antibodies would then be exploited for immunofluorescence analyses of the Cir proteins in *P. chabaudi* infected RBCs taken from a rodent model.

References

Afonso, A., Neto, Z,. Castro, H., Lopes, D., Alves A.C., Tomás, A.M., Rosário, V.D. (2010). *Plasmodium chabaudi chabaudi* malaria parasites can develop stable resistance to atovaquone with a mutation in the cytochrome b gene. *Malaria Journal* (9), 135.

Andrews, S.C., Berks, B.C., McClay, J., Ambler, A., Quail, M.A., Golby, P., Guest, J.R. (1997). A 12-cistron *Escherichia coli* operon (*hyf*) encoding a putative proton-translocating formate hydrogenlyase system. *Microbiology* (Pt 11), 3633-3647.

Arinaminpathy, Y., Khurana, E., Engelman, D.M., Gerstein, M.B. (2009). Computational analysis of membrane proteins: the largest class of drug targets. *Drug Discovery Today* (23-24), 1130-1135.

Bannai, H., Tamada, Y., Maruyama, O., Nakai, K., and Miyano, S. (2002). Extensive feature detection of N-terminal protein sorting signals. *Bioinformatics* (2), 298-305.

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

Barrett, M.P., Gilbert, I.H. (2006). Targeting of toxic compounds to the trypanosome's interior. *Advances in Parasitology* (63), 125-183.

Beckham, K.S., Potter, J.A., Unkles, S.E. (2010). Formate-nitrite transporters: optimisation of expression, purification and analysis of prokaryotic and eukaryotic representatives. *Protein Expression and Purification* (2), 184-189.

Ben Mamoun, C., Gluzman, I.Y., Goyard, S., Beverley, S.M., Goldberg, D.E. (1999). A set of independent selectable markers for transfection of the human malaria parasite *Plasmodium falciparum*. *Proceedings of the National Academy of Sciences USA* (15), 8716-8720.

Bender, A., van Dooren, G.G., Ralph, S.A., McFadden, G.I., Schneider, G. (2003). Properties and prediction of mitochondrial transit peptides from *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* (132), 59-66.

Bergsten, J. (2005). A review of long-branch attraction. *Cladistics* (21), 163-193.

Blomme, B., Lheureux, P., Gerlo, E., Maes, V. (2001). Cobas Mira S endpoint enzymatic assay for plasma formate. *Journal of Analytical Toxicology* (2), 77-80.

Bray, P.G., Ward, S.A., Ginsburg, H. (1999). Na+/H+ Antiporter, Chloroquine Uptake and Drug Resistance: Inconsistencies in a Newly Proposed Model. *Parasitology Today* (9), 360-363.

Butcher, G.A. (2007). Development of malaria blood-stage vaccines: learning from mosquitoes. *Transactions of the Royal Society of Tropical Medicine and Hygiene* (6), 530-531.

Carlton, J.M., Adams, J.H., Silva, J.C., Bidwell, S.L., Lorenzi, H., Caler, E., Crabtree, J., Angiuoli, S.V., Merino, E.F., Amedeo, P., Cheng, Q., Coulson, R.M., Crabb, B.S., del Portillo, H.A., Essien, K., Feldblyum, T.V., Fernandez-Becerra, C., Gilson, P.R., Gueye, A.H., Guo, X., Kang'a, S., Kooij, T.W., Korsinczky, M., Meyer, E.V., Nene, V., Paulsen, I., White, O., Ralph, S.A., Ren, Q., Sargeant, T.J., Salzberg, S.L., Stoeckert, C.J., Sullivan, S.A., Yamamoto, M.M., Hoffman, S.L., Wortman, J.R., Gardner, M.J., Galinski, M.R., Barnwell, J.W., Fraser-Liggett, C.M. (2008). Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature* (7214), 757-763.

Carter, N.S., Ben Mamoun, C., Lui, C., Silva, E.O., Landfear, S.M., Goldberg, D.E., Ullman, B. (2000). Isolation and functional characterisation of the PfNT1 nucleoside transporter gene from *Plasmodium falciparum. Journal of Biological Chemistry* (275), 10683-10691.

Charpian, S., Przyborski, J.M. (2008). Protein transport across the parasitophorous vacuole of *Plasmodium falciparum*: into the great wide open. *Traffic* (2), 157-165.

Chollet, C., Baliani, A., Wong, P.E., Barrett, M.P., Gilbert, I.H. (2009). Targeted delivery of compounds to *Trypanosoma brucei* using melamnie motifs. *Bioorganic and Medicinal Chemistry* (17), 2512-2523.

Clegg, S., Yu, F., Griffiths, L., Cole, J.A. (2002). The roles of the polytopic membrane proteins NarK, NarU and NirC in *Escherichia coli* K-12: two nitrate and three nitrite transporters. *Molecular Microbiology* (44), 143-155.

Cole, J.A., Coleman, K.J., Compton, B.E., Kavanagh, B.M., Keevil, C.W. (1974). Nitrite and ammonia assimilation by anaerobic continuous cultures of *Escherichia coli*. *Journal of General Microbiology* (1), 11-22.

Cooper, R.A., Hartwig, C.L., and Ferdig, M.T. (2005). *pfcrt* is more than the *Plasmodium falciparum* chloroquine resistance gene: a functional and evolutionary perspective. *Acta Tropica* (94), 170-180.

Crabb, B.S., Rug, M., Gilberger, T.W., Thompson, J.K., Triglia, T., Maier, A.G., and Cowman, A.F. (2004). Transfection of the human malaria parasite *Plasmodium falciparum*. *Methods in Molecular Biology* (270), 263-276.

Cunningham, D., Lawton, J., Jarra, W., Preiser, P., Langhorne, J. (2010). The *pir* multigene family of *Plasmodium*: antigenic variation and beyond. *Molecular and Biochemical Parasitology* (2), 65-73.

Cunningham, D.A., Fonager, J., Jarra, W., Carret, C., Preiser, P.R., Langhorne, J. (2009). Rapid changes in transcription profiles of the *Plasmodium yoelii yir* multigene family in clonal populations: lack of epigenetic memory? *PLoS ONE* (1), e4285.

Cunningham, D.A., Jarra, W., Koernig, S., Fonager, J., Fernandez-Reyes, D., Blythe, J.E., Waller, C., Preiser, P.R., Langhorne, J. (2005). Host immunity modulates transcriptional changes in a multigene family (*yir*) of rodent malaria. *Molecular Microbiology* (3), 636-647.

Czyzewski, B.K., Wang, D.N. (2012). Identification and characterization of a bacterial hydrosulphide ion channel. *Nature* (7390), 494-497.

Daley, D.O., Rapp, M., Granseth, E., Melén, K., Drew, D., von Heijne, G. (2005). Global topology analysis of the *Escherichia coli* inner membrane proteome. *Science* (5726), 1321-1323.

Darnell, J.E., Lodish, H., Berk, A., Zipursky, L., Matsudaria, P., Baltimore, D. (2000). *Molecular Cell Biology. 4th edition*: W.H.Freeman & Co Ltd.

Das, P., Lahiri, A., Lahiri, A., Chakravortty, D. (2009). Novel role of the nitrite transporter NirC in Salmonella pathogenesis: SPI2-dependent suppression of inducible nitric oxide synthase in activated macrophages. *Microbiology* (Pt8), 2476-2489.

del Portillo, H.A., Fernandez-Becerra, C., Bowman, S., Oliver, K., Preuss, M., Sanchez, C.P., Schneider, N.K., Villalobos, J.M., Rajandream, M.A., Harris, D., Pereira da Silva, L.H., Barrell, B., Lanzer, M. (2001). A superfamily of variant genes encoded in the subtelomeric region of *Plasmodium vivax*. *Nature* (410), 839-842.

del Portillo, H.A., Lanzer, M., Rodriguez-Malaga, S., Zavala, F., Fernandez-Becerra, C. (2004). Variant genes and the spleen in *Plasmodium vivax* malaria. *International Journal of Parasitology* (13-14), 1547-1554.

Deloménie, C., Foti, E., Floch, E., Diderot, V., Porquet, D., Dupuy, C., Bonaly, J. (2007). A new homolog of FocA transporters identified in cadmium-resistant *Euglena gracilis*. *Biochemical and Biophysical Research Communications* (358), 455-461.

Dobson, P.D., Kell, D.B. (2008). Carrier-mediated cellular uptake of pharmaceutical drugs: an exception or the rule? *Nature Reviews Drug Discovery* (7), 205-220.

Downie, M.J., el Bissati, K., Bobenchik, A.M., Nic Lochlainn, L., Amerik, A., Zufferey, R., Kirk, K., Ben Mamoun, C. (2010). PfNT2, a permease of the equilibrative nucleoside transporter family in the endoplasmic reticulum of *Plasmodium falciparum*. *The Journal of Biological Chemistry* (27), 20827-20833.

Drew, M.E., Banerjee, R., Uffman, E.W., Gilbertson, S., Rosenthal, P.J., Goldberg, D.E. (2008). *Plasmodium* food vacuole plasmepsins are activated by falcipains. *Journal of Biological Chemistry* (19), 12870-12876.

Duraisingh, M.T., Triglia, T., Cowman, A.F. (2002). Negative selection of *Plasmodium falciparum* reveals targeted gene deletion by double crossover recombination. *International Journal of Parasitology* (1), 81-89.

Dzikowski, R., Frank, M., Deitsch, K. (2006). Mutually exclusive expression of virulence genes by malaria parasites is regulated independently of antigen production. *PLoS Pathogens* (3), e22.

Ebbinghaus, P., Krücken, J. (2011). Characterization and tissue-specific expression patterns of the *Plasmodium chabaudi cir* multigene family. *Malaria Journal* (10), 272.

el Bissati, K., Downie, M.J., Kim, S., Horowitz, M., Carter, N., Ullman, B., Ben Mamoun, C. (2008). Genetic evidence for the role of PfNT1 in the transport and utilisation of xanthine, guanine, guanosine and adenine by *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* (161), 130-139.

el Bissati, K., Zufferey, R., Witola, W.H., Carter, N.S., Ullman, B., Ben Mamoun, C. (2006). The plasma membrane permease PfNT1 is essential for purine salvage in the human malaria parasite *Plasmodium falciparum*. *Proceedings of the National Academy of Sciences USA* (24), 9286-9291.

Erasmus, D. (n.d.). BrainyQuote.com. Retrieved July 1, 2012, from BrainyQuote.com. Web site: http://www.brainyquote.com/quotes/quotes/d/desiderius148997.html

Faguy, D.M. (2000). The controlled chaos of shifty pathogens. *Current Biology* (10), R498-R501.

Falke, D., Schulz, K., Doberenz, C., Beyer, L., Lilie, H., Thiemer, B., Sawers, R.G. (2009). Unexpected oligomeric structure of the FocA formate channel of *Escherichia coli* : a paradigm for the formate-nitrite transporter family of integral membrane proteins. *FEMS Microbiology Letters* (1), 69-75.

Fang, W., Vega-Rodeíguez, J., Ghosh, A.K., Jacobs-Lorena, M., Kang, A. St Leger, R.J. (2011). Development of transgenic fungi that kill human malaria parasites in mosquitoes. *Science* (6020), 1074-1047.

Fernandez-Becerra, C., Yamamoto, M.M., Vencio, R.Z.N., Lacerda, M., Rosanas-Urgell. A., del Portillo, H.A. (2008). *Plasmodium vivax* and the importance of the subtelomeric multigene *vir* family. *Cell* (25), 44-51.

Fernandez, E., Galvan, A. (2007). Inorganic nitrogen assimilation in *Chlamydomonas*. *Journal of Experimental Botany* (9), 2279-2287.

Fidock, D.A., Nomura, T., Talley, A.K., Cooper, R.A., Dzekunov, S.M., Ferdig, M.T., Ursos, L.M., Sidhu, A.B., Naudé, 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. *Molecular Cell* (4), 861-671.

