Improved computational model of the malaria metabolic network and flux analysis for drug target prediction

by

Francis Isidore Garcia Totañes

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

THE UNIVERSITY OF LEEDS

Faculty of Biological Sciences

School of Molecular and Cellular Biology

March 2017

Intellectual property and publication statement

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Acknowledgements

Abstract

In recent years, genome scale metabolic models have become an important tool for identifying potential drug targets against pathogens. These are particularly important where cultivation and genetic manipulation (conditional knockouts) are difficult. Malaria is a globally important disease infecting 212 million cases and causing more than 400,000 deaths in 2015. The resistance of the parasite to all antimalarial drugs on the market emphasises the urgent need to identify new drug targets. There are a few malaria metabolic models that have already been developed; however, these models are limited in terms of network size or input of accurate experimentally derived metabolomics and biomass data. With extensive curation and utilisation of parasite-specific constraints in the improvement of existing metabolic network models, a highly curated metabolic network model of *Plasmodium falciparum*, iFT342, was developed here. The model has updated gene and reaction annotations as well as additional species identifiers that will facilitate ease in comparison with other models. The model has no dead-end metabolites (compared to 5 to 39% for other highly curated models) and has the highest percentage of live reactions. With the addition of experimentally measured biomass composition and metabolite fluxes for glucose and 18 amino acids, iFT342 was able to model *in vitro* parasite growth in restricted glucose environment with remarkable fidelity. In addition, through single gene knockout analysis, the model was able to significantly enrich the number of experimentally validated essential genes (true positives) in the predicted essential gene set, and had the highest percentage of true positive predictions compared with other malaria models. Finally, as proof of concept, inhibition of parasite growth was demonstrated using gemcitabine, which targets UMP-CMP kinase, a novel target predicted by the model. Gemcitabine inhibited parasite growth in a dose-dependent fashion exhibiting an IC_{50} in the low micromolar range and blocked the development of the parasite from the trophozoite to the schizont stage.

Table of contents

List of figures

List of tables

Abbreviations

ACP	Acyl-carrier protein
ADP	Adenosine diphosphate
aLipDH	Apicoplastic dihydrolipoyl dehydrogenase
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
BIGG	Biochemical Genetic and Genomic knowledgebase
BRENDA	BRaunschweig ENzyme DAtabase
cAMP	Cyclic AMP
CDP	Cytidine diphosphate
ChEBI	Chemical Entities of Biological Interest
CID	PubMed compound identifier
CMP	Cytidine monophosphate
CoA	Coenzyme A
COBRA	COnstraint-Based Reconstruction and Analysis
COBRApy	COnstraint-Based Reconstruction and Analysis for Python
CTP	Cytidine triphosphate
dCMP	Deoxycytidine monophosphate
DHFR	Dihydrofolate reductase
DHODH	Dihydroorotate dehydrogenase
DHPS	Dihydropteroate synthase
DNA	Deoxyribonucleic acid
dTMP	Deoxythymidine monophosphate
dUMP	Deoxyuridine monophosphate
EC	Enzyme Commission
ExPASy	Expert Protein Analysis System database
FBA	Flux balance analysis
FDA	Food and Drug Administration
FMOC	9-fluorenylmethoxycarbonyl chloride
FVA	Flux variability analysis
G6PD	Glucose 6-phosphate dehydrogenase
GDP	Guanosine diphosphate
gDW	Gram per dry weight
GLUT1	Glucose transporter 1
GLPK	GNU Linear Programming Kit
GO	Gene ontology
GSM	Genome scale metabolic model
GTP	Guanosine triphosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HGXPRTase	Hypoxanthine-guanine-xanthine phosphoribosyltransferase
HPLC	High performance liquid chromatography

IC$_{50}$	Half maximal inhibitory concentration
ID	Identifier
IMP	Inosine monophosphate
InChI	International Chemical Identifier
IUPAC	International Union of Pure and Applied Chemistry
KEGG	Kyoto Encyclopedia of Genes and Genomes
Ki	Inhibition/dissociation constant
Km	Michaelis constant
MIC$_{50}$	Minimum concentration resulting in at least 50% growth inhibition
MOMA	Minimisation of metabolic adjustments
MPA	3-mercaptopropionic acid
MPMP	Malaria Parasite Metabolic Pathways database
mRNA	Messenger ribonucleic acid
MW	Molecular weight
NADH	Reduced nicotinamide adenine dinucleotide
NADPH	Reduced nicotinamide adenine dinucleotide phosphate
NHSBT	National Health Service Blood and Transplant Unit
NPPs	New permeability pathways
OMP	Orotidine monophosphate
OPA	O-phthaldialdehyde
ORF	Open reading frame
PfENT1	*P. falciparum* equilibrative nucleoside transporter 1
PfHT1	*P. falciparum* hexose transporter 1
PfLDH	*P. falciparum* lactate dehydrogenase
6PGD	6-phosphogluconate dehydrogenase
PlasmoDB	*Plasmodium falciparum* Genome Database
PPP	Pentose phosphate pathway
PPV	Positive predictive value
PTFE	Polytetrafluoroethylene
RBC	Red blood cell
RNA	Ribonucleic acid
RPMI	Roswell Park Memorial Institute
SBML	Systems Biology Markup Language
SD	Standard deviation
SEM	Standard error of the mean
SIB	Swiss Institute of Bioinformatics
SID	PubMed compound identifier
SMILES	Simplified Molecular-Input Line-Entry System
TCA	Tricarboxylic acid
TPR	True positive rate
tRNA	Transfer ribonucleic acid
UHPLC	Ultrahigh performance liquid chromatography
UMP	Uridine monophosphate

UniProt	Universal Protein Resource
UTP	Uridine triphosphate
UV	Ultraviolet
XML	eXtensible Markup Language
XMP	Xanthosine monophosphate

Chapter 1   General introduction

1.1   Malaria metabolic pathways and drug targets

Malaria is an arthropod-borne disease caused by protozoan parasites in the genus *Plasmodium*. *Plasmodium* falls under the Apicomplexan phylum, a group of obligate parasites known for the red algae-derived organelle called the apicoplast (Arisue & Hashimoto, 2015). It infects a variety of species such as lizards (Hicks & Schall, 2014), birds (Medeiros *et al*, 2013) and mammals (Perkins & Schaer, 2016). There are four main species that naturally infect humans: *P. falciparum*, *P. malariae*, *P. ovale* and *P. vivax*, of which *P. falciparum* results in the most fatality (Snounou *et al*, 1993; Wernsdorfer, 2012). There have been reports of naturally acquired infection in humans by simian malaria, *P. cynomolgi* and *P. knowlesi* (Ta *et al*, 2014; White, 2008). It is estimated that about 50% of the world's population is at risk of malaria. A total of 104 countries and territories are currently endemic for malaria, most of which are in the tropics and subtropics (World Health Organization, 2014; Feachem *et al*, 2010). In 2015, about 212 million worldwide contracted the disease, resulting in 429,000 deaths (World Health Organization, 2016).

1.1.1   Malaria life cycle

Malaria has a complex life cycle that can distinctly be divided into three phases: the sporogonic phase which occurs in the mosquito vector, the exo-erythrocytic phase which occurs in the host liver and the erythrocytic phase/cycle which occurs in the blood (Schuster, 2002). In the sporogonic phase, male and female gametocytes, known as microgametocytes and macrogametocytes, respectively, are taken up, for human malarias, by the anopheline mosquito vector from an infected human host during a blood meal (Figure 1.1a). The gametes combine to form a zygote in the stomach of the insect. The zygote penetrates into the midgut of the mosquito where it develops into an oocyst. The oocyst undergoes a series of mitoses and subsequent cellular differentiation to form sporozoites. The oocyst ruptures and releases the sporozoites which migrate towards the salivary glands (Vaughan, 2007). During its next blood meal, the mosquito introduces sporozoites into the human blood stream (Figure 1.1b). This infective stage then reaches the liver where it invades hepatocytes and begins the exo-erythrocytic phase. In the liver cell, the parasite develops into a mature schizont and divides asexually into thousands of merozoites. The merozoites are eventually released into the blood where they invade erythrocytes (Frevert, 2004). The erythrocytic cycle starts with the merozoite entering the red blood cell (RBC) (Figure 1.1c). As the merozoite penetrates the host cell, it forms a membrane layer around it which seals off when the invasion is complete. This forms the