Fidock, D.A., Wellems, T.E. (1997). Transformation with human dihydrofolate reductase renders malaria parasites insensitive to WR99210 but does not affect

the intrinsic activity of proguanil. *Proceedings of the National Academy of Sciences USA* (20), 10931-10936.

Fischer, K., Chavchich, M., Huestis, R., Wilson, D.W., Kemp, D.J., Saul, A. (2003). Ten families of variant genes encoded in subtelomeric regions of multiple chromosomes of *Plasmodium chabaudi*, a malaria species that undergoes antigenic variation in the laboratory mouse. *Molecular Microbiology* (5), 1209-1223.

Fonager, J., Cunningham, D., Jarra, W., Koernig, S., Henneman, A.A., Langhorne, J., Preiser, P. (2007). Transcription and alternative splicing in the yir multigene family of the malaria parasite *Plasmodium y. yoelii*: identification of motif suggesting epigenetic and post-transcriptional control of RNA expression. *Molecular and Biochemical Parasitology* (156), 1-11.

Foth, B.J., Ralph, S.A., Tonkin, C.J., Struck, N.S., Fraunholz, M., Roos, D.S., Cowman, A.F., McFadden, G.I. (2003). Dissecting apicoplast targeting in the malaria parasite *Plasmodium falciparum*. *Science* (5607), 705-708.

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

Frank, M., Deitsch, K. (2006). Activation, silencing and mutually exclusive expression within the var gene family of *Plasmodium falciparum*. *International Journal of Parasitology* (9), 975-985.

Fulton, J.D., Spooner, D.F. (1956). The *in vitro* rerspiratory metabolism of erythrocytic forms of *Plasmodium berghei*. *Parasitology V*, 58-78.

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

Gardner, M.J., Shallom, S.J., Carlton, J.M., Salzberg, S.L., Nene, V., Shoaibi, A., Ciecko, A., Lynn, J., Rizzo, M., Weaver, B., Jarrahi, B., Brenner, M., Parvizi, B., Tallon, L., Moazzez, A., Granger, D., Fujii, C., Hansen, C., Pederson, J., Feldblyum, T., Peterson, J., Suh, B., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., White, O., Cummings, L.M., Smith, H.O., Adams, M.D., Venter, J.C., Carucci, D.J., Hoffman, S.L., Fraser, C.M. (2002b). Sequence of *Plasmodium falciparum* chromosomes 2, 10, 11 and 14. *Nature* (6906), 531-534.

Ginsburg, H. (2010). Metabolism: Malaria parasite stands out. *Nature* (466), 702-703.

Girolamo, F.D., Reggi, C., Birago, C., Pizzi, E., Lalle, M., Picci, L., Pace, T., Bachi, A., de Jong, J., Janse, C.J., Waters, A.P., Sargiacomo, M., Ponzi, M. (2008). *Plasmodium* lipid rafts contain proteins implicated in vesicular trafficking and signalling as well as members of the PIR superfamily, potentially implicated in host immune system interactions. *Proteomics* (8), 2500-2513.

Goodman, C.D., Su, V., McFadden, G.I. (2007). The effects of antibacterials on the malaria parasite *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* (152), 181-191.

Gramaglia, I., Sobolewski, P., Meays, D., Contreras, R., Nolan, J.P., Frangos, J.A., Intaglietta, M., van der Heyde, H.C. (2006). Low nitric oxide bioavailability contributes to the genesis of experimental cerebral malaria. *Naure Medicine* (12), 1417-1422.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology* (3), 307-321.

Günther, S. (2007). *Lipoic acid protein ligases in Plasmodium spp.* (PhD), University of Glasgow.

Günther, S., Matuschewski, K., and Müller, S. (2009). Knockout studies reveal an important role of *Plasmodium* lipoic acid protein ligase A1 for asexual blood stage parasite survival. *PLoS ONE* (5), e5510.

Hall, N., Pain, A., Berriman, M., Churcher, C., Harris, B., Harris, D., Mungall, K., Bowman, S., Atkin, R., Baker, S., Barron, A., Brooks, K., Buckee, C.O., Burrows, C., Cherevach, I., Chillingworth, C., Chillingworth, T., Christodoulou,

Z., Clark, L., Clark, R., Corton, C., Cronin, A., Davies, R., Davis, P., Dear, P., Dearden, F., Doggett, J., Feltwell, T., Goble, A., Goodhead, I., Gwilliam, R., Hamlin, N., Hance, Z., Harper, D., Hauser, H., Hornsby, T., Holroyd, S., Horrocks, P., Humphray, S., Jagels, K., James, K.D., Johnson, D., Kerhornou, A., Knights, A., Konfortov, B., Kyes, S., Larke, N., Lawson, D., Lennard, N., Line, A., Maddison, M., McLean, J., Mooney, P., Moule, S., Murphy, L., Oliver, K., Ormond, D., Price, C., Quail, M.A., Rabbinowitsch, E., Rajandream, M.A., Rutter, S., Rutherford, K.M., Sanders, M., Simmonds, M., Seeger, K., Sharp, S., Smith, R., Squares, R., Squares, S., Stevens, K., Taylor, K., Tivey, A., Unwin, L., Whitehead, S., Woodward, J., Sulston, J.E., Craig, A., Newbold, C., Barrell, B.G. (2002). Sequence of *Plasmodium falciparum* chromosomes 1, 3-9 and 13. *Nature* (6906), 527-531.

Hall, N., Karras, M., Raine, J.D., Carlton, J.M., Kooij, T.W., Berriman, M., Florens, L., Janssen, C.S., Pain, A., Christophides, G.K., James, K., Rutherford, K., Harris, B., Harris, D., Churcher, C., Quail, M.A., Ormond, D., Doggett, J., Trueman, H.E., Mendoza, J., Bidwell, S.L., Rajandream, M.A., Carucci, D.J., Yates, J.R. 3rd., Kafatos, F.C., Janse, C.J., Barrell, B., Turner, C.M., Waters, A.P., Sinden, R.E. (2005). A comprehensive survey of the *Plasmodium* lifecycle by genomic, transcriptomic and proteomic analyses. *Science* (307), 82-86.

Hansen, M., Kun, J.F., Schultz, J.E., Beitz, E. (2002). A single, bi-functional aquaglyceroporin in blood-stage *Plasmodium falciparum* malaria parasites. *The Journal of Biological Chemistry* (7), 4874-4882.

Hayashi, M., Yamada, H., Mitamura, T., Horii, T., Yamamoto, A., Moriyama, Y. (2000). Vacuolar H(+)-ATPase localized in plasma membranes of malaria parasite cells, *Plasmodium falciparum*, is involved in regional acidification of parasitized erythrocytes. *Journal of Biological Chemistry* 44), 34353-34358.

Hedfalk, K., Pettersson, N., Oberg, F., Hohmann, S., Gordon, E. (2008). Production, characterisation and crystallisation of the *Plasmodium falciparum* aquaporin. *Protein Expression and Purification* (59), 69-78.

Herrera-Ortiz, A., Martínez-Barnetche, J., Smit, N., Rodriguez, M.H., Lanz-Mendoza, H. (2010). The effect of nitric oxide and hydrogen peroxide in the activation of

the systemic immune response of Anopheles albimanus infected with *Plasmodium berghei*. *Developmental and Comparative Immunology* (1), 44-50.

Herrera, S., Perlaza, B.L., Bonelo, A., Arevalo-Herrera, M. (2002). *Aotus* monkeys; their great value for anti-malaria vaccines and drug testing. *International Journal of Parasitology* (32), 1625-1635.

Hill, A. (2012). *Vaccine designs for intracellular parasites.* Paper presented at the Keynote Institute Lecture Series, Institute of Infection, Immunity and Inflammation., University of Glasgow.

Hiller, N.L., Bhattacharjee, S., van Ooij, C., Liolios, K., Harrison, T., Lopez-Estraño, C., Haldar, K. (2004). A host-targeting signal in virulence proteins reveals a secretome in malarial infection. *Science* (5703), 1934-1937.

Hofmann, K., Stoffel, W. (1993). TMbase - A database of membrane spanning proteins segments. *Biological Chemistry Hoppe-Seyler* (166), MF C-35.

Holder, A.A. (2009). Malaria vaccines: where next? *PLoS Pathogens*.(10), e1000638.

Hope, R. (2005). *Detailed Method for Membrane Preparation.* University of Leeds.

Howard, A.F.V., Koenraadt, C.J.M., Farenhorst, M., Knols, B.G.J., Takken, W. (2010). Pyrethroid resistance in *Anopheles gambiae* leads to increased susceptibility to the entomopathogenic fungi *Metarhizium anisopliae* and *Beauveria bassiana*. *Malaria Journal* (9), 168.

Hyman, R.W., Fung, E., Conway, A., Kurdi, O., Mao, J., Miranda, M., Nakao, B., Rowley, D., Tamaki, T., Wang, F., Davis, R.W. (2002). Sequence of *Plasmodium falciparum* chromosome 12. *Nature* (6906), 534-537.

Hyde, J.E. . (2007). Drug-resistant malaria – an insight. *FEBS Journal*.(4), 4688-4698.

Inoue, H., Nojima, H., Okayama, H. (1990). High efficiency transformation of *Escherichia coli* with plasmids. *Gene* (96), 23-28.

Janssen, C.S., Barrett, M.P., Turner, C.M.R., Phillips, R.S. (2002). A large gene family for putative variant antigens shared by human and rodent malaria parasites. *Proceedings of the Royal Society London* (269), 431-435.

Janssen, C.S., Phillips, R.S., Turner, C.M.R., Barrett, M.P. (2004). *Plasmodium* interspersed repeats: the major multigene superfamily of malaria parasites. *Nucleic Acids Research* (32), 5712-5720.

Jemmely, N.Y., Niang, M., Preiser, P.R. (2010). Small variant surface antigens and *Plasmodium* evasion of immunity. *Future Microbiology* (4), 663-682.

Jennings, M.L., Cui, J. (2008). Chloride homeostasis in *Saccharomyces cerevisiae*: high affinity influx, V-ATPase-dependent sequestration, and identification of a candidate Cl- sensor. *Journal of General Physiology* (4), 379-391.

Jia, W., Tovell, N., Clegg, S., Trimmer, M., Cole, J. (2009). A single channel for nitrate uptake, nitrite export and nitrite uptake by *Escherichia coli* NarU and a role for NirC in nitrite export and uptake. *The Biochemical Journal* (1), 297-304.

Joët, T., Krishna, S. (2004). The hexose transporter of *Plasmodium falciparum* is a worthy drug target. *Acta Tropica.*(3), 371-374.

Junge, F., Schneider, B., Reckel, S., Schwarz, D., Dötsch, V., Bernhard, F. (2008). Large-scale production of functional membrane proteins. *Cellular and Molecular Life Sciences* (11), 1729-1755.

Kang, M., Lisk, G., Hollingworth, S., Baylor, S.M., Desai, S.A. . (2005). Malaria parasites are rapidly killed by dantrolene derivatives specific for the plasmodial surface anion channel. *Molecular Pharmacology* (68), 34-40.

Katoh, K., Misawa, K., Kuma, K., Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* (14), 3059-3066.

Keane, T.M., Creevey, C.J., Pentony, M.M., Naughton, T.J., McInerney, J.O. (2006). Assessment of methods for amino acid matrix selection and their use on empirical data shows that *ad hoc* assumptions for choice of matrix are not justified. *BMC Evolutionary Biology* (6), 29.

Kell, DB. (2012). *The cellular uptake of pharmaceutical drugs: a problem not of biophysics but of systems biology*. Paper presented at the Glasgow Polyomics Launch Symposium., University of Glasgow.

Kim, S.W., Fushinobu, S., Zhou, S., Wakagi, T., Shoun, H. (2010). The possible involvement of copper-containing nitrite reductase (NirK) and flavohemoglobin in denitrification by the fungus *Cylindrocarpon tonkinense*. *Bioscience, Biotechnology and Biochemistry* (7), 1403-1407.