parasitophorous vacuole membrane that serves as a secondary layer around the parasite (Cowman & Crabb, 2006). In the RBC, the parasite develops from the merozoite to the early trophozoite stage (also known as the ring stage), followed by the late trophozoite stage and finally the schizont stage. The schizont divides into multiple merozoites, which invade other RBCs. Some trophozoites develop into gametocytes which are taken up by the mosquito, thus completing the whole life cycle.
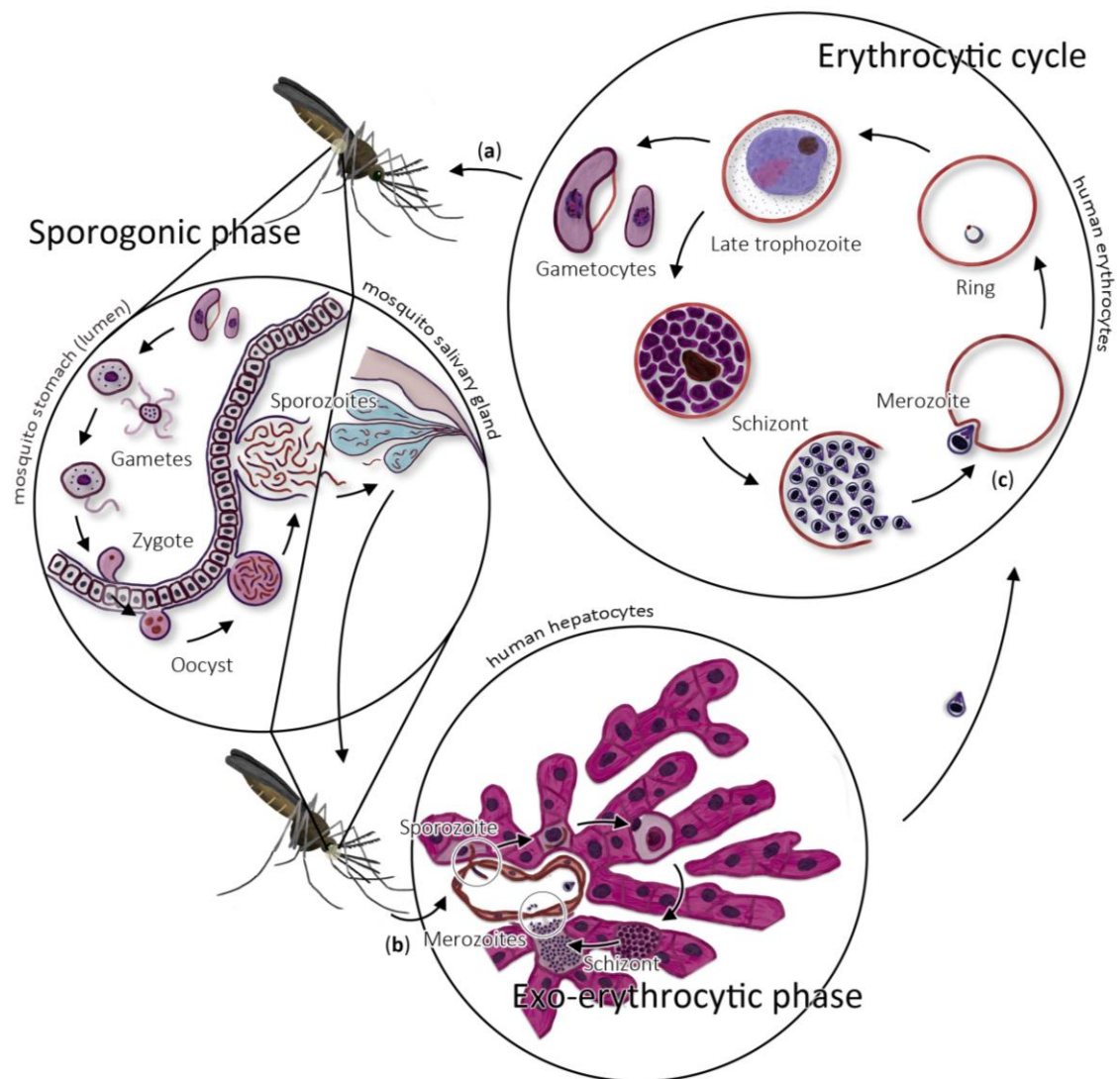
Figure 1.1   *Plasmodium falciparum* life cycle
The *P. falciparum* life cycle is divided into three phases: the sporogonic phase, which occurs in the anopheline mosquito vector; the exo-erythrocytic phase, which occurs in the host liver; and finally the erythrocytic phase/cycle, which occurs in the red blood cell.

1.1.2   Carbon metabolism

The erythrocytic cycle is mainly responsible for the symptoms brought about by the disease and has been extensively investigated for the development of antimalarial drugs and vaccines (Bozdech *et al*, 2003). During this cycle which takes about 48 hours, the parasite is quite

metabolically active and requires a lot of nutrients from the RBC and the host serum (Schuster, 2002). The new permeability pathways (NPPs), which transport solutes into the RBC, allow this increased demand to be met. As the parasite develops from ring to schizont, the host cell permeability increases allowing greater influx of important nutrients (Ginsburg *et al*, 1986b). These NPPs have been postulated to either come from parasite transporters incorporated into the RBC membrane or as a result of parasite-mediated alterations in host membrane proteins (Ginsburg & Stein, 2004). Because the parasite lacks the ability to store carbohydrates as an energy source (Scheibel & Miller, 1969; Olszewski & Llinás, 2011), *P. falciparum* is largely dependent on the presence of glucose, with this sugar being the parasite's main energy and carbon source (Roth, 1990; Slavic *et al*, 2010). As shown in Figure 1.2, the parasite imports glucose present within the red blood cell, brought in via the erythrocyte glucose transporter (GLUT1) and the NPPs (Kirk & Lehane, 2014). Glucose freely diffuses through the parasitophorous vacuole membrane then passes through a glucose transporter (PfHT1) in the parasite membrane (Woodrow *et al*, 2000). Compared with uninfected RBCs, infected cells have been shown to have up to a hundred-fold increase in glucose uptake (Jensen *et al*, 1983; Roth, 1990). PfHT1 has been extensively characterised and was shown to be functionally different from its human counterpart, making it a potential drug target (Joet *et al*, 2003; Joët & Krishna, 2004). Compound 3361, a glucose analogue that inhibits this transporter, has been shown to be effective *in vitro* against the parasite (Patel *et al*, 2008; Saliba *et al*, 2004). Compared with uninfected cells, there is a lower percentage of glucose that is converted to lactate through glycolysis in the infected cell, suggesting the utilisation of the sugar in the production of biomass components, apart from energy production (adenosine triphosphate, ATP). For example, glucose 6-phosphate can be metabolised to ribose 5-phosphate for the production of nucleotides, or to mannose 6-phosphate for glycosylphosphatidylinositol anchor biosynthesis. Further down the glycolytic pathway, pyruvate can be metabolised to produce acetyl-coenzyme A (CoA) for lipid metabolism or incorporation into the tricarboxylic acid cycle (Olszewski & Llinás, 2011).
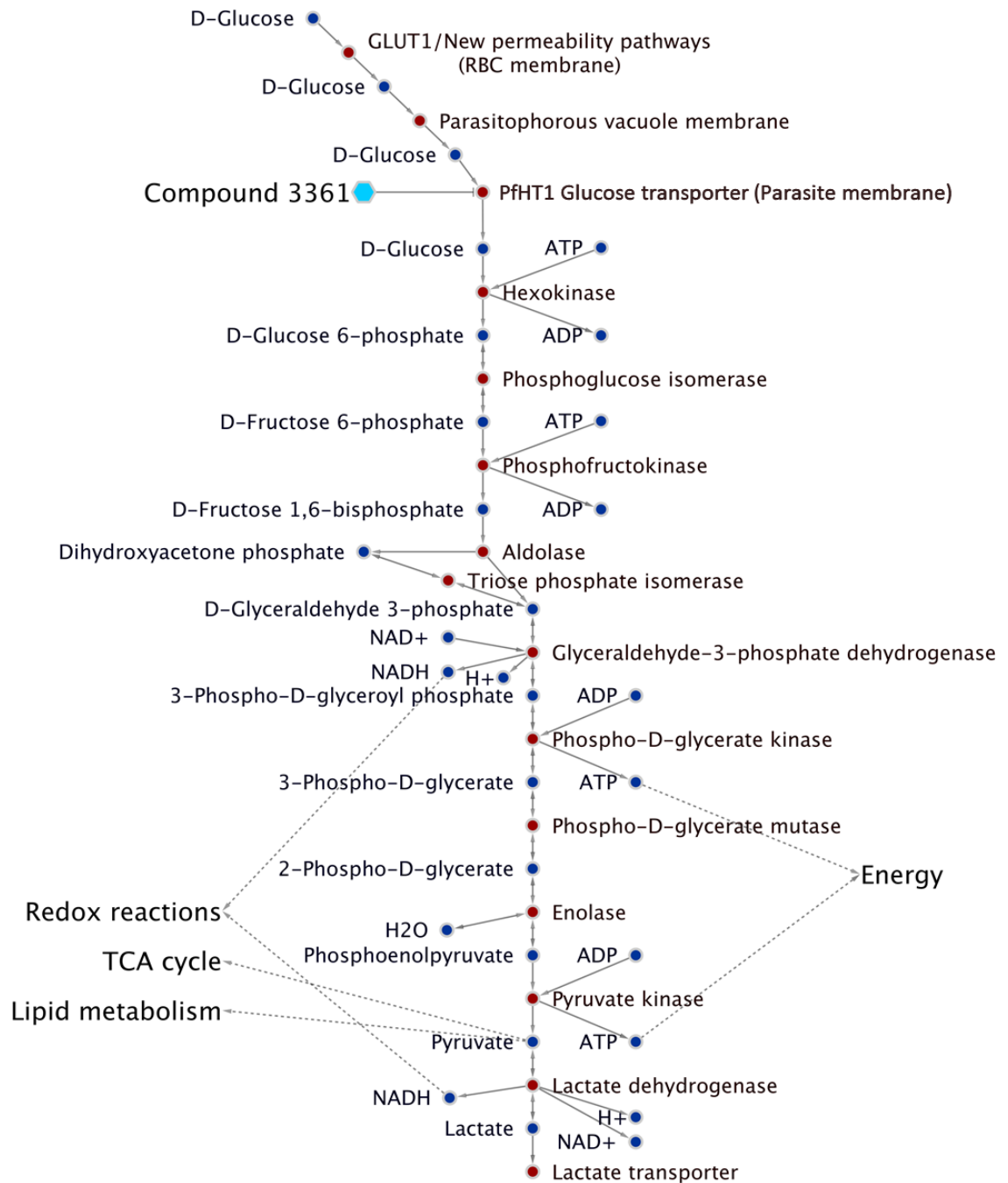
Figure 1.2   The glycolytic pathway
The glycolytic pathway, also known as glycolysis, involves a series of reactions (in red) that result in the transport and conversion of glucose into different metabolites (in blue) and finally to lactate. Glucose enters the red blood cell through GLUT1 and the new permeability pathways, then diffuses through the parasitophorous vacuole, and finally through the hexose transporter (PfHT1) in the parasite membrane. This pathway produces energy in the form of ATP, as well as reducing equivalents in the form of reduced nicotinamide adenosine dinucleotide (NADH). Compound 3361, shown in cyan, is a known inhibitor of PfHT1.