Kissinger, J.C., Brunk, B.P., Crabtree, J., Fraunholz, M.J., Gajria, B., Milgram, A.J., Pearson, D.S., Schug, J., Bahl, A., Diskin, S.J., Ginsburg, H., Grant, G.R.,

Gupta, D., Labo, P., Li, L., Mailman, M.D., McWeeney, S.K., Whetzel, P., Stoeckert, C.J., Roos, D.S. (2002). The *Plasmodium* genome database. *Nature* (6906), 490-492.

Klotz, F.W., Scheller, L.F., Seguin, M.C., Kumar, N., Marletta, M.A., Green, S.J., Azad, A.F. (1995). Co-localization of inducible-nitric oxide synthase and *Plasmodium berghei* in hepatocytes from rats immunized with irradiated sporozoites. *Journal of Immunology* (7), 3391-3395.

Koka, S., Lang, C., Niemoeller, O.M., Boini, K.M., Nicolay, J.P., Huber, S.M., Lang, F. (2008). Influence of NO synthase inhibitor L-NAME on parasitemia and survival of *Plasmodium berghei* infected mice. *Cellular Physiology and Biochemistry*.(5-6), 481-488.

Koteswara Reddy, G., Nagamalleswara Rao, K., Phani Rama Krishna, B., Aravind, S. (2010). *In silico* identification of potential therapeutic targets in *Clostridium botulinum* by the approach subtractive genomics *International Journal of Bioinformatics Research* (2), 16-16.

Krishna, S., Woodrow, C.J., Staines, H.M., Haynes, R.K., Mercereau-Puijalon, O. (2006). Re-evaluation of how artemisinins work in light of emerging evidence of in vitro resistance. *Trends in Molecular Medicine* (5), 200-205.

Krnajski, Z., Gilberger, T.W., Walter, R.D., Cowman, A.F., Müller, S. (2002). Thioredoxin reductase is essential for the survival of *Plasmodium falciparum* erythrocytic stages. *Journal of Biological Chemistry* (29), 35970-25975.

Krungkrai, J., Yuthavong, Y., Webster, H.K. (1985). Guanosine triphosphate cyclohydrolase in *Plasmodium facliparum* and other *Plasmodium* species. *Molecular and Biochemical Parasitology* (17), 265-276.

Kyes, S.A., Rowe, J.A., Kriek, N., Newbold, C.I. (1999). Rifins: a second family of clonally variant proteins expressed on the surface of red cells infected with *Plasmodium falciparum*. *Proceedings of the National Academy of Sciences USA* (16), 9333-9338.

Lakshmanan, V., Bray, P.G., Verdier-Pinard, D., Johnson, D.J., Horrocks, P., Muhle, R.A., Alakpa, G.E., Hughes, R.H., Ward, S.A., Krogstad, D.J., Sidhu, A.B, Fidock, D.A. (2005). A critical role for PfCRT K76T in *Plasmodium falciparum*

verapamil-reversible chloroquine resistance. *The EMBO Journal* (24), 2294-2305.

Lambros, C., and Vanderberg, J.P. (1979). Synchronization of *Plasmodium falciparum* erythrocytic stages in culture. *Journal of Parasitology*.(65), 418-420.

Lasonder, E., Ishihama, Y., Andersen, J.S., Vermunt, A.M., Pain, A., Sauerwein, R.W., Eling, W.M., Hall, N., Waters, A.P., Stunnenberg, H.G., Mann, M. (2002). Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. *Nature* (6906), 573-542.

Lawton, J., Brugat, T., Yan, Y.X., Reid, A.J., Böhme, U., Otto, T.D., Pain, A., Jackson, A., Berriman, M., Cunningham, D., Preiser, P., Langhorne, J. (2012). Characterisation and gene expression analysis of the *cir* milti-gene family of *Plasmodium chabaudi* (AS). *BMC Genomics* (13), 125.

Le Roch, K.G., Zhou, Y., Blair, P.L., Grainger, M., Moch, J.K., Haynes, J.D., De La Vega, P., Holder, A.A., Batalov, S., Carucci, D.J., Winzeler, E.A. (2003). Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* (5639), 1503-1508.

Lengeler, C., Armstrong-Schellenberg, J., D'Alessandro, U., Binka, F., Cattani, J. (2006). Relative versus absolute risk of dying reduction after using insecticide treated bednets in Africa. *Tropical Medicine and International Health* (3), 286-290.

Lim, L., McFadden, G.I. (2010). The evolution, metabolism and functions of the apicoplast. *Philosophical Transactions of the Royal Society of London, Series B, Biological Science* (365), 749-763.

López-Barragán, M.J., Lemieux, J., Quiñones, M., Williamson, K.C., Molina-Cruz, A., Cui, K., Barillas-Mury, C., Zhao, K., Su, X.Z. (2011). Directional gene expression and antisense transcripts in sexual and asexual stages of *Plasmodium falciparum*. *BMC Genomics* (12):587.

Lü, W., Du, J., Schwarzer, N.J., Gerbig-Smentek, E., Einsle, O., Andrade, S.L.A. (2012). The formate channel FocA exports the products of mixed-acid fermentation. *Proceedings of the National Academy of Sciences USA*, 1204201109v1204201101-1201204201.

Lü, W., Du, J., Wacker, T., Gerbig-Smentek, E., Andrade, S.L.A., Einsle, O. (2011). pH-dependent gating in a FocA formate channel. *Science* (6027), 352-354.

Lui, H., Naismith, J.H. (2009). A simple and efficient expression and purification system using two newly constructed vectors. *Protein Expression and Purification* (63), 102-111.

Mackinnon, M.J., Walker, P.R., Rowe, J.A. (2002). *Plasmodium chabaudi*: rosetting in a rodent malaria model. *Experimental Parasitology* (2-3), 121-128.

Maier, A.G., Braks, J.A., Waters, A.P., and Cowman, A.F. (2006). Negative selection using yeast cytosine deaminase/uracil phosphoribosyl transferase in *Plasmodium falciparum* for targeted gene deletion by double crossover recombination. *Molecular and Biochemical Parasitology* (150), 118-121.

Maier, A.G., Cooke, B.M., Cowman, A.F., Tilley, L. (2009). Malaria parasite proteins that remodel the host erythrocyte. *Nature Reviews Microbiology* (7), 341-354.

Manoil, C. (1991). Analysis of membrane protein topology using alkaline phosphatase and beta-galactosidase gene fusions. *Methods in Cell Biology* (34), 61-75.

Marchetti, R. (2012). *A monocarboxylate transporter in the intraerythrocytic malaria parasite, Plasmodium falciparum.* Poster presented at, Molecular Approaches to Malaria 2012. Lorne, Australia.

Mariscal, V., Moulin, P., Orsel, M., Miller, A.J., Fernandez, E., Galvan, A. (2006). Differential regulation of the *Chlamydomonas* Nar 1 gene by carbon and nitrogen. *Protist* (157), 421-433.

Marti, M., Good, R.T., Rug, M., Knuepfer, E., Cowman, A.F. (2004). Targeting malaria virulence and remodelling proteins to the host erythrocyte. *Science* (306), 1930-1933.

Martin, R.E., Ginsburg, H., Kirk, K. (2009b). Membrane transport proteins of the malaria parasite. *Molecular Microbiology* (3), 519-528.

Martin, R.E., Henry, R.I., Abbey, J.L., Clements, J.D., Kirk, K. (2005). The 'permeome' of the malaria parasite: an overview of the membrane transport proteins of *Plasmodium falciparum*. *Genome Biology* (6), R26.

Martin, R.E., Marchetti, R.V., Cowman, A.I., Howitt, S.M., Bröer, S., Kirk, K. (2009a). Chloroquine transport via the malaria parasite's chloroquine resistance transporter. *Science* (5948), 1680-1682.

Melén, K., Krogh, A., von Heijne, G. (2003). Reliability measures for membrane protein topology prediction algorithms. *Journal of Molecular Biology* (3), 735-744.

Ménard, R. (2000). The journey of the malaria sporozoite through its hosts: two parasite proteins lead the way. . *Microbes and Infection* (2), 633-642.

Merino, E.F., Fernandez-Becerra, C., Durham, A.M., Ferreira, J.E., Tumilasci, V.F., d'Arc-Neves, J., da Silva-Nunes, M., Ferreira, M.U., Wickramarachchi, T., Udagama-Randeniya, P., Handunnetti, S.M., Del Portillo, H.A. (2006). Multi-character population study of the *vir* subtelomeric multigene superfamily of *Plasmodium vivax*, a major human parasite. *Molecular and Biochemical Parasitology* (149), 10-16.

Miroux, B., Walker, J.E. (1996). Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of Molecular Biology* (3), 289-298.

Mitaku, S., Hirokawa, T., and Tsuji T. (2002). Amphiphilicity index of polar amino acids as an aid in the characterization of amino acid preference at membrane-water interfaces. *Bioinformatics* (18), 608-616.

Möller, S., Croning, M.D., Apweiler, R. (2001). Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* (7), 649-653.

Mota, M.M., Rodriguez, A. (2004). Migration through host cells: the first steps of *Plasmodium* sporozoites in the mammalian host. *Cellular Microbiology* (12), 1119-1118.

Mullin, K., Lim, L., Ralph, S.A., Spurk, T.P., Handman, E., McFadden, G.I. (2006). Membrane transporters in the relict plastid of malaria parasites. *Proceedings of the National Academy of Sciences USA* (103), 9571-9577.

Mwakalinga, S.B., Wang, C.W., Bengtsson, D.C., Turner, L., Dinko, B., Lusingu, J.P., Arnot, D.E., Sutherland, C.J., Theander, T.G., Lavstsen, T. (2012). Expression of a type B RIFIN in *Plasmodium falciparum* merozoites and gametes. *Malaria Journal* (11), 429.

Mwangangi, J.M., Kahindi, S.C., Kibe L.W., Nzovu, J.G., Luethy, P., Githure, I.J., Mbogo C.M. (2011). Wide-scale application of Bti/Bs biolarvicide in different

aquatic habitat types in urban and peri-urban Malindi, Kenya. *Parasitology Research* (108), 1355-1363.

Mycek, M.J., Harvey, R.A., Champe, P.C. (2000). *Lippincott's Illustrated Reviews: Pharmacology 2nd Edition*: Lippincott, Williams and Wilkins, London.

Nair, S., Miller, B., Barends, M., Jaidee, A., Patel, J., Mayxay, M., Newton, P., Nosten, F., Ferdig, M.T., Anderson, T.J.C. (2008). Adaptive copy number evolution in malaria parasites. *PloS Genetics* (10), e1000243.

Nallamsetty, S., Waugh, D.S. (2007). A generic protocol for the expression and purification of recombinant proteins in *Escherichia coli* using a combinatorial His6-maltose binding protein fusion tag. *Nature Protocols* (2), 383-391.

Naughton, J.A., Nasizadeh, S., Bell, A. (2010). Downstream effects of haemoglobinase inhibition in *Plasmodium falciparum*-infected erythrocytes. *Molecular and Biochemical Parasitology.*(2), 81-87.

Newby, Z.E.R., O'Connell, J, 3rd., Robles-Colmenares, Y., Khademi, S., Miercke, L.J., Stroud, R.M. (2008). Crystal structure of the aquaglyceroporin PfAQP from the malarial parasite *Plasmodium falciparum*. *Nature Structural and Molecular Biology* (6), 619-625.

Newstead, S., Kim, H., von Heijne, G., Iwata, S., Drew, D. (2007). High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences USA* (35), 13939-13941.

Niang, M., Yan Yam, X., and Preiser, P.R. (2009). The *Plasmodium falciparum* STEVOR multigene family mediates antigenic variation of the infected erythrocyte. *PLoS Pathogens* (5) e1000307.

Nosten, F., White, N.J. (2007). Artemisinin-based combination treatment of *falciparum* malaria. *The American Journal of Tropical Medicine and Hygiene* (6 Suppl), 181-192.

Nozawa, A., Fujimoto, R., Matsuoka, H., Tsuboi, T., Tozawa, Y. (2011). Cell-free synthesis, reconstitution, and characterization of a mitochondrial dicarboxylate-tricarboxylate carrier of *Plasmodium falciparum*. *Biochemical and Biophysical Research Communications* (3), 612-617.

Olszewski, K.L., Llinás, M. (2011). Central carbon metabolism of *Plasmodium* parasites. *Molecular and Biochemical Parasitology*(2), 95-103.

Olszewski K.L., Mather M.W., Morrisey J.M., Garcia B.A., Vaidya A.B., Rabinowitz J.D., Llinás, M. (2010). Branched tricarboxylic acid metabolism in *Plasmodium falciparum*. *Nature* (466), 774-778.

Ostera, G., Tokumasu, F., Oliveira, F., Sa, J., Furuya, T., Teixeira, C., Dvorak, J. (2008). *Plasmodium falciparum*: food vacuole localization of nitric oxide-derived species in intraerythrocytic stages of the malaria parasite. *Experimental Parasitology* (1), 29-38.

Ostera, G., Tokumasu, F., Teixeira, C., Collin, N., Sa, J., Hume, J., Kumar, S., Ribeiro, J., Lukat-Rodgers, G.S., Rodgers, K.R. (2011). *Plasmodium falciparum*: nitric oxide modulates heme speciation in isolated food vacuoles. *Experimental Parasitology* (1), 1-8.

Packard, R.M. (2007). *The Making of a Tropical Disease: A Short History of Malaria*: Johns Hopkins University Press, Baltimore.

Page, R.D.M. (1996). TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* (12), 357-358.

Pain, A., Böhme, U., Berry, A.E., Mungall, K., Finn, R.D., Jackson, A.P., Mourier, T., Mistry, J., Pasini, E.M., Aslett, M.A., Balasubrammaniam, S., Borgwardt, K., Brooks, K., Carret, C., Carver, T.J., Cherevach, I., Chillingworth, T., Clark, T.G., Galinski, M.R., Hall, N., Harper, D., Harris, D., Hauser, H., Ivens, A., Janssen, C.S., Keane, T., Larke, N., Lapp, S., Marti, M., Moule, S., Meyer, I.M., Ormond, D., Peters, N., Sanders, M., Sanders, S., Sargeant, T.J., Simmonds, M., Smith, F., Squares, R., Thurston, S., Tivey, A.R., Walker, D., White, B., Zuiderwijk, E., Churcher, C., Quail, M.A., Cowman, A.F., Turner, C.M.R., Rajandream, M.A., Kocken, C.H., Thomas, A.W., Newbold, C.I., Barrell, B.G., Berriman, M. (2008). The genome of the simian and human malaria parasite *Plasmodium knowlesi*. *Nature*.(7214), 799-803.

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

Parker, M.D., Hyde, R.J., Yao, S.Y., McRobert, L., Cass, C.E., Young, J.D., McConkey, G.A., Baldwin, S.A. (2000). Identification of a nucleoside/nucleobase transporter from *Plasmodium falciparum*, a novel target for antimalarial chemotherapy. *Biochemical Journa* .(349), 67-75.

Pasternak, N.D., Dzikowski, R. (2009). PfEMP1: an antigen that plays a key role in the pathogenicity and immune evasion of the malaria parasite *Plasmodium falciparum*. . *International Journal of Biochemistry and Cell Biology* (40), 1463-1466.

Patzewitz, E. (2009). *Glutathione metabolism of Plasmodium falciparum.* (PhD), University of Glasgow.

Paulsen, I.T., Sliwinski, M.K., Nelissen, B., Goffeau, A., Saier, M.H. Jr. (1998). Unified inventory of established and putative transporters encoded within the complete genome of Saccharomyces cerevisiae. *FEBS Letters* (1-2), 116-125.

Petersen, T.N., Brunak, s., von Heijne, G., Nielsen H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* (8), 785-786.

Petter, M., Haeggström, M., Khattab, A., Fernandez, V., Klinkert, M.Q., Wahlgren, M. (2007). Variant proteins of the *Plasmodium falciparum* RIFIN family show distinct subcellular localization and developmental expression patterns. *Molecular and Biochemical Parasitology* (1), 51-56.

Phillips, R.S. (2001). Current Status of malaria and potential for control. *Clinical Microbiology Reviews* (14), 208-226.

Ponpuak, M., Klemba, M., Park, M., Gluzman, I.Y., Lamppa, G.K., Goldberg, D.E. (2007). A role for falcilysin in transit peptide degradation in the *Plasmodium falciparum* apicoplast. *Molecular Microbiology* (2), 314-334.

Promeneur, D., Liu, Y., Maciel, J., Agre, P., King, L.S., Kumar, N. (2007). Aquaglyceroporin PbAQP during intraerythrocytic development of the malaria parasite *Plasmodium berghei*. *Proceedings of the National Academy of Sciences USA* (7), 2211-2216.

Raj, D.K., Mu, J., Jiang, H., Kabat, J., Singh, S., Sullivan, M., Fay, M.P., McCutchan, T.F., and Su, X.Z. (2008). Disruption of a *Plasmodium falciparum* multidrug resistance-associated protein (PFMRP) alters its fitness and transport of

antimalarial drugs and glutathione. *Journal of Biological Chemistry* (12), 7687-7696.

Ralph, S.A. (2007). Subcellular multitasking – multiple destinations and roles for the *Plasmodium* falcilysin protease. *Molecular Microbiology* (2), 309-313.

Rapp, M., Drew, D., Daley, D.O., Nilsson, J., Carvalho, T., Melén, K., de Gier, J.W., von Heijne, G. (2004). Experimentally based topology models for *E. coli* inner membrane proteins. *Protein Science* (4), 937-945.

Rath, A., Glibowicka, M., Nadeau, V.G., Chen, G., Deber, C.M. (2008). Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proceedings of the National Academy of Sciences USA* (6), 1760-1765.

Read, A.F., Lynch, P.A., Thomas, M.B. (2009). How to make evolution-proof insecticides fo malaria control. *PLoS Biology* (7), e1000058.

Reece, S.E., Thompson, J. (2008). Transformation of the rodent malaria parasite *Plasmodium chabaudi* and generation of a stable fluorescent line PcGFPCON. *Malaria Journa* (7), 183.

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.

Rexach, J., Fernandez, E., Galvlan, G. (2000). The *Chlamydomonas* reinhardtii Nar 1 gene encodes a chloroplast membrane protein involved in nitrite transport. *The Plant Cell* (12), 1441-1453.

Roberts, L.S., Janovy, J. (2006). *Foundations of Parasitology 7th Edition.*: McGraw Hill, London.

Roepe, P.D., Ferdig, M.T. (2009). *P. falciparum* Na+ / H+ Exchanger (PfNHE) Function and Quinine Resistance (QNR). *Molecular and Biochemical Parasitology* (1), 1-3.

Rosenthal, P.J., Meshnick, S.R. (1996). Haemoglobin catabolism and iron utilization by malaria parasites. *Molecular and Biochemical Parasitology* (83), 131-139.

Ross, R. (1902). Researches on Malaria. *Nobel Lectures, Physiology or Medicine 1901-1921.*: Elsevier Publishing Company. Amsterdam, 1967.

Rossmann, R., Sawers, G., Böck, A. (1991). Mechanism of regulation of the formate-hydrogenlyase pathway by oxygen, nitrate, and pH: definition of the formate regulon. *Molecular Microbiology* (11), 2807-2814.

Rycovska, A., Hatahet, L., Fendler, K., Michel, H. (2012). The nitrite transport protein NirC from *Samlonella typhimurium* is a nitrite/proton antiporter. *Biochima et Biophysica Acta* (1818), 1342-1350.

Saier, M.H. Jr., Eng, B.H., Fard, S., Garg, J., Haggerty, D.A., Hutchinson, W.J., Jack, D.L., Lai, E.C., Liu, H.J., Nusinew, D.P., Omar, A.M., Pao, S.S., Paulsen, I.T., Quan, J.A., Sliwinski, M., Tseng, T.T., Wachi, S., Young, G.B. (1999). Phylogenetic characterization of novel transport protein families revealed by genome analyses. *Biochimica et Biophysica Acta* (1), 1-56.

Salcedo-Sora, J.E., Ochong, E., Beveridge, S., Johnson, D., Nzila, A., Biagini, G.A., Stocks, P.A., O'Neill, P.M., Krishna, S., Bray, P.G., Ward, S.A. (2011). The molecular basis of folate salvage in *Plasmodium falciparum*: characterization of two folate transporters. *Journal of Biological Chemistry* (52), 44659-44668.

Sanchez, C.P., Dave, A., Stein, W.D., Lanzer, M. (2010). Transporters as mediators of drug resistance in *Plasmodium falciparum*. *International Journal of Parasitology* (10), 1109-1118.

Saridaki, T., Sanchez, C.P., Pfahler, J., Lanzer, M. (2008). A conditional export system provides new insights into protein export in *Plasmodium falciparum*-infected erythrocytes. *Cellular Microbiology* (12), 2483-2495.

Sawers, G. (2005). Formate and its role in hydrogen production in *Escherichia coli*. *Biochemical Society Transactions* (Pt 1), 42-46.

Schauer, N.L., Brown, D.P., Ferry, J.G. (1982). Kinetics of formate metabolism in *Methanobacterium formicicum* and *Methanospirillum hungatei*. *Applied and Environmental Microbiology* (3), 549-554.

Scheibel, L.W., Pflaum W.K. (1970). Carbohydrate metabolism in *Plasmodium knowlesi*. *Comparative Biochemistry and Physiology* (37), 543-553.

Scherf. A, Lopez-Rubio, J.J., Riviere, L. (2008). Antigenic variation in *Plasmodium falciparum*. *Annual Reviews of Microbiology* (62), 45-47.

Schreiber, N., Brattig, N., Evans, J., Agbenyega, T., Horstmann, R.D., May, J., Klinkert, M. (2006). Cerebral malaria s associated with IgG2 and IgG4 antibody

responses to recombinant *Plasmodium falciparum* RIFIN antigen. *Microbes and Infection* (8), 1269-1279.

Serirom, S., Raharjo, W.H., Chotivanich, K., Loareesuwan, S., Kubes, P., Ho, M. (2003). Anti-adhesive effect of nitric oxide on *Plasmodium falciparum* cytoadherence under flow. *The American Journal of Pathology* (5), 1651-1660.

Silvie, O., Franetich, J.F., Rénia, L., Mazier, D. (2004). Malaria sporozoite: migrating for a living. *Trends in Molecular Medicine* (3), 97-100.

Sinden, R.E., Canning, E.U., Bray, R.S., Smalley, M.E. (1978). Gametocyte and gamete development in *Plasmodium falciparum*. *Proceedings of the Royal Society London. Series B: Biological Sciences* (1145), 375-399.

Singh, G., Prakash, S. (2011). Evaluation of culture filtrates of *Culicinomyces clavisporus*: Mycoadulticide for *Culex quinquefasciatus*, *Aedes aegypti* and *Anopheles stephensi*. *Parasitology Research* (1), 267-272.

Slavic, K., Straschil, U., Reininger, L., Doerig, C., Morin, C., Tewari, R., Krishna, S. (2010). Life cycle studies of the hexose transporter of *Plasmodium* species and genetic validation of their essentiality. *Molecular Microbiology* (6), 1402-1413.

Spycher, C., Rug, M., Klonis, N., Ferguson, D.J.P., Cowman, A.F., Beck, H.P., Tilley, L. (2006). Genesis of and Trafficking to the Maurer's Clefts of *Plasmodium falciparum*-Infected Erythrocytes. *Molecular and Cellular Biology* (11), 4074-4085.

Staines, H.M., Powell, T., Thomas, S.L., Ellory, J.C. (2004). *Plasmodium falciparum*-induced channels. *International Journal of Parasitology* (6), 665-673.

Stark, M.J. (1987). Multicopy expression vectors carrying the lac repressor gene for regulated high-level expression of genes in *Escherichia coli*. *Gene* (2-3), 255-267.

Sturm, A., Amino, R., van de Sand, C., Regen, T., Retzlaff, S., Rennenberg, A., Krueger, A., Pollok, J.M., Ménard, R., Heussler, V.T. (2006). Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids. *Science* (5791), 1287-1291.