The pentose phosphate pathway (PPP) is important in generating ribose 5-phosphate from glucose 6-phosphate for use in the production of nucleotides (Figure 1.3). The pathway also produces reduced nicotinamide adenine dinucleotide phosphate (NADPH) which is an important reducing equivalent especially during oxidative stress (Barrett, 1997; Stincone *et al*, 2015). The importance of this pathway in *Plasmodium* is exemplified in the case of glucose 6-phosphate dehydrogenase (G6PD) deficiency in the human host wherein resistance against malaria has been observed. This suggests that the parasite at some point in its lifecycle is dependent on the capacity of the host to produce NADPH and maintain redox balance (Ruwende & Hill, 1998). PPP has two branches: the oxidative branch and the non-oxidative branch, both of which produce ribose 5-phosphate. The oxidative branch generates ribose 5-phosphate from glucose 6-phosphate through a series of four reactions, beginning with G6PD and finally with the isomerisation of ribulose 5-phosphate to ribose 5-phosphate (Bozdech & Ginsburg, 2005). It is important to note that the first two steps in this branch are facilitated by a bifunctional G6PD/6-phosphogluconolactonase enzyme (Clarke *et al*, 2001). This enzyme is different from human G6PD (Jortzik *et al*, 2011) and is likely to be essential for parasite survival based on attempted gene knockout experiments, as well as enzyme and *in vitro* growth inhibition assay with ellagic acid, making it a promising antimalarial target (Allen *et al*, 2015). This oxidative branch generates two moles of NADPH for every mole of glucose 6-phosphate. On the other hand, the non-oxidative branch utilises fructose 6-phosphate and glyceraldehyde 3-phosphate from glycolysis to produce ribose 5-phosphate. This branch is reversible and metabolites can be redirected to meet the needs of the parasite. The non-oxidative branch may be turned off in the case where the parasite requires both metabolites for nucleic acid synthesis and NAPDH. On the other hand, if the parasite is in need of energy, ribulose 5-phosphate may be consumed by the reverse of the non-oxidative branch to produce fructose 6-phosphate and glyceraldehyde 3-phosphate that can be shunted back to glycolysis for the production of ATP and NADH (Bozdech & Ginsburg, 2005).
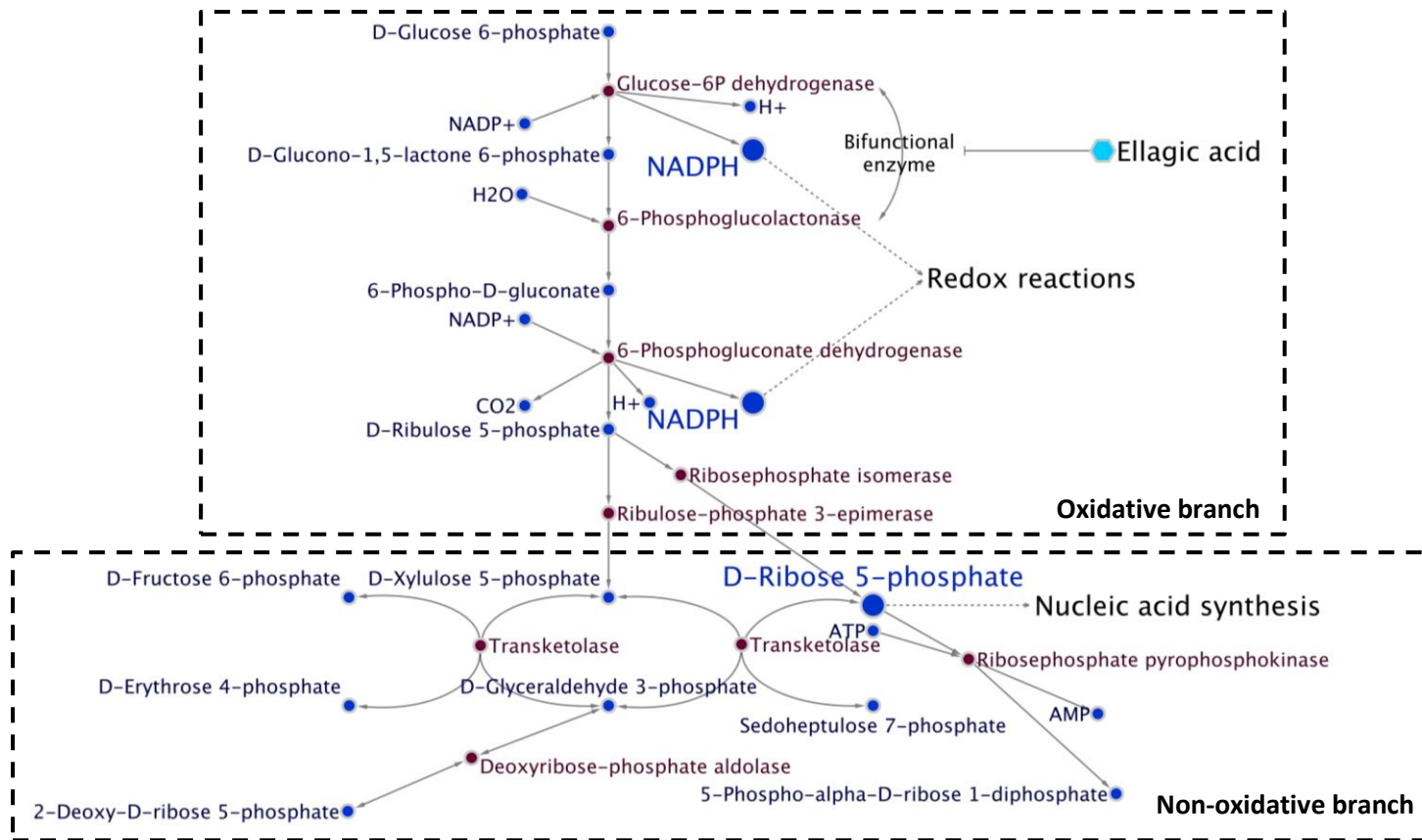
Figure 1.3   The pentose phosphate pathway

The pentose phosphate pathway can be divided into two parts, the oxidative branch and the non-oxidative branch. Both branches are capable of producing ribose 5-phosphate, which is utilised in nucleic acid synthesis. In addition, the oxidative branch produces NADPH that participates in redox reactions. Reactions are shown here in red while metabolites are shown in blue. Ellagic acid (cyan) is a known inhibitor of the bifunctional enzyme, glucose 6-phosphate dehydrogenase/6-phosphogluconate dehydrogenase. In large fonts are important metabolite products of this pathway.

1.1.3 Mitochondrial metabolism

There are a number of indispensable metabolic pathways that occur in the mitochondria. The tricarboxylic acid (TCA) cycle occurs in the parasite's mitochondria and is important in the generation of NADH and ubiquinol (Figure 1.4). In addition, succinyl-CoA from the TCA cycle can be utilised for the production of haem which is important in the generation of cytochrome for the mitochondrial electron transport chain. All enzymes involved in the canonical TCA cycle are present in the *Plasmodium* genome except for pyruvate dehydrogenase, which is located in the apicoplast (Gardner *et al*, 2002). Pyruvate dehydrogenase exists as a 4-subunit enzyme complex, the pyruvate dehydrogenase complex, which converts pyruvate to acetyl-CoA that can then be fed into the TCA cycle or to fatty acid synthesis pathway (Foth *et al*, 2005). It is currently speculated that oxoglutarate (Figure 1.4a) is the main metabolite that is fed into the TCA cycle instead of acetyl-CoA (Figure 1.4b) in the canonical cycle. ^{13}C labelling studies showed that labelled glucose only contributes to a small proportion of the acetyl-CoA that enters the TCA cycle. On the other hand, labelled glutamine which is converted to glutamate was observed to contribute to the labelling of TCA metabolites such as succinate, fumarate, malate and citrate. At the beginning of the canonical TCA cycle, acetyl-CoA from glucose condenses with oxoglutarate to form citrate; however, ^{13}C-glutamine was shown to contribute more to citrate labelling than ^{13}C-glucose (Olszewski & Llinás, 2011; Olszewski *et al*, 2010; Vaidya & Mather, 2009).