Suppmann, B., Sawers, G. (1994). Isolation and characterisation of hypophosphite resistant mutants of *Eschericia coli*: identification of the FocA protein encoded

by the *pfl* operon, as a putative formate transporter. *Molecular Microbiology* (11), 965-982.

Symington, V.F. (2009). *Structure and function of nitrate and nitrite transporters, NrtA and NitA, from Aspergillus nidulans.* (PhD.), University of St Andrews.

Thomas, T.C., McNamee, M.G. (1990). Purification of Membrane Proteins. In J. N. Abelson, Simon, M.I., Deutscher, M.P. (Ed.), *Guide to Protein Purification, Volume 182 (Methods in Enzymology).* Academic Press,.

Tonkin, C.J., Carret, C.K., Duraisingh, M.T., Voss, T.S., Ralph, S.A., Hommel, M., Duffy, M.F., Silva, L.M., Scherf, A., Ivens, A., Speed, T.P., Beeson, J.G., Cowman, A.F. (2009). *Sir2* paralogues cooperate to regulate virulence genes and antigenic variation in *Plasmodium falciparum*. *PLoS Biology* (4), e84.

Tonkin, C.J., van Dooren, G.G., Spurck, T.P., Struck, N.S., Good, R.T., Handman, E., Cowman, A.F. and McFadden, G.I. (2004). Localization of organellar proteins in *Plasmodium falciparum* using a novel set of transfection vectors and a new immunofluorescence fixation method. *Molecular and Biochemical Parasitology* (137), 13-21.

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

Tren, R., Bate, R. (2000). *When politics kills: Malaria and the DDT story*. Retrieved 05 July 2012, from Competitive Enterprise Institute. http://cei.org/PDFs/malaria.pdf.

Tusnády, G.E., Simon, I. (1998). Principles Governing Amino Acid Composition of Integral Membrane Proteins: Applications to Topology Prediction. *Journal of Molecular Biology* (238), 489-506.

Tuteja, R. (2007). Malaria - an overview *FEBS Journal.*(18), 4670-4679.

Umlas, J. Fallon, J.N. (1971). New thick-film technique for malaria diagnosis. Use of saponin stromatolytic solution for lysis. *The American Journal of Tropical Medicine and Hygiene* (20), 527-529.

Unkles, S.E., Karabika, E., Symington, V.F., Cecile, J.L., Rouch, D.A., Akhtar, N., Cromer, B.A., Kinghorn, J.R. (2012). Alanine scanning mutagenesis of a high-affinity nitrate transporter highlights the requirement for glycine and asparagine

residues in the two nitrate signature motifs. *The Biochemical Journal* (BJ20120631).

Unkles, S.E., Symington, V.F., Kotur, Z., Wang, Y., Siddiqi, M.Y., Kinghorn, J.R., Glass, A.D. (2011). Physiological and biochemical characterization of AnNitA, the *Aspergillus nidulans* high-affinity nitrite transporter. *Eukaryotic Cell* (12), 1724-1732.

Valderramos, S.G., Scanfeld, D., Uhlemann, A.C., Fidock, D.A., Krishna, S. (2010). Investigations into the role of the *Plasmodium falciparum* SERCA (PfATP6) L263E mutation in artemisinin action and resistance. *Antimicrobial Agents and Chemotherapy* (9), 3842-3852.

van Dooren, G.G., Marti, M., Tonkin, C.J., Stimmler, L.M., Cowman, A.F., McFadden, G.I. (2005). Development of the endoplasmic reticulum, mitochondrion and apicoplast during the asexual life cycle of *Plasmodium falciparum*. *Molecular Microbiology* (2), 104-419.

Vanderberg, J.P., Weiss, M.M., Mack, S.R. (1977). In vitro cultivation of the sporogonic stages of *Plasmodium*: a review. *Bulletin of the World Health Organization* (2-3), 377-392.

Vickerman, K. (2005). The lure of life cycles: Cyril Garnham and the malaria parasites of primates. *Protist* (156), 433-499.

von Heijne, G. (1992). Membrane Protein Structure Prediction: Hydrophobicity Analysis and the 'Positive Inside' Rule. *Journal of Molecular Biology* (225), 487-494.

Waight, A.B., Love, J., Wang, D.N. (2010). Structure and mechanism of a pentameric formate channel. *Nature Structural and Molecular Biology* (1), 31-37.

Waldo, G.S., Standish, B.M., Berendzen, J., Terwilliger, T.C. (1999). Rapid protein-folding assay using green fluorescent protein. *Nature Biotechnology* (7), 691-695.

Waller, K.L., Muhle, R.A., Ursos, L.M., Horrocks, P., Verdier-Pinard, D., Sidhu, A.B., Fujioka, H., Roepe, P.D., Fidock, D.A. (2003). Chloroquine resistance modulated in vitro by expression levels of the *Plasmodium falciparum* chloroquine resistance transporter. *Journal of Biological Chemistry* (35), 33593-33601.

Wang, C.W., Magistrado, P.A., Nielsen, M.A., Theander, T.G., Lavstsen, T. (2009b). Preferential transcription of conserved rif genes in two phenotypically distinct *Plasmodium falciparum* parasite lines. *International Journal of Parasitology* (6), 655-664.

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

Wang, Y., Huang, Y., Wang, J., Cheng, C., Huang, W., Lu, P., Xu, Y.N., Wang, P., Yan, N., Shi, Y. (2009a). Structure of the formate transporter FocA reveals a pentameric aquaporin-like channel. *Nature* (7272), 467-472.

Wang, Y., Li, W., Siddiqi, Y., Symington, V.F., Kinghorn, J.R., Unkles, S.E., Glass, A.D. (2008). Nitrite transport is mediated by the nitrite-specific high-affinity NitA transporter and by nitrate transporters NrtA, NrtB in *Aspergillus nidulans*. *Fungal Genetics and Biology* (2), 94-102.

Warrens, A.N., Jones, M.D., Lechler, R.I. (1997). Splicing by overlap extension by PCR using asymmetric amplification: an improved technique for the generation of hybrid proteins of immunological interest. *Gene* (1), 29-35.

Waters, A. P. (2002). Orthology between the genomes of *Plasmodium falciparum* and rodent malaria parasites: possible practical applications. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences* (357), 55-63.

Weinberg, J.B., Lopansri, B.K., Mwaikambo, E., Granger, D.L. (2008). Arginine, nitric oxide, carbon monoxide, and endothelial function in severe malaria. *Current Opinion in Infectious Disease* (5), 468-475.

White, N.J. (2004). Antimalarial drug resistance. *The Journal of Clinical Investigation* (113), 1084-1092.

White, W.B., Ferry, J.G. (1992). Identification of formate dehydrogenase-specific mRNA species and nucleotide sequence of the *fdhC* gene of *Methanobacterium formicicum*. *Journal of Bacteriology* (15), 4997-5004.

Whitley, P., Zander, T., Ehrmann, M., Haardt, M., Bremer, E., von Heijne, G. (1994). Sec-independent translocation of a 100-residue periplasmic N-terminal tail in the *E. coli* inner membrane protein proW. *The EMBO Journal* (19), 4653-4651.

Wilde, O. (n.d.). goodreads.com. Retrieved August 9, 2012, from goodreads.com. Web site: http://www.goodreads.com/quotes/248952-the-smallest-act-of-kindness-is-worth-more-than-the

Winzeler, E.A. (2008). Malaria research in the post-genomic era. *Nature* (455), 751-756.

Woodrow, C.J., Penny, J.I., Krishna, S. (1999). Intraerythrocytic *Plasmodium falciparum* expresses a high affinity facilitative hexose transporter. *The Journal of Biological Chemistry* (11), 7272-7277.

Wu, Y., Sifri, C.D., Lei, H.H., Su, X.Z. and Wellems, T.E. (1995). Transfection of *Plasmodium falciparum* within human red blood cells. *Proceedings of the National Academy of Sciences USA* (92), 973-977.

Zhang, X., Tolzmann, C.A., Melcher, M., Haas, B.J., Gardner, M.J., Smith, J.D., Feagin, J.E. (2011). Branch point identification and sequence requirements for intron splicing in *Plasmodium falciparum*. *Eukaryotic Cell* (11), 1422-1428.

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* (9), 40.

Zhu, G., Marchewka, M.J., Keithly, J.S. (2000). *Cryptosporidium parvum* appears to lack a plastid genome. *Microbiology* (146), 315-321.

Zuegge, J., Ralph, S., Schmuker, M., McFadden, G.I., Schneider, G. (2001). Deciphering apicoplast targeting signals - feature extraction from nuclear-encoded precursors of *Plasmodium falciparum* apicoplast proteins. *Gene* (280), 19-26.

Appendix 1: Bacterial Growth Media

2xTY	16 g/L tryptone, 10 g/L yeast extract, 5 g/L sodium chloride.
JC Minimal Salts Medium	4 g/L potassium dihydrogen phosphate, 10.5 g/L potassium hydrogen phosphate, 1 g/L ammonium sulphate, 0.05 g/L magnesium chloride, 2.5 g/L LB, 10 µM sodium molybdate, 10 µM sodium selenate, 10 ml/L sulphur free salts. (Jia *et al*, 2009)
LB Agar	1.2 % agar (w/v) in LB medium.
Luria Bertani (LB) medium	25 g/L LB powder (Invitrogen).
M9 Media	200 ml/L M9 minimal salts, 2 mM magnesium sulphate, 0.1 mM calcium chloride, 0.4 % (w/v) glycerol, pH 6.8.
M9 Minimal Salts	64 g/L disodium hydrogen phosphate, 15 g/L potassium dihydrogen phosphate, 12.5 g/L sodium chloride, 25 g/L ammonium chloride.
Super Optimal Broth (SOB)	20 g/L tryptone, 5 g/L yeast extract, 10 mM sodium chloride, 2.5 mM potassium chloride, 0.5 mM magnesium sulphate, 0.5 mM magnesium chloride.
SOB with Catabolite repression	SOB medium with 20 mM glucose.

Sulphur free salts	82 g/L magnesium chloride, 10 g/L manganese chloride, 4 g/L iron(III)chloride, 1 g/L calcium chloride, 20 ml hydrochloric acid (Cole *et al*, 1974).
TGYEP	10 g/L tryptone, 5 g/L yeast extract, 5 g/L dipotassium hydrogen phosphate, 12 g/L potassium dihydrogen phosphate, pH 6.7.
Werkmans Minimal Medium	50 mM disodium hydrogen phosphate, 50 mM potassium dihydrogen phosphate, 1 mM magnesium sulphate, 0.1 mM calcium chloride, 15 mM ammonium sulphate, 80 mM glucose, pH 6.5 (Suppmann and Sawers, 1994).

Appendix 2: Oligonucleotide Primers

Name	Sequence (5'->3')	Insert/Purpose	L	RES
NitAopt F	CGAACTGCAGTTCCCTTTCTTC	Full-length Opt NitA for pV5GPD	22	*Pst*I
Y212K R	CGGAGAAGACTTTGCCAGCGCCGTC	Opt NitA Y212K mutation	25	
Y212K F	GACGGCGCTGGCAAAGTCTTCTCCG	Opt NitA Y212K mutation	25	
NitAopt R	ACGGTATCGATAAGCTTGATATCGAATTCGACGTCCTAGGT	Full-length Opt NitA for pV5GPD	41	*Cla*I
PcCir2Dom F	GCACCCATGGCCAAAGATCTGTGCGAT	PcCir2 Domain for pEHISTEV and pEHISGEFTEV	27	*Nco*I
PcCir2Dom R	CGTGCTCGAGTCAGTTTTTAAAGTTATCATAATCTTTG	PcCir2 Domain for pEHISTEV and pEHISGEFTEV	38	*Xho*I
PcCir2Dom2 F	GCACCCATGGCAAACAACGATAGCCAGAAA	PcCir2 Domain2 for pEHISTEV and pEHISGEFTEV	30	*Nco*I
PcCir GFPe F	GCAGCTCGCGGATGGCCAAAGATCTGTGCGATG	Opt Cir2 for pGFPe and pHA-1	33	*Xho*I
PcCir GFPe R	CGTCGGTACCCCCACATAGCTCGCCATTTTG	Opt Cir2 for pGFPe and pHA-1	31	*Kpn*I
PcCir2 40 R	GCTCGGTACCCCGTTGGTCGGGCAATAGG	Opt Cir2 40 for pGFPe and pHA-1	29	*Kpn*I
PcCir2 106 R	GCTCGGTACCCCAAAGGTCTGGTTCGGATG	Opt Cir2 106 for pGFPe and pHA-1	30	*Kpn*I
PcCir2 210 R	CGTCGGTACCCCGTTTTTAAAGTTATCATAATCTTTC	Opt Cir2 210 for pGFPe and pHA-1	37	*Kpn*I
PcCir2 254 R	GCTCGGTACCCCGCTAATATCTTCGCTGTTATG	Opt Cir2 256 for pGFPe and pHA-1	33	*Kpn*I
PcCir2 Δ73 F	CCACCTCGCGGATGAAAAACAACGATAGCCAG	Opt Cir2 Δ73 for pGFPe and pHA-1	32	*Xho*I
NitAopt TTQ F	CCAGGAATTCGCATATGCCGCCCAACAACAGCAAA	Full-length Opt NitA for TTQ18	35	*Eco*RI
NitAopt V5 TTQ R	GAGGAAGCTTCTAGGTGGAGTGAAGACCGAGGAG	Full-length Opt NitA with V5 tag for TTQ18	34	*Hind*III
FocA F	CCGGAATTCGCATATGAAAGCTGACAACCCTTTTG	Full-length EcFocA for TTQ18	35	*Eco*RI
FocA R	AAAACTGCAGCATGGTGGTCGTTTTCACGCAGG	Full-length EcFocA for TTQ18	33	*Pst*I
NirC F	CCGGAATTCGCATATGTTACAGACACTAT	Full-length EcNirC for TTQ18	29	*Eco*RI
NirC R	AAAACTGCAGCACCGGCAGCCGTTTCAGTT	Full-length EcNirC for TTQ18	30	*Pst*I
5′ KO check int F	GCATAAGCTTGTACAATTGAGTTAAGTGGATGG	Check integration of 5′ pCC KO	33	*Hind*III
5′ KO check int R	CCAATAGATAAAATTTGTAGAG	Check integration of 5′ pCC KO	22	
3′ KO check int F	CATATGTTAAATATTTATTTCTC	Check integration of 3′ pCC KO	23	
3′ KO check int R	GGACTCTAGAGGATATCAGCGCAAACATGTGATGGA	Check integration of 3′ pCC KO	36	*Xba*I
PfNitA KO 750 5 F	CAGGCCGCGGGAATGAATACATACTAAGAGCAGC	PfNitA 5' 724 bp for pCC1/4	34	*Sac*II

Name	Sequence (5'->3')	Insert/Purpose	L	RES
PfNitA KO 750 5 R	GAACCTTAAGGTGCTCATGTCAAATGGGTAC	PfNitA 5' 724 bp for pCC1/4	31	*Afl*II
PfNitA KO 750 3 F	CGAGCCATGGGCTGGTGCAATTGTTTTGGG	PfNitA 3' 756 bp for pCC1/4	30	*Nco*I
PfNitA KO 750 3 R	GATCCCTAGGGGTGATTTGAACAACAGGACAATG	PfNitA 3' 756 bp for pCC1/4	34	*Avr*II
PfNitA KO 500 5 F	CAGGCCGCGGGCCACCAAATAATTCCAAATATG	PfNitA 5' 508 bp for pCC1/4	33	*Sac*II
PfNitA KO 500 5 R	GAACCTTAAGCTCATCGATGCTCCTACGATTTC	PfNitA 5' 508 bp for pCC1/4	33	*Afl*II
PfNitA KO 500 3 F	CGAGCCATGGGTTAGGAAATTACTTAAGCTC	PfNitA 3' 542 bp for pCC1/4	31	*Nco*I
PfNitA KO 500 3 R	GATCCCTAGGGTATTTTATATGCCATTAGCTC	PfNitA 3' 542 bp for pCC1/4	32	*Avr*II
PfNitACtml GW F	CACCCTTATTTATCCTGGAGCATTTAC	C - terminal PfNitA for Gateway™ Cloning	27	
PfNitA GW R	ATTTCGTAATTCTATAGATAAA	C - terminal PfNitA for Gateway™ Cloning	22	
PfNitA F1	CACAAATTAAGAGAAATCGTAGG	Check integration of C-terminal PfNitA fragment	23	
GW int R1	CGTTTTACAACGTCTCGAGGG	Check integration of C-terminal PfNitA fragment	21	
GFP mut2 R	CGCGAAAGTAGTGACAAGTG	Check integration of C-terminal PfNitA fragment with GFP tag	20	
NitAopt GW F	CACCATGCCGCCCAACAACAG	Full-length Opt NitA for Gateway™ Cloning	21	
NitAopt GW R	GGTGGAGTCAAGACCGAG	Full-length Opt NitA for Gateway™ Cloning	18	
CD F	ATCCCTTCCTTACTACAGAATAGG	Amplification of CD	24	
CD R	ATCAGATTGTTGGTTGAAGAAGG	Amplification of CD	23	
hDHFR F	GCGCGGATCCATGCATGGTTCGCTAAAC	Amplification of hDHFR	28	
hDHFR R	GCGCGCTTAATCATTCTTCTCATATAC	Amplification of hDHFR	27	
PfNitA pHH1 KO F	GCAGAGATCTGACTTTTATATTATAAAGTGTATGC	Amplification of NitA for pHH1 based knock-out and control	35	*Bgl*II
PfNitA pHH1 KO R	GCAGCTCGAGTTAGATAAAATATATGATAAAATATATTCCTATTGG	Amplification of NitA for pHH1 based knock-out	46	*Xho*I
PfNitA pHH1 KOcon R	GCAGCTCGAGGCCATTAGCTCAATTTCGTAATTCTATAG	Amplification of NitA for pHH1 based knock-out control	39	*Xho*I

L= number of nucleotides. RES = Restriction Endonuclease Site, underlined in sequence.

Appendix 3: Accession Numbers

Accession numbers for the genes referred to in this text were obtained from the National Centre for Biotechnology Information (NCBI)[49] or PlasmoDB[50] databases and are as follows:

Organism	Protein	Number
P. falciparum	NitA, putative formate-nitrite transporter	PFC0725c
E. coli	FocB	CAQ32863.1
P. knowlesi	putative transporter	PKH_082480
P. berghei	putative formate-nitrite transporter	PBANKA_041440
P. chabaudi	putative formate-nitrite transporter	PCHAS_041530
P. yoelii	putative formate-nitrite transporter	PY06388
P. vivax	putative transporter	PVX_095405
E. coli	FocA	CAQ31432.1
H. influenzae	formate transporter	YP_247898.1
S. typhimurium	NirC	AAA27040.1
E. coli	NirC	CAQ33687.1
S. boydii	NirC	YP_001882042.1
B. subtilis	formate-nitrite transporter	NP_391685.1
S. aureus	formate-nitrite transporter	YP_039761.1
C. sticklandii	fdhc protein	YP_003937095.1
M. thermautotrophicus	fdhc protein	AAC44819.1
M. formicicim	probable formate transporter	P35839.1
B. cereus	fdhC gene product	YP_085067.1
C. albicans	transporter family	CAA21934.1
S. cerevisiae	YHL008C	AAS56428.1
T. gondii	putative formate-nitrite transporter	XP_002369729.1
E. gracilis	FocA-like	AAS66886.1
B. bovis	formate-nitrite transporter family protein	XP_001608703.1
T. annulata	formate-nitrite transporter	XP_951799.1
C. reinhardtii	formate-nitrite transporter Nar 1.3	XP_001701278.1
C. reinhardtii	formate-nitrite transporter Nar 1.4	XP_001700148.1
V. cholerae	Chain D, Pentameric Formate Channel	3KLZ_D
A. nidulans	formate-nitrite family transporter	CBF78263.1
C. difficile	Chain E, Crystal Structure Of Hsc K16s	3TE2_E
S. typhimurium	formate transporter	YP_005232043.1
P. chabaudi	Cir2	PCHAS_000290

[49] http://www.ncbi.nlm.nih.gov/guide/. Accessed on 10 August 2012.
[50] http://plasmodb.org/plasmo/. Accessed on 10 August 2012.

Appendix 4: Abbreviations

General Abbreviations

APS	ammonium persulphate
ATP	adenosine triphosphate
BCA	bicinchoninic acid
BCIP	5-bromo-4-chloro-3-indolyl phosphate
BCKDH	branched chain alpha-keto acid dehydrogenase
bis	N, N' – methylene bisacrylamide
bp	base pair (s)
Bla	blasticidin
BSA	bovine serum albumen
BSD	blasticidin-S-deaminase
Bs	*Bacillus sphaericus*
bti	*Bacillus thuringiensis var. israelensis*
CBB	Coomassie Brilliant Blue
CD	circular dichroism spectrophotometry
cDNA	complementary deoxyribonucleic acid
C-terminal	carboxy terminal
CQ	chloroquine
^{o}C	degrees centigrade
Da	Daltons
DABCO	1,4-Diazabicyclo[2.2.2]octane
DAPI	4', 6-diamidino-2-phenylindole dihydrochloride
DDT	dichlorodiphenyltrichlorethane
DHFR	dihydrofolate reductase
DHOD	dihydroorotate dehydrogenase
DHPS	dihydropteroate synthase
DIC	differential interference contrast
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
DSM	drug selectable marker
DTT	dithiothreitol
DV	digestive vacuole
DVM	digestive vacuole membrane

DW	distilled water
ECL	enhanced chemiluminescence
EDTA	ethylenediaminetetraacetic acid
EGTA	ethyleneglycoltetraacetic acid
ETC	electron transport chain
FITC	Fluorescein isothiocyanate
FNT	formate-nitrite transporter
g	gram (s)
g	gravity
gDNA	genomic deoxyribonucleic acid
GFP	green fluorescent protein
GOI	gene of interest
h	hour (s)
HAc	haemagglutinin
HAT	human African trypanosomiasis
Hb	haemoglobin
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His	Histadine (tag)
HVR	hypervariable region
IB (s)	inclusion body (ies)
IDC	intraerythrocytic developmental cycle
IE	infected erythrocyte (s)
IFA	immunofluorescence analysis
IPP	isopentenyl pyrophosphate
IPTG	isopropyl β-D-1-thiogalactopyranoside
iRBC	infected red blood cell (s)
k	kilo
kb	kilobases
KO	knock out
kpsi	kilo pound-force per square inch
L	litre
LB	Luria Bertani
LBA	long branch attraction
LDS	lithium dodecyl sulphate
M	molar
MCS	multicloning site
MCT	monocarboylate transporter

µg	micro gram (s)
µl	micro litre (s)
µM	micro Molar
mg	milli gram (s)
ml	milli litre (s)
mM	milli Molar
min (s)	minute (s)
MOPS	3-(N-morpholino) propanesulfonic acid
MQ	milliQ
NED	N-1-naphthylethylenediamine dihydrochloride
nm	nanometers
NO	nitric oxide
N-terminal	amino terminal
OD	optical density
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEP	phospho*enol*pyruvate
PEXEL	*Plasmodium* export element
PhoA	alkaline phosphatase
PIPES	2-[4-(2-sulfoethyl)piperazin-1-yl]ethanesulfonic acid
PIR	*Plasmodium* interspersed repeats
pNPP	p-nitrophenyl phospate
PPM	parasite plasma membrane
PV	parasitophorus vacuole
PVM	parasitophorus vacuole membrane
PVDF	polyvinylidene fluoride
PVM	parasitophorus vacuole membrane
RBC	red blood cell (s)
RBCm	red blood cell membrane
rpm	revolutions per minute
RPMI	Roswell Park Memorial Institute
RT	room temperature
s	second (s)
SDS	sodium dodecyl sulphate
SEM	standard error of the mean