Electrons from the TCA cycle are then brought into the inner mitochondrial membrane where they participate in the electron transport chain. In many organisms, the electron transport chain involves a series of redox reactions that creates a proton gradient and drives the synthesis of ATP; however, the parasite genome does not encode for the necessary enzyme for this (Gardner *et al*, 2002). Instead, the electron transport chain is utilised for the conversion of ubiquinol to ubiquinone. A number of dehydrogenases transfer electrons from different sources (e.g., TCA cycle) through the reduction of ubiquinone and this in turn creates the proton gradient, as shown in Figure 1.5. This proton gradient then drives the recycling of ubiquinol to ubiquinone through the action of cytochrome bc1 (Painter *et al*, 2007; Nixon *et al*, 2013). Proguanil, an antimalarial drug, inhibits parasite growth in two ways: (1) it is converted in the human liver to cycloguanil which is a potent inhibitor of dihydrofolate reductase (DHFR); and (2) at high micromolar concentrations, proguanil disrupts the electron membrane potential. Proguanil is given in combination with atovaquone, which targets cytochrome bc1. Together, they prevent the oxidation of ubiquinol to ubiquinone, which is necessary for pyrimidine synthesis (Srivastava & Vaidya, 1999; Guiguemde *et al*, 2012).
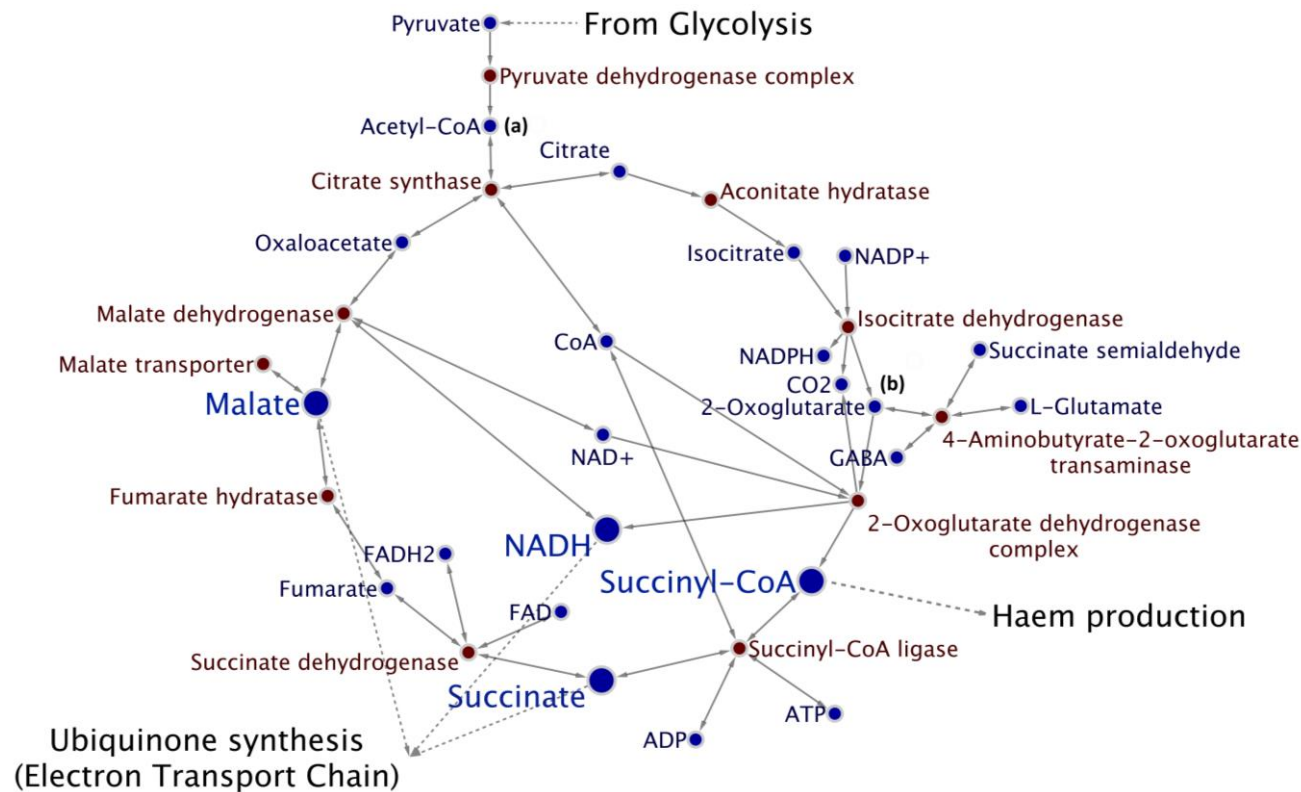
Figure 1.4   The tricarboxylic acid cycle

The tricarboxylic acid cycle, or the TCA cycle, is responsible for the production of metabolites that participate in the electron transport chain. It also produces succinyl-CoA that goes into the production of haem, and eventually cytochrome. The metabolites of a canonical TCA cycle are normally obtained from pyruvate from glycolysis, with acetyl-CoA (a) as the main metabolite being fed into the cycle. However, recent studies show that more of the metabolites participating in the TCA cycle are obtained from glutamate, making 2-oxoglutarate (b) as the main metabolite that enters into the TCA cycle. Shown in red are the reactions and in blue are the metabolites. Important metabolic products are shown in large fonts.
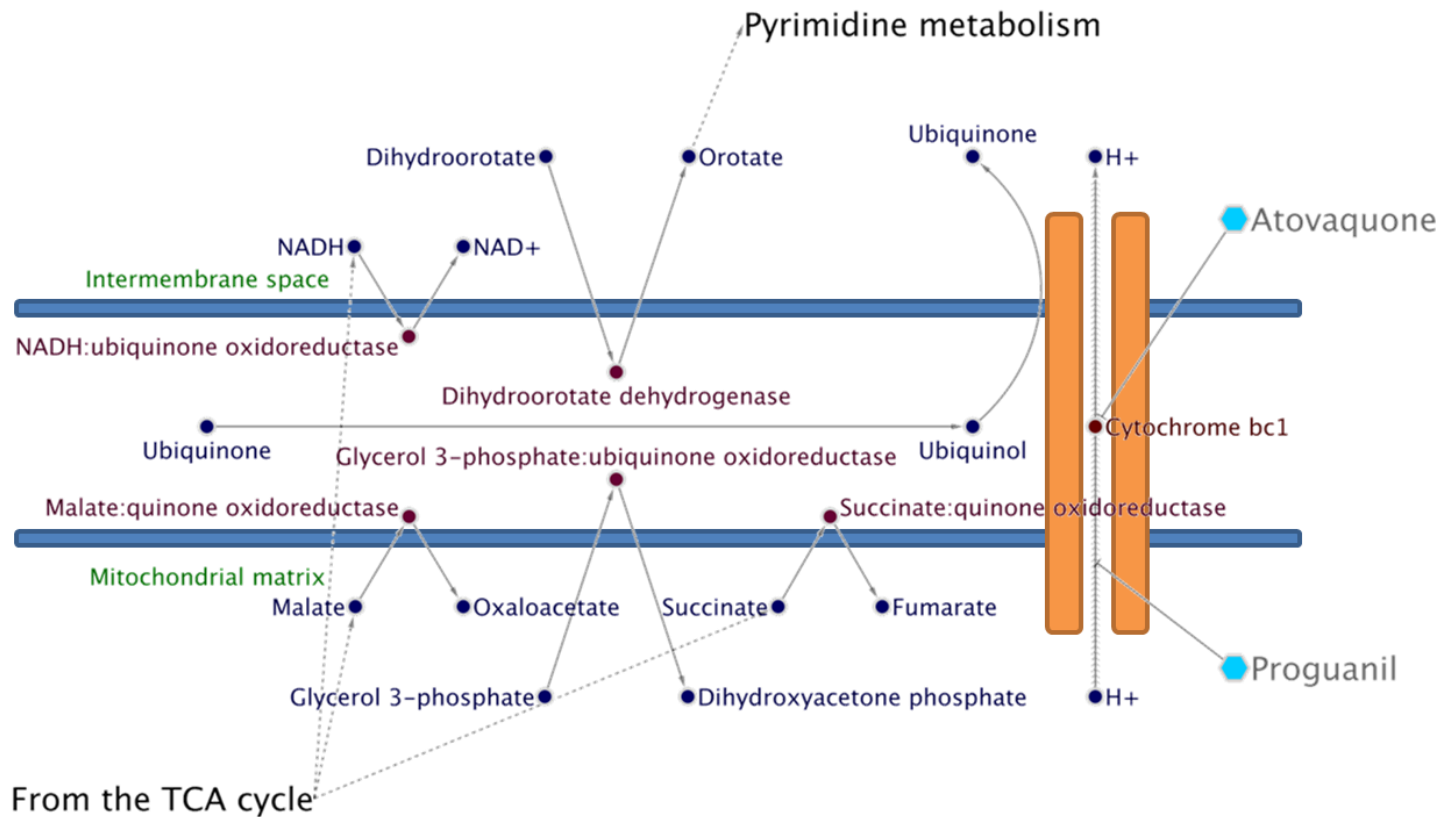
Figure 1.5   The electron transport chain
Typically, the electron transport chain generates a proton gradient between the intermembrane space and the matrix of the mitochondria that drives the phosphorylation of ADP to ATP. In *Plasmodium* this gradient is formed as a result of dehydrogenase/oxidoreductase reactions in the inner mitochondrial membrane, and ultimately facilitates the production of ubiquinone via cytochrome bc1. Atovaquone inhibits cytochrome bc1 while proguanil disrupts the proton gradient, both preventing the production of ubiquinone. This in turn inhibits the production of pyrimidines.

1.1.4 Nucleotide metabolism

The malaria parasite has the necessary mechanism for *de novo* pyrimidine synthesis but lacks pyrimidine salvage pathways. Uridine monophosphate (UMP), which is the precursor for cytidine monophosphate (CMP), deoxycytidine monophosphate (dCMP) and deoxythymidine monophosphate (dTMP), is synthesised from the metabolism of aspartate and glutamine and follows the canonical pathway (Figure 1.6). Glutamine is converted to carbamoyl phosphate and combined with aspartate to be further metabolised by dihydroorotase and dihydroorotate dehydrogenase (DHODH) to produce orotate, a pyrimidine carboxylic acid. DHODH requires ubiquinone as an electron acceptor for the reaction to proceed. With the addition of phosphoribosyl pyrophosphate followed by decarboxylation, UMP is produced (Ginsburg, 2006; Hyde, 2007). The dependence of *Plasmodium* on the *de novo* synthesis of pyrimidines has made this pathway an important target for antimalarials (Heikkilä *et al*, 2007; Krungkrai & Krungkrai, 2016). Cycloguanil, pyrimethamine and trimethoprim are antimalarial drugs that target DHFR. Inhibition of DHFR prevents the conversion of dihydrofolate to tetrahydrofolate, the latter being an important cofactor in the methylation of dUMP to dTMP (Olliaro, 2001). Sulfadoxine, an inhibitor of dihydropteroate synthase (DHPS) which in turn inhibits the production of a dihydrofolate precursor (dihydropteroate), was often administered together with pyrimethamine for the treatment of uncomplicated malaria (Dorsey *et al*, 2002). As previously mentioned, drugs such as atovaquone and proguanil disrupt the electron transport chain. This disruption lowers the ubiquinone supply in the mitochondria and indirectly inhibits the DHODH reaction. The DHODH reaction, which has been identified as essential to the parasite (McRobert & McConkey, 2002; Painter *et al*, 2007), is currently being exploited for the development of new antimalarials. DSM265, a long-duration DHODH inhibitor for example has been advanced to phase I clinical trials (Phillips *et al*, 2015).

Similar to other Apicomplexan parasites, *Plasmodium* is incapable of synthesising purine nucleotides *de novo* thus it needs to acquire precursors from the host (Chaudhary *et al*, 2004; Downie *et al*, 2008). The NPPs allow permeation of precursors, which are then transported through the parasite plasma membrane via the equilibrative nucleoside transporter 1 (PfENT1) (Figure 1.7). These transporters are responsible for bringing in nucleoside and nucleobases such as adenine, adenosine, guanine, hypoxanthine and xanthine into the parasite (Hyde, 2007). Apart from these purine bases, *Plasmodium* has the capacity to utilise 5'-methylthioinosine and 5'-methylthioadenosine from polyamine metabolism to produce hypoxanthine (Downie *et al*, 2008). Hypoxanthine is an important nucleobase that can further be converted into different purine nucleotides (Ginsburg, 2006; Hyde, 2007). Allopurinol, a drug normally used for the treatment of gout, has been shown to inhibit hypoxanthine-guanine-xanthine

phosphoribosyltransferase (HGXPRTase) (Downie *et al*, 2008). A combination of allopurinol and quinine resulted in significantly lowered mean parasite clearance time compared with treatment with quinine alone (Sarma *et al*, 1998). Immucillin-H, an immunosuppressant, which targets both human and malaria purine nucleoside phosphorylase, inhibits the conversion of inosine to hypoxanthine in the parasite. When the compound was modified to preferentially target the parasite enzyme, it was shown to inhibit parasite growth in the absence of hypoxanthine (Ting *et al*, 2005).
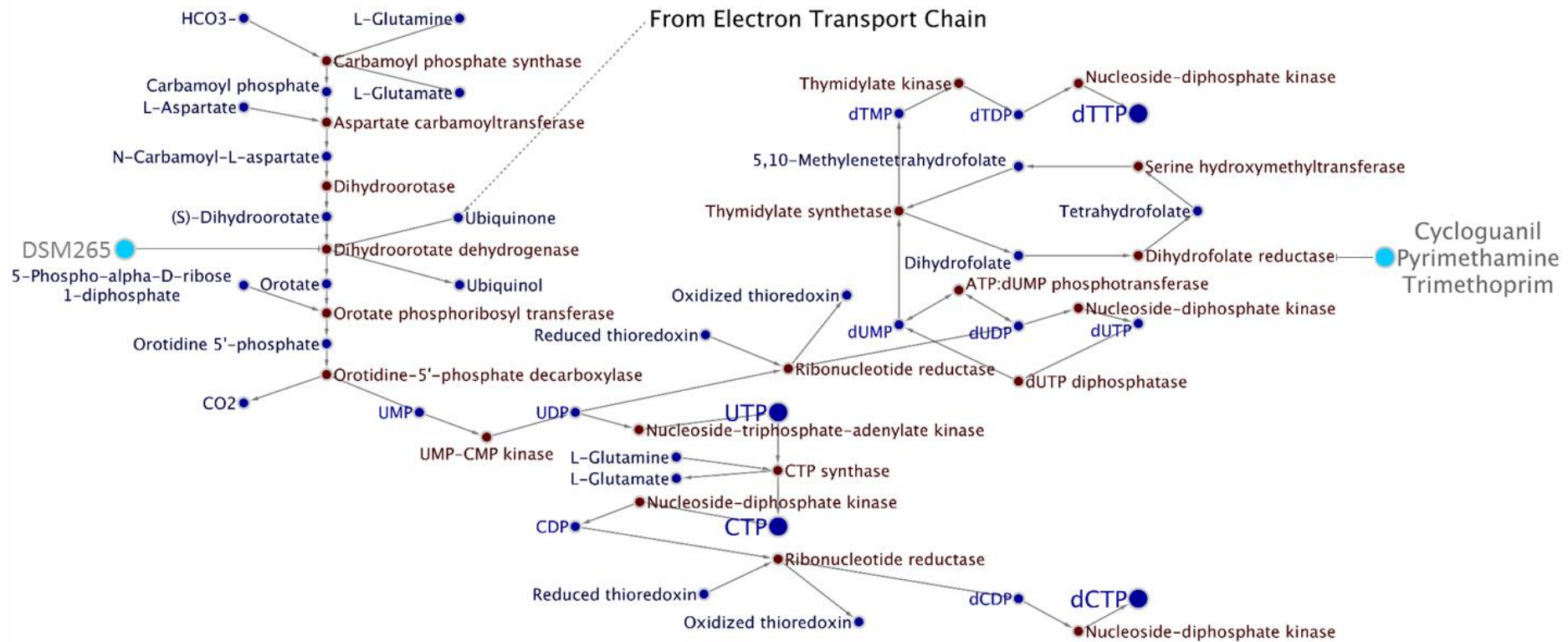
Figure 1.6   *De novo* synthesis of pyrimidines

Pyrimidines are obtained from glutamine and aspartate and the eventual production of orotidine monophosphate (OMP), which is the precursor for all pyrimidine nucleotides. Decarboxylation of OMP produces uridine monophosphate (UMP), which can be phosphorylated to generate di- and triphosphate counterparts. Deoxyuridine monophosphate (dUMP), obtained from the reduction and dephosphorylation of uridine diphosphate (UDP), serves as a precursor for thymidine deoxynucleotides (via thymidylate synthetase), while uridine triphosphate (UTP) serves as a precursor for cytidine nucleotides (via cytidine triphosphate (CTP) synthase). Important inhibitors of pyrimidine synthesis include DSM265, which acts on dihydroorotate dehydrogenase (DHODH) and prevents the production of orotate. Cycloguanil, pyrimethamine and trimethoprim inhibit dihydrofolate reductase, limiting the amount of dihydrofolate necessary for thymidylate synthase reaction. This in turn blocks the production of thymidine deoxynucleotides.
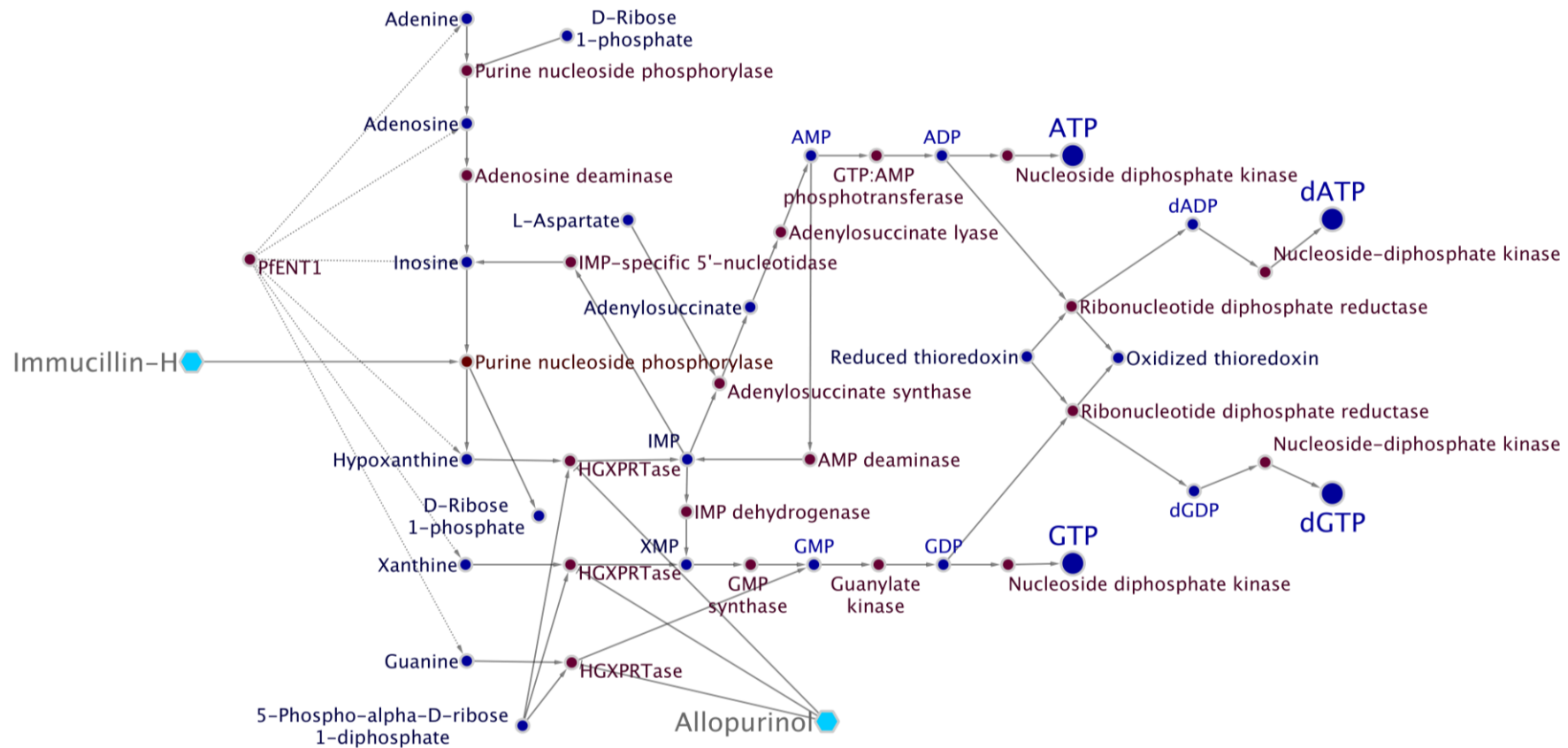
Figure 1.7   Purine salvage in *Plasmodium*

Plasmodium is capable of obtaining purine nucleosides and bases from its host via the nucleoside transporter, PfENT1. Adenosine nucleotides originate from inosine monophosphate (IMP), which comes from the phosphoribosylation of hypoxanthine. Guanosine nucleotides, on the other hand, can either come from xanthine monophosphate (XMP), produced either from the oxidation of IMP or from xanthine or guanine through the action of hypoxanthine-guanine-xanthine phosphoribosyltransferase (HGXPRTase). Known inhibitors that affect this pathway are shown in cyan. Reactions are shown in red and metabolites in blue.

1.1.5 *Amino acid metabolism*

The intraerythrocytic parasite is also dependent on external sources of amino acids. The loss of enzymes responsible for the *de novo* synthesis of amino acids through evolution has resulted in auxotrophic dependence on a number of amino acids (Payne & Loomis, 2006). Parasites digest about 60 to 80% of the haemoglobin to obtain most amino acids (Francis *et al*, 1997; Krugliak *et al*, 2002). Haemoglobin is imported into the parasite through a double-membrane invagination of the parasitophorous vacuole membrane and the plasma membrane called the cytostome (Figure 1.8). The cytostome eventually pinches off to form a vesicle that fuses with the digestive vacuole where digestion via proteases occurs (Milani *et al*, 2015). However, haemoglobin cannot provide enough cysteine, glutamate, glutamine and methionine, thus the parasite obtains these either from human serum or the culture media (Divo *et al*, 1985; Francis *et al*, 1997). Furthermore, isoleucine is not present in haemoglobin thus the parasite relies solely on getting it from the serum; and the absence of isoleucine has been shown to result in slowed growth (Babbitt *et al*, 2012). The parasite has a number of aspartic and cysteine proteases for haemoglobin digestion and the redundancy of these proteases allow the parasite to balance the utilisation of amino acids from haemoglobin and those from the surrounding environment. It has been shown that there are also strain specific nuances when it comes to the importance of amino acids obtained from the external environment. In some cases, parasites have been shown to grow in culture media in the absence of all amino acids except for isoleucine, while for others, methionine in the media was also necessary (Liu *et al*, 2006). Protease inhibitors have been shown to inhibit parasite growth *in vitro* and *in vivo* (Pérez *et al*, 2013; Greenbaum *et al*, 2004); however, knockout of cysteine and aspartic protease genes was not lethal and still permitted growth in amino acid rich media (Liu *et al*, 2006).

Plasmodium is known to digest haemoglobin and release most of the digested amino acids from the red blood cell (Zarchin *et al*, 1986) suggesting that the parasite digests haemoglobin for other purposes aside from obtaining amino acids, at the expense of producing toxic ferriprotoporphyrin IX or haem (Francis *et al*, 1997). In a study that looked into haemoglobin digestion and utilisation in different strains of *P. falciparum*, up to 65% of haemoglobin was digested by the parasite on average but only 15% of which was taken up by the parasite for protein synthesis. Some suggest that this allows more room for the parasite and prevents build-up of internal volume within the RBC that could result in premature cell lysis (Krugliak *et al*, 2002; Allen & Kirk, 2004). Others suggest that the parasite controls the colloid osmotic pressure in the host cell to counter the effect of increasing permeability of the RBC membrane as the parasite develops in order to prevent premature lysis of the host cell. By breaking down haemoglobin and releasing the amino acids into the surrounding environment, the colloid

osmolarity inside the RBC is lowered, preventing the movement of permeable cations and water inward (Lew *et al*, 2003; Waldecker *et al*, 2017).

Detoxification of haem from haemoglobin digestion is key to the survival of the parasite. This is done by haem detoxification protein that converts haem into crystalline haemozoin which is inert. Haemozoin appears as a dark pigment that can easily be appreciated when observing Giemsa-stained infected blood smears under the microscope (Fitch, 2004; Francis *et al*, 1997). Chloroquine was developed in the 1930s through the addition of a diethylaminoisopentylamino side chain on chloroquinoline and became one of the World Health Organization's drugs of choice against malaria. It acts by preventing haem polymerisation, thus taking advantage of the toxicity of haem to the parasite (Schlitzer, 2007). Other quinolines such as quinine and mefloquine prevent the fusion of vesicles that carry haemoglobin with the digestive vacuole (Fitch, 2004). Lumefantrine also prevents haemozoin formation (Combrinck *et al*, 2013) and is currently administered together with artemisinins (Tilley *et al*, 2016). Artemisinin derivatives release oxidative radicals after activation through the cutting of the endoperoxide bond by iron-containing reactive species such as haem (Tilley *et al*, 2016). It has been speculated that artemisinins alkylate haem and prevents haemozoin formation (O'Neill *et al*, 2010). Others suggest that the mode of action of reactive artemisinin derivatives involves the formation of covalent bonds with proteins (Yang *et al*, 1994), disruption of parasite membrane (Hartwig *et al*, 2009) or disruption of the mitochondrial electron transport chain (Li *et al*, 2005).
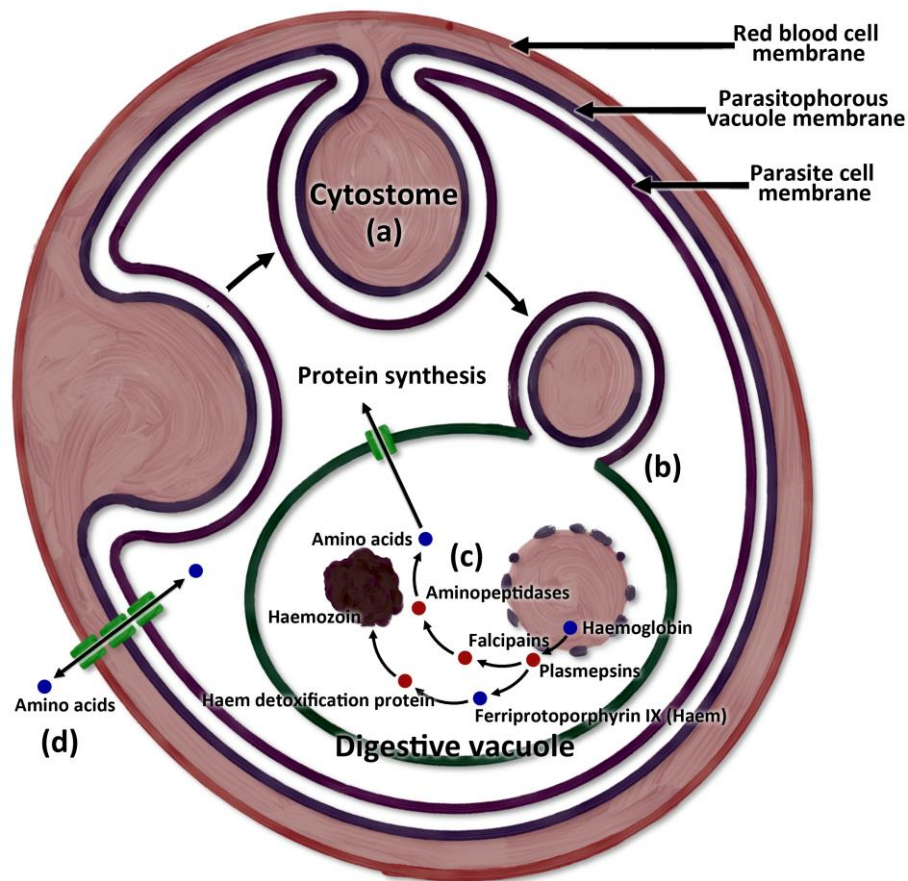
Figure 1.8 Haemoglobin transport and digestion
Haemoglobin is brought into the parasite cytosol through the formation of the double-membrane cytostome (a), which is formed by the invagination of the parasitophorous vacuole membrane and parasite membrane. This fuses with the digestive vacuole (b) where the inner membrane is digested releasing haemoglobin for digestion by several proteases and peptidases (c). Amino acids from haemoglobin are then utilised for protein synthesis. Amino acids not found in haemoglobin (i.e., isoleucine) as well as those that are of limited quantity, are obtained from the external environment. Excess amino acids from haemoglobin digestion are excreted into the external environment (d). Digestion of haemoglobin produces ferriprotoporphyrin IX (haem), which is toxic to the parasite. This is detoxified by the crystallisation of haem to haemozoin through the action of haem detoxification protein.

1.1.6 Fatty acid metabolism

Similar to nucleotide metabolism, *Plasmodium* is capable of both *de novo* synthesis and salvage of fatty acids. The parasite synthesises fatty acids in the apicoplast through the type II fatty acid synthesis pathway which is similar to prokaryotic fatty acid synthesis (Ralph *et al*, 2004; Goodman & McFadden, 2007). Fatty acid synthesis involves the repeated addition of two-carbon units onto a growing fatty acid chain. During each addition, malonyl-acyl carrier protein (ACP) condenses with a fatty acid-ACP to form a β-ketoacyl-ACP through the action of FabB/F. The terminal ketone is then reduced by FabG to an alcohol which is dehydrated to give a double bond between the two carbon atoms (alkene group) by FabZ. This alkene end is then reduced by FabI, forming an additional two-carbon to the fatty acid chain (Tarun *et al*, 2009). Specific

compounds that target this fatty acid synthesis pathway have been shown to be effective against Apicomplexan parasites such as *Toxoplasma* and *Plasmodium*. Examples of these compounds are thiolactomycin and triclosan which respectively inhibits the initial condensation of malonyl-ACP and fatty acid-ACP (FabB/F), and the final reduction of the alkene group (FabI) (Gornicki, 2003). In addition, hexachlorophene, a known anthelminthic drug, has been shown to inhibit FabG as well as *in vitro* growth of *P. falciparum* at low micromolar concentrations (Wickramasinghe *et al*, 2006). These suggest that the type II fatty acid synthesis pathway is essential to *Plasmodium* (Mazumdar & Striepen, 2007); however, gene deletion studies have shown that the pathway is not essential in blood and gametocyte stages (Yu *et al*, 2008; Vaughan *et al*, 2009). This implies that the pathway may have a different and more important role. It is suggested that the parasite regulates *de novo* synthesis and salvage depending on the availability of fatty acids in the serum (Botté *et al*, 2013). Fatty acid metabolism is summarised in Figure 1.9.
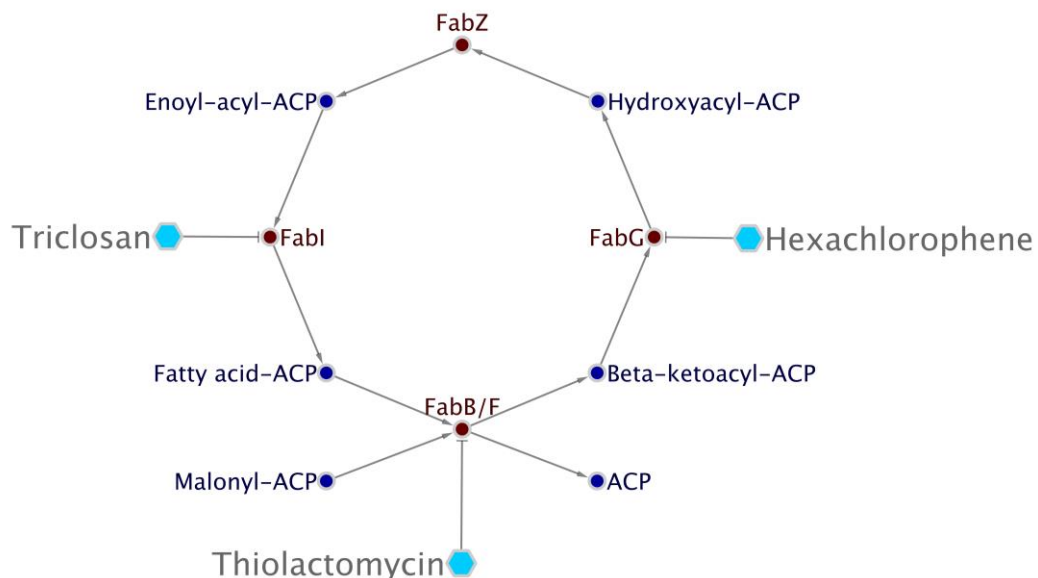
Figure 1.9 Fatty acid synthesis
Fatty acid synthesis involves the repeated addition of a two-carbon chain into a growing fatty acid chain. Malonyl-ACP condenses with a fatty acid-ACP, forming a ketone. This ketone is reduced to form a hydroxyl group which is then dehydrated. A second reduction step results in an elongated fatty acid. Drugs such as triclosan, thiolactomycin and hexachlorophene are known to inhibit specific enzymes involved in fatty acid synthesis.

1.1.7 Drug resistance

The use of chloroquine as the first line antimalarial for decades has resulted in the development of drug resistance. Chloroquine resistance has been linked to the expression of a mutant transmembrane domain protein that is able to shuttle chloroquine out of the digestive

vacuole (Schlitzer, 2007). The resistance arose from an amino acid residue mutation in what previously was a metabolite transporter that possibly transported chloride or amino acids out of the vacuole into the cytosol (Mita *et al*, 2009; Martin & Kirk, 2004). Through the years, resistance to other drugs have also been identified, such as sulfadoxine and pyrimethamine. Single and multiple point mutations in the amino acid sequence of DHPS and DHFR are linked to reduced parasite inhibition by these drugs, respectively (Triglia *et al*, 1997; Mita *et al*, 2009). Currently, there is already widespread resistance of the parasite to a number of antimalarial drugs. As a response to this issue, the World Health Organization recommended a combination therapy with an artemisinin-based compound together with a structurally unrelated drug (e.g., amodiaquine, lumefantrine, mefloquine or sulfadoxine-pyrimethamine) as the first-line treatment against malaria in 2000 (Nosten & White, 2007; Denis *et al*, 2006). However, sentinel surveillance and randomised control trials on drug efficacy against *P. falciparum* in the Thai-Cambodian border done as early as 2001 revealed emerging resistance to combination therapy (Denis *et al*, 2006; Alker *et al*, 2007; Miotto *et al*, 2013; Dondorp *et al*, 2009). Long-term or incorrect use of these drugs often leads to the development of drug resistance and the ineffectiveness of first line drugs indicates the need for more effective second line medications (Asante *et al*, 2010). Given the current limitations of existing treatment regimens, it is imperative to identify and develop newer and safer compounds against malaria. Several effective compounds have recently been identified through phenotypic screening; however, the targets of a number of these compounds remain unknown (Creek *et al*, 2016). Knowledge on the specific targets of these compounds and an understanding of the metabolic flux through the pathways may aid in further identifying and developing new antimalarial compounds (Spangenberg *et al*, 2013).

1.2 Genome scale metabolic model and flux balance analysis

1.2.1 Metabolic network reconstruction

A mathematical network of the different metabolic processes that occur in an organism is what defines a genome scale metabolic model (Monk *et al*, 2014). With the increasing availability of annotated genome data, the reconstruction of these metabolic networks has become easier. This has led to the increase in the total number of genome scale models since the reconstruction of the first model in 1999 (Kim *et al*, 2012; Edwards & Palsson, 1999). One of the main objectives of the development of these *in silico* models is hypothesis generation. These models provide a cheaper and less time consuming alternative to wet lab experiments and can surpass some technical difficulties that make experimental approaches unfeasible (Kell &

Goodacre, 2014). These models may be utilised to obtain further understanding of organisms, and to provide annotations for previously undefined genes or reactions (Chavali *et al*, 2008). These models have also been used for the bioengineering of organisms that produce or over-produce specific metabolites, most often molecules of nutritional or pharmaceutical importance (Kim *et al*, 2012; Navid, 2011). As these models take into consideration the consumption and production of different metabolites, they have also been used to identify metabolite markers for disease diagnosis (Kell & Goodacre, 2014). Most importantly, in relation to pathogenic organisms, analysis of and simulations using genome scale models have led to the identification of novel drug targets (Chavali *et al*, 2008; Plata *et al*, 2010; Huthmacher *et al*, 2010; Li *et al*, 2011).

A metabolic network reconstruction normally begins with the analysis of the organism's genome and identifying all possible genes that are eventually translated to proteins. Open reading frames (ORFs) are identified and proteins associated to these ORFs are evaluated for their protein function and location. Proteins/enzymes involved in metabolic reactions are identified and their corresponding metabolic reactions are pieced together to create a preliminary model. Utilising data on these proteins, the compartmental location of the corresponding reactions can be determined. Gaps in the metabolic pathways can then be identified and filled in using information from closely related organisms. The connection from gene to protein to reaction is incorporated into the model, and is also known as the gene-protein-reaction relationship. This relationship ensures incorporation of metabolic reactions and pathways associated with the organism's genes. Furthermore, model evaluation can utilise these connections when simulating gene knockouts. The general scheme for the development of a genome scale metabolic model is shown in Figure 1.10.

Genome scale metabolic models often fall under the stoichiometric type of model in systems biology (Bruggeman & Westerhoff, 2007). Compared to the two other types of models, i.e., regulatory and kinetic models, stoichiometric models aim to provide a more complete and holistic representation of a given organism by utilising large data sets (e.g., genomics, transcriptomics, proteomics and metabolomics). Stoichiometric models are the least complicated of the three models as it does not require input on reaction effectors and kinetic parameters. These stoichiometric models work under the steady state assumption which states that metabolite concentrations reach a point where no change occurs. The time it takes for steady state to be achieved depends on the type of reaction being modelled. In the case of enzymatic reactions, these occur rapidly in response to alterations in the environment, and steady state is achieved almost instantaneously (Palsson, 2015). Given that genome scale

metabolic models normally represent enzymatic reactions, it is therefore appropriate to consider the steady state assumption.

This is in contrast with kinetic models that require parameters such as Michaelis-Menten constant (Khodayari *et al*, 2014), which can be quite difficult when modelling organisms in a much larger scale. While kinetic models tend to focus on core metabolism, stoichiometric models can expand their coverage to incorporate as much information that can be obtained from available data.
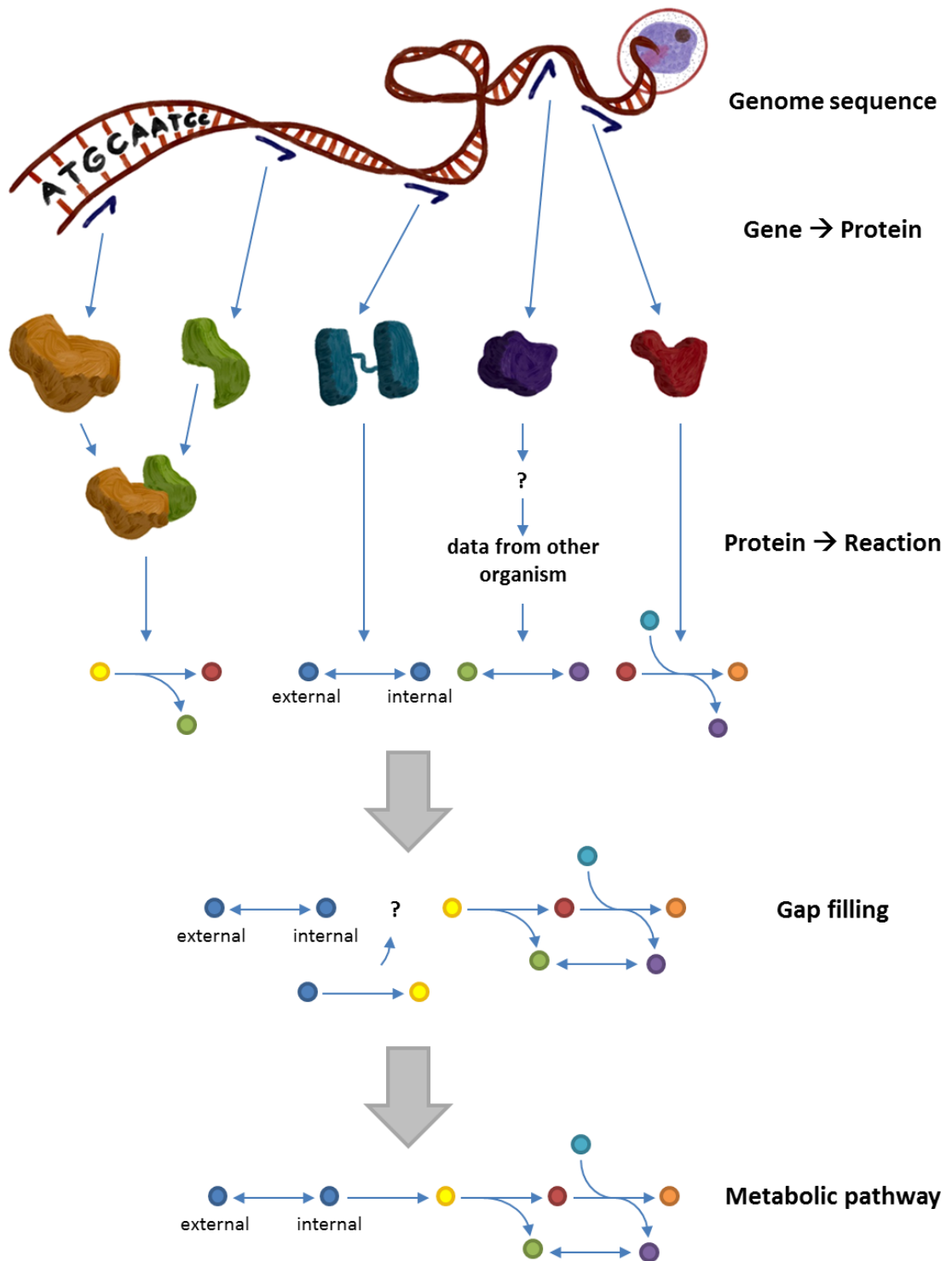
Figure 1.10 Genome scale metabolic model reconstruction flow chart
Genome scale metabolic model reconstruction begins with looking at all open reading frames in the genome sequence. Using available gene annotations, genes that are related to proteins that are involved in metabolic/enzymatic reactions are collected. Proteins that lack annotations are checked against similar proteins in other closely related organisms. The metabolic reactions are put together to form metabolic pathways using available information.

1.2.2   Important databases for malaria metabolic network reconstruction

At the moment there are several databases that can provide gene annotations, protein functions and enzyme reaction data. Among those widely accessed for the development and improvement of the model in this project is PlasmoDB. Also known as the *Plasmodium falciparum* Genome Database, PlasmoDB is part of the EuPathDB Bioinformatics Resource Center which is a group of genome databases on eukaryotic pathogens (Aurrecoechea *et al*, 2013). PlasmoDB is an online resource that is mainly focussed on the *Plasmodium falciparum* 3D7 strain; however, it has also included comprehensive data on other related species (PlasmoDB, 2001). Other malaria species included in this database are human malaria *P. vivax*; murine malarias *P. yoelii*, *P. berghei* and *P. chabaudi*; primate malarias *P. knowlesi* and *P. reichenowi*; and avian malaria *P. gallinaceum*. The database provides extensive linkages between the parasite genome and published experimental data. Transcript and protein expression, gene ontology (GO) term annotation, Enzyme Commission (EC) classification numbers and related metabolic pathways are examples of data associated with the *Plasmodium* genome presented in this database (Aurrecoechea *et al*, 2009). Accessing a gene of interest will not only provide data associated with the gene, but also external links to other relevant online databases such as the Expert Protein Analysis System database (ExPASy) (Artimo *et al*, 2012), the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa & Goto, 2000) and the Malaria Parasite Metabolic Pathways database (MPMP) (Ginsburg, 2006). It also has links to publications in PubMed. A feature of PlasmoDB that was quite useful in this project is that whole genome data can be downloaded alongside information mentioned above, old and new gene IDs, as well as reactants, products and reaction directionality/reversibility for genes associated with metabolic reactions (Aurrecoechea *et al*, 2009).

The Malaria Parasite Metabolic Pathways online database contains more than 120 metabolic pathway maps that are mostly relevant to the erythrocytic stages of the *P. falciparum* 3D7 strain (Ginsburg & Abdel-Haleem, 2016; Ginsburg, 2006). Many of these pathways were obtained from KEGG and streamlined to represent only those that are applicable to the parasite. In the absence of gene association data, some pathways were added when biochemical data that suggest functionality of the given pathway were available. The pathway maps present metabolite names that are linked to KEGG, and reaction names and EC numbers that are linked to PlasmoDB, ExPASy, GeneDB and the BRaunschweig ENzyme DAtabase (BRENDA). When available, localisation data on a given enzyme/reaction, including fluorescence microscopy images and links to publications, are provided. Similar to PlasmoDB, this website provides links to PubMed. It also provides "transcription clocks" that represent stage-dependent transcription within the 48-hour erythrocytic cycle. Apart from the common pathways like the glycolytic

pathway or the mitochondrial TCA cycle, the database also includes malaria-specific metabolic functions such as membrane and intracellular traffic, haemoglobin digestion and apicoplast function.

The Kyoto Encyclopedia of Genes and Genomes has a genome collection of more than 5,000 organisms that were obtained from RefSeq and GenBank (Kanehisa *et al*, 2016). There are a total of nine complete *Plasmodium* genome data, three of which are strains of *P. falciparum*. More than just a repository of gene and genome data, KEGG is a conglomeration of 16 databases. One of which is the KEGG Reaction database which contains different biochemical reactions. An entry in the Reaction database includes the KEGG reaction ID, reaction equation, EC number of the enzyme involved in the reaction and pathway involvement. Entries for these reactions are presented with links to other KEGG databases. KEGG Compound database contains metabolites and molecules and their associated attributes (e.g., KEGG metabolite ID, name, formula, molecular weight, structure and links to reactions and pathways they are involved in). KEGG Pathway database can be used to look at metabolic pathways in general or based on a reference organism.
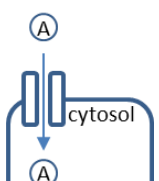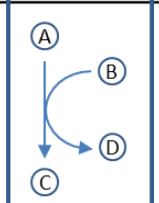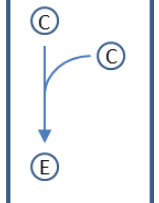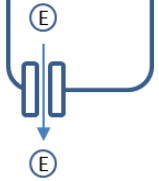
GeneDB is an online database hosted by the Sanger Institute that currently has genome data on 41 pathogenic organisms, including *P. falciparum* (Logan-Klumpler *et al*, 2012). The database works in collaboration with EuPathDB and regularly forwards annotations to EuPathDB databases such as PlasmoDB and TriTrypDB (i.e., a database on trypanosomatids). The development of GeneDB has focussed on the manual curation of trypanosomatids and *Plasmodium* species utilising more than 600 international peer-reviewed journal publications and producing more than 11,000 annotations. It also automatically receives user comments from these EuPathDB databases, and these comments are constantly reviewed and incorporated into their database as necessary. More than 1,600 comments from members of the scientific community have been incorporated into the database in order to update the annotations. The database allows users to access the organism's genome and can focus on specific genes. Annotated genes are provided with information on gene names, gene type, chromosomal location, gene ontology as well as protein data. The website provides links to other external databases such as PlasmoDB, MPMP and PubMed.

Provided by the Swiss Institute of Bioinformatics (SIB) together with 20 other collaborating institutions in Switzerland, ExPASy was initially developed to provide proteomics data but was eventually expanded to include genomics, transcriptomics, systems biology, phylogeny, population genetics and biophysics (Artimo *et al*, 2012). When conducting a query or a search, the resource returns a summary of total number of hits from and corresponding links to 14 SIB databases and 8 other external databases. The BRaunschweig ENzyme DAtabase,

or BRENDA, contains information on enzymes from about 11,300 organisms. Enzyme information that can be retrieved from this database include associated molecules (e.g., substrates, products, inhibitors and cofactors), kinetic parameters, IC_{50} values and localisation. Similar to the previously described databases, BRENDA also has links to PubMed references (Schomburg *et al*, 2004; Placzek *et al*, 2017).

1.2.3 Flux balance analysis

Using available information, a genome scale metabolic model can represent the gene-protein-reaction relationship, as previously mentioned. Figure 1.11 is a simple representation of a metabolic model, illustrating in the third column a given metabolic pathway from the entry of metabolite A into the cytosol until the exit of E into the external environment. The first column shows the genes that are translated into their corresponding protein in the second column. These proteins are linked to the corresponding reaction in the next column. The last column summarises the different metabolites that participate in each reaction. In the metabolites column, each metabolite involved in a given reaction has a stoichiometric coefficient based on the correct reaction equation. A positive number represents the production of a metabolite, while a negative number represents the consumption of the said metabolite in the given reaction. In reaction R2, for example, 2 moles of metabolite C are required to produce 1 mole of metabolite E. The relationship between reactions and metabolites is represented by the stoichiometric matrix, conventionally represented by the variable S (Thiele *et al*, 2013; Orth *et al*, 2010; Edwards & Palsson, 1999). In this matrix, the columns (m) represent the reactions and the rows (n) represent the metabolites (Figure 1.12). Stoichiometric coefficients for each metabolite are written under each of the reaction the metabolite is involved in. Going back to the previous example, reaction R2, here we can see metabolite C has a stoichiometric