SOB	super optimal broth
SOC	super optimal broth with catabolite repression
SOD	superoxide dismutase
TAE	trisacetate-EDTA
TBS	tris-buffered saline
TBST	tris-buffered saline tween-20
TCA	tricarboxylic acid
TE Buffer	tris-EDTA
TEMED	N, N, N', N' – tetramethylethylenediamine
TEV	tobacco etch virus (protease site)
Tris	tris(hydroxymethyl)aminomethane
TRITC	tetramethyl rhodamine iso-thiocyanate
Tween-20	polyoxyethylene sorbitan
U	unit (s)
UTR	untranslated region
UV	ultraviolet
VSA	variant surface antigen (s)
VTS	vacuolar targeting sequence
v/v	volume per volume
w/v	weight per volume
WHO	World Health Organisation
WR 99210	Walter Reed Institute anti-DHFR antifolate 99210
2D	two dimensional
3D	three dimensional
3'	three prime
5'	five prime

Protein Abbreviations

PfAQP	*Plasmodium falciparum* aquaglyceroporin
PfCRT	*Plasmodium falciparum* chloroquine resistance transporter
PfEMP-1	*Plasmodium falciparum* erythrocyte membrane protein-1
PfFT1 and 2	*Plasmodium falciparum* folate transporters 1 and 2
PfMDR	*Plasmodium falciparum* multi drug resistance transporter
PfMRP	*Plasmodium falciparum* multi resistance-associated protein
PfNHE	*Plasmodium falciparum* Na^+/H^+ exporter
PfHT	*Plasmodium falciparum* hexose transporter
PfNitA	*Plasmodium falciparum* formate/nitrite transporter, putative
PfNT/PfENT	*Plasmodium falciparum* nucleoside transporter
YHL008c	FNT protein of *S. cerevisiae*
PfNitA	FNT protein of *Plasmodium falciparum*
NirC	FNT proteins of *E. coli* and *Salmonella* and *Thermofilum pendens*
Nar1	FNT protein of *Chlamydomonas reinhardtii*
PcCir2	*Plasmodium chabaudi* Cir2
FocA/B	FNT protein of *E. coli* and *V. cholerae*
EgFth	FNT protein of *Euglena gracilis*

Amino Acid Abbreviations

A	ala	alanine
R	arg	arginine
N	asn	asparagine
D	asp	aspartate
C	cys	cysteine
E	glu	glutamate
Q	gln	glutamine
G	gly	glycine
H	his	histidine
I	ile	isoleucine
L	leu	leucine
K	lys	lysine
M	met	methionine
F	phe	phenylalanine
P	pro	proline
S	ser	serine
T	thr	threonine
W	trp	tryptophan
Y	try	tyrosine
V	val	valine

Appendix 5: Data Collected from BLASTP Database Searches.

Query: *S. cerevisiae*, AAS56428.1.

Species	Gene Anotation	e-value	bit score	max ident
Ogataea parapolymorpha	formate transporter 1, putative	8e-111	352	57 %
Candida glabrata	hypothetical protein	0.0	542	47 %
Candida albicans	transporter family	8e-116	365	45 %
Spathaspora passalidarum	putative formate transporter 1	8e-128	395	44 %
Wickerhamomyces ciferrii	putative transporter	5e-139	429	40 %

Query: *P. falciparum,* PFC0725c.

Species	Gene Anotation	e-value	bit score	max ident
P. vivax	transporter, putative	6e-148	421	78 %
P. knowlesi	transporter, putative	9e-162	458	73 %
P. berghei	transporter, putative	1e-154	440	71 %
P. yoelli	formate/nitrite transporter, putative	2e-153	437	70 %
P. chabaudi	transporter, putative	2e-74	230	67 %

Query: *E. coli* NirC, CAQ33687.1.

Species	Gene Anotation	e-value	bit score	max ident
Shigella flexneri	putative nitrite transporter	0.0	533	99 %
Salmonella enterica	Nitrite transporter NirC	3e-177	500	93 %

Query: *E. coli* FocA, CAQ31432.1.

Species	Gene Anotation	e-value	bit score	max ident
Shigella flexneri	putative formate transporter 1	0.0	577	99 %
Salmonella enterica	formate transporter	0.0	562	96 %

Appendix 6: Synthetic Codon-Optimised Gene Sequences

pfnita_optEc

ATGCCTCCGAATAATTCAAAATATGTTCTGGATCCGGTGAGCATTAAAAGCGTTTGTGGTGGTGAAGAAA
GCTATATTCGTTGCGTGGAATATGGCAAAAAAAAAGCCCATTATAGCAACCTGAATCTGCTGGCAAAAGC
AATTCTGGCAGGTATGTTTGTTGGTCTGTGTGCACATGCAAGCGGTATTGCCGGTGGTCTGTTCTATTAT
CATAAACTGCGTGAAATTGTTGGTGCCAGCATGAGCGTTTTTGTTTATGGTTTTACCTTTCCGATCGCCT
TCATGTGTATTATTTGTACCGGTAGCGACCTGTTTACAGGTAATACCCTGGCAGTTACCATGGCACTGTA
TGAGAAAAAAGTTAAACTGCTGGATTATCTGCGCGTGATGACCATTAGCCTGTTTGGTAATTATGTGGGT
GCAGTTAGCTTTGCCTTTTTTGTTAGCTATCTGAGCGGTGCCTTTACCAATGTTCATGCCGTTGAAAAAA
ACCATTTTTTTCAGTTTCTGAACGATATCGCCGAAAAAAAAGTGCATCACACCTTTGTTGAATGTGTTAG
CCTGGCCGTTGGTTGCAACATTTTTGTTTGTCTGGCCGTTTATTTTGTGCTGACCCTGAAAGATGGTGCA
GGTTATGTGTTTAGCGTGTTTTTTGCCGTTTATGCATTTGCCATTGCCGGTTATGAACATATTATCGCCA
ACATCTATACCCTGAATATTGCCCTGATGGTGAACACCAAAATTACCGTTTATCAGGCCTACATCAAAAA
TCTGCTGCCGACCCTGCTGGGTAACTATATTGCGGGTGCAATTGTTCTGGGTCTGCCGCTGTATTTTATC
TATAAAGAGCACTACTATAACTTTGAACGCAGCAAACGCGATAATAATGATGCACAGATGAAAAGCCTGA
GCATTGAACTGCGTAAT

pfnita_opt

ATGCCGCCCAACAACAGCAAATACGTCCTTGATCCCGTCAGCATCAAATCCGTCTGCGGCGGCGAGGAGA
GCTACATCCGATGCGTCGAATATGGTAAGAAGAAAGCCCACTATTCCAACCTGAACCTGCTCGCTAAGGC
TATCCTCGCCGGCATGTTCGTCGGCCTCTGCGCCCACGCCTCCGGCATCGCCGGCGGCCTCTTCTATTAC
CACAAGCTCCGCGAGATCGTCGGCGCCTCTATGTCCGTCTTCGTTTACGGTTTCACATTCCCTATCGCCT
TCATGTGCATCATCTGCACCGGTTCCGATCTCTTCACCGGTAACACTCTTGCCGTTACGATGGCTCTCTA
CGAGAAGAAGGTCAAGCTCCTGGACTACCTGCGCGTTATGACTATCTCTCTTTTCGGCAACTACGTCGGC
GCCGTCTCCTTCGCTTTCTTCGTCTCCTACCTCTCTGGCGCCTTCACCAACGTTCACGCCGTCGAGAAGA
ACCACTTCTTCCAATTCCTTAACGACATCGCCGAGAAGAAAGTTCACCACACTTTCGTCGAGTGCGTTTC
GCTCGCTGTCGGCTGCAACATCTTCGTCTGCCTCGCTGTCTACTTCGTCCTGACCCTCAAGGACGGCGCT
GGCTACGTCTTCTCCGTCTTCTTCGCTGTCTACGCTTTCGCCATCGCCGGTTACGAGCATATCATCGCTA
ACATATATACCCTCAACATCGCCCTCATGGTCAACACCAAGATCACCGTCTACCAGGCCTACATCAAGAA
CCTCCTGCCCACTCTGCTCGGCAACTACATCGCCGGTGCCATCGTTCTCGGCCTTCCCCTTTACTTCATC
TACAAGGAACATTATTACAACTTCGAACGCTCGAAGCGCGATAACAACGACGCCCAGATGAAGTCTCTCA
GCATCGAGCTCCGCAAC

pccir2

ATGAACAAAGATCTGTGCGATGTGATTAAAGGCATTGATGATCTGATTGAAGTGGAAGTGAAAGCGGAAG
GCATTGAAACCATTCGTGATGAACTGTTTAACACCTATTGCCCGACCAACAAAGGCGGCAAAATGCGTCA
GCAGGGCCAGGATGGCTGCTGCATTGGCTATAGCGAAACCGTGATTAGCGCGTTTATTCATCTGCAAGAA
ACCCTGAAAAACAACGATAGCCAGAAAAAACTGGATCGTGATAAACTGGCGCAGTATGCGATTCTGTGGC
TGAGCTATAAAATTAACCAGCATCCGAACCAGACCTTTGGCACCAACGATATCTACAACAACTTCAAACA
GTACGGCTACTGGAACCGTGAACATAACAACTACATCGAACAGATCAAGAAATACGTGGATATCAAAGAT
ATGACCAAACTGCATGAAGCGTTTATTCTGCTGTGCAACATGTATACCGAAATTGATGAAAACAAAAGCA
ACTGCACCAAATGCAGCCAGAAAGCGAGCGAATTTGTGAAAAAATTTGAAATTCTGAACGATGATCCGAA
CCATATTAAAGATAGCCCGTATAGCCAGATTCTGCTGACCCTGAGCAAAGATTATGATAACTTTAAAAAC
TGCTGCAACAAGAAAAAGGGTGAAAGCTGCGATTTTCCGAGCCTGCCGCAGATTAGCCCAAAGAAAAGCT
TTGCGCAGAACAGCCTGGAAAGCCCGGGCCATACCAGCGGCCATAACAGCGAAGATATTAGCAGCAAAAG
CCCGATGGCGAACAAACTGATTCCGGGCCTGCTGATTTTTGCGGCGATTCCGGTGTTTCTGGGCATTGCG
TATAAATATAGCCTGTTTGGCTTTGATAAACAGCGTCATCGTCAGTATCTGCGTGAAAAAATTAAGAAAA
TCAAAAACAAAATGGCGAGCTATGTGGC

Appendix 7: Raw Data

Appendix 7.1.1

Nitrite uptake assay - calculated nitrite concentration [mM] in the medium at various time points in each of the replicates for the control study.

replicate	sample	30	60	120	240
1	4018	76	63	25	9
	4520	80	88	70	43
	Control	86	89	85	89
2	4018	83	62	34	10
	4520	80	87	81	49
	Control	81	82	79	81
3	4018	66	49	9	9
	4520	89	88	81	75
	Control	62.308	63.077	66.154	68.426
4	4018	76	57	10	9
	4520	83	88	81	75
	Control	66.154	68.462	65.385	68.462
5	4018	58.462	43.846	7.692	6.923
	4520	68.426	67.962	61.538	57.692
6	4018	50.769	37.692	6.923	6.923
	4520	63.846	67.962	64.615	57.692
7	4018	63.846	47.692	25.154	7.692
	4520	61.538	67.962	53.846	33.077
8	4018	58.462	48.462	19.231	6.923
	4520	61.538	66.923	62.308	37.692

Nitrite uptake assay - calculated nitrite concentration [mM] in the medium at various time points in each of the five replicates.

replicate	sample	0	30	60	90	120
1	4018	100.952	102.857	76.190	47.619	28.271
	4520	102.857	100.952	101.905	100.000	90.476
	NirC	109.524	78.095	70.476	62.857	53.333
	FocA	85.714	85.905	75.428	72.286	68.762
	NitA	107.619	104.762	104.762	90.470	102.857
	NitA/Ec	83.164	86.849	84.217	85.963	84.911
	HK	100.952	113.333	107.619	110.476	112.381
	control	105.714	105.714	103.810	100.952	102.857

2	4018	90.476	76.190	56.190	24.762	9.524
	4520	89.524	81.905	81.905	81.905	76.190
	NirC	80.952	80.952	66.667	64.762	60.952
	FocA	69.119	69.936	60.436	57.543	54.441
	NitA	106.667	92.381	91.429	84.762	77.143
	NitA/Ec	81.991	84.500	78.716	77.375	69.473
	HK	86.667	86.667	91.429	89.524	90.476
	control	80.952	87.619	86.667	80.000	88.571
3	4018	94.286	87.619	74.286	59.048	50.476
	4520	86.667	89.524	88.571	85.714	79.048
	NirC	80.952	83.810	75.238	62.857	59.048
	FocA	92.381	100.000	76.190	64.762	66.667
	NitA	88.571	92.381	91.429	81.905	82.857
	NitA/Ec	75.205	77.896	78.711	84.800	87.898
	HK	82.857	88.571	80.952	85.714	82.857
	control	93.333	86.667	88.571	93.333	85.714
4	4018	88.571	59.048	14.286	8.571	9.524
	4520	116.190	106.667	98.095	93.333	95.238
	NirC	88.571	69.524	57.143	42.857	31.429
	FocA	87.619	70.476	62.857	52.381	44.762
	NitA	85.714	86.667	86.667	90.476	78.095
	NitA/Ec	78.450	80.791	79.978	78.576	76.813
	HK	91.429	92.381	95.238	98.095	98.095
	control	80.000	80.952	81.905	81.905	82.857
5	4018	75.238	77.143	60.000	34.286	12.381
	4520	77.143	90.476	88.571	80.952	76.190
	NirC	72.381	71.429	54.286	40.952	35.238
	FocA	72.381	95.238	85.714	90.952	86.667
	NitA	78.095	89.524	92.381	84.762	85.714
	NitA/Ec	78.955	79.300	80.358	77.089	73.221
	HK	80.000	89.524	97.143	83.810	87.619
	control	69.524	86.667	85.714	76.190	81.905

Appendix 7.1.2

Nitrite uptake assay statistics.

Descriptives

VAR

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
1.00	5	22.0352	17.70613	7.91842	.0501	44.0203
2.00	5	83.4284	8.84749	3.95672	72.4428	94.4140
3.00	5	48.0000	13.74518	6.14703	30.9331	65.0669
4.00	5	68.7548	15.74102	7.03960	49.2097	88.2999
5.00	5	85.3332	10.40230	4.65205	72.4170	98.2494
6.00	5	78.4606	7.78263	3.48050	68.7972	88.1240
7.00	5	94.2856	11.52741	5.15521	79.9724	108.5988
8.00	5	88.3808	8.50230	3.80235	77.8238	98.9378
Total	40	71.0848	25.70385	4.06414	62.8643	79.3053

ANOVA

VAR

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	20957.063	7	2993.866	19.919	.000
Within Groups	4809.768	32	150.305		
Total	25766.832	39			

Appendix 7.2.1

Formate β-galactosidase assay - calculated β-galactosidase activity (Miller Units) at various time points in each of the replicates for the control study.

	0 mM			0.1 mM			0.5 mM		
1	2147.37			3123.08			3350		
2	2463.16	1580.95	1657.89	3507.69	2150	3947.37	3683.33	2866.67	6307.69
3	2066.67	2189.47	3631.58	3485.71	2714.29	4666.67	3750	2753.85	5500
4	2562.96	2216.67	3230.77	4327.27	3833.33	5214.29	2368.42	3783.33	4187.5
20	2176.67	1994.03		1547.83	1587.88		1276.64	1191.6	
24	693.33			824.56			788.43		

	1 mM			2 mM			RM201		
1	3366.67			3366.67			3165.22		
2	3716.67	3316.67	6153.85	2477.78	3183.33	4733.33	1978.08	1602.78	5411.76
3	2941.18	2584.62	5142.86	3100	4175	5266.67	1013.86	1230.93	5520
4	3476.92	3533.33	4928.57	3185.71	4511.11	5500	1278.57	1239.62	4710.45
20	1033.66	1240.37		1472.16	1242.2		1289.72	430.94	
24	833.04			906.42			578.29		

Formate β-galactosidase assay - calculated β-galactosidase activity (Miller Units) at various time points in each of the five replicates.

2 h		1	2	3	4	5
	RM	2206.061	1895.455	1146.667	1509.302	2038.360
	Foc	3924.324	3961.290	4082.353	2724.138	1146.667
	Pf/Ec	2192.727	2400.000	2482.759	2073.333	1731.034
	Cir	1140.909	2009.091	2561.905	1544.000	1365.217
	Nir	2328.571	3208.333	2969.231	3053.846	3666.667
3 h		1	2	3	4	5
	RM	1137.391	1156.098	1731.915	1253.846	1623.810
	Foc	3158.974	3126.829	5670.588	3220.000	3244.444
	Pf/Ec	1850.794	1720.000	2311.765	2228.571	1737.500
	Cir	904.000	1610.169	2535.714	1566.667	1592.593
	Nir	3488.889	3936.842	2700.000	2792.000	3063.636
4 h		1	2	3	4	5
	RM	1125.180	1091.473	1407.273	1043.165	1378.430
	Foc	3390.698	4120.000	5326.316	2793.750	3420.000
	Pf/Ec	1888.525	1632.353	3166.667	1626.667	1857.895
	Cir	677.143	1642.623	2266.667	1913.514	1646.154
	Nir	2608.333	2727.273	3236.364	3272.727	3894.737

Appendix 7.2.2

Formate β-galactosidase assay statistics.

2 h test:

Test of Homogeneity of Variances

VAR

Levene Statistic	df1	df2	Sig.
3.808	4	20	.018

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of VAR00006 is the same across categories of VAR00004.	Independent-Samples Kruskal-Wallis Test	.018	Reject the null hypothesis.
2	The medians of VAR00006 are the same across categories of VAR00004.	Independent-Samples Median Test	.008	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

3 h test:

Descriptives

VAR

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
1.00	5	1380.6120	277.57507	124.13534	1035.9570	1725.2670
2.00	5	3684.1670	1111.43253	497.04774	2304.1412	5064.1928
3.00	5	1969.7260	280.36950	125.38505	1621.6013	2317.8507
4.00	5	1641.8286	581.48471	260.04787	919.8200	2363.8372
5.00	5	3196.2734	515.11714	230.36739	2556.6710	3835.8758
Total	25	2374.5214	1087.28712	217.45742	1925.7113	2823.3315

ANOVA

VAR

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2.040E7	4	5098752.372	12.783	.000
Within Groups	7977629.523	20	398881.476		
Total	2.837E7	24			

4 h test:

Descriptives

VAR

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
1.00	5	1209.1042	170.55665	76.27525	997.3301	1420.8783
2.00	5	3810.1528	969.01913	433.35853	2606.9566	5013.3490
3.00	5	2034.4214	644.66078	288.30106	1233.9693	2834.8735
4.00	5	1629.2202	590.35556	264.01503	896.1969	2362.2435
5.00	5	3147.8868	512.16032	229.04506	2511.9558	3783.8178
Total	25	2366.1571	1144.16303	228.83261	1893.8698	2838.4444

ANOVA

VAR

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2.344E7	4	5860151.110	14.691	.000
Within Groups	7978012.245	20	398900.612		
Total	3.142E7	24			

Appendix 7.3 GFP assay

Truncation/ Replicate	Final OD(600)	Fluorescence units	Truncation/ Replicate	Final OD(600)	Fluorescence units
$PcCir2^{40}$	0.129	9830	$PcCir2^{254}$	0.115	8098
1	0.133	8369	1	0.136	8098
	0.108	9719		0.104	8139
	0.109	9994		0.110	10023
2	0.097	9988	2	0.109	10021
	0.111	9991		0.115	10025
	0.134	9980		0.141	10014
3	0.122	9992	3	0.102	10017
	0.136	9984		0.127	10016
	0.137	9906		0.109	9933
4	0.141	9888	4	0.116	9930
	0.127	9899		0.118	9930
	0.119	9993		0.118	10021
5	0.147	9979	5	0.116	10021
	0.130	9993		0.102	10027
$PcCir2^{106}$	0.115	4163	$PcCir2^{\Delta 73}$	0.196	3720
1	0.136	5433	1	0.227	4990
	0.117	5127		0.180	4802
	0.108	2675		0.189	3991
2	0.109	2739	2	0.130	2521
	0.123	3630		0.192	3780
	0.143	3507		0.168	3061
3	0.179	3805	3	0.168	3103
	0.180	4075		0.147	2916
	0.119	3122		0.180	2712
4	0.133	3508	4	0.094	1058
	0.147	3371		0.156	2249
	0.116	3196		0.218	3972
5	0.120	2953	5	0.170	3161
	0.136	3160		0.218	4114
$PcCir2^{210}$	0.181	8792	$PcCir2^{312}$	0.183	4099
1	0.192	10057	1	0.152	3339
	0.119	10058		0.205	4070
	0.196	10090		0.181	3675
2	0.179	10149	2	0.135	2517
	0.191	10149		0.124	2470
	0.159	8856		0.145	2714
3	0.160	9470	3	0.176	2822
	0.175	10180		0.151	2786
	0.159	9113		0.183	2634
4	0.172	8720	4	0.170	2702
	0.178	10189		0.167	2654
	0.172	8043		0.189	3177
5	0.221	10112	5	0.209	3359
	0.205	10108		0.194	3199

Appendix 7.4 PhoA assay

Truncation/ Replicate		OD600	OD420	Truncation/ Replicate		OD600	OD420
$PcCir2^{40}$		0.139	0.002	$PcCir2^{254}$		0.124	0.003
	1	0.139	0.004		1	0.158	0.002
		0.127	0.002			0.16	0.002
		0.145	0.003			0.115	0.001
	2	0.118	0.004		2	0.122	0.002
		0.122	0.002			0.117	0.002
		0.072	0.002			0.075	0.002
	3	0.076	0.002		3	0.074	0.001
		0.085	0.002			0.11	0.003
		0.124	0.003			0.134	0.003
	4	0.119	0.002		4	0.123	0.001
		0.109	0.002			0.088	0.004
		0.094	0.001			0.098	0.004
	5	0.102	0.002		5	0.125	0.003
		0.118	0.002			0.076	0.002
$PcCir2^{106}$		0.138	0.003	$PcCir2^{\Delta 73}$		0.126	0.007
	1	0.161	0.005		1	0.112	0.005
		0.158	0.004			0.104	0.005
		0.173	0.003			0.105	0.005
	2	0.132	0.001		2	0.112	0.005
		0.145	0.001			0.101	0.004
		0.096	0.002			0.071	0.005
	3	0.074	0.001		3	0.086	0.006
		0.107	0.004			0.077	0.003
		0.12	0.004			0.093	0.005
	4	0.105	0.002		4	0.106	0.007
		0.115	0.003			0.098	0.005
		0.095	0.003			0.068	0.007
	5	0.121	0.004		5	0.084	0.006
		0.071	0.003			0.063	0.005
$PcCir2^{210}$		0.137	0.004	$PcCir2^{309}$		0.121	0.013
	1	0.141	0.005		1	0.136	0.013
		0.152	0.004			0.154	0.015
		0.111	0.002			0.098	0.013
	2	0.104	0.002		2	0.123	0.014
		0.12	0.002			0.111	0.011
		0.093	0.003			0.075	0.011
	3	0.086	0.003		3	0.068	0.005
		0.092	0.002			0.082	0.011
		0.102	0.001			0.128	0.015
	4	0.116	0.002		4	0.092	0.011
		0.114	0.002			0.091	0.01
		0.1	0.003			0.079	0.013
	5	0.081	0.001		5	0.076	0.008
		0.079	0.001			0.076	0.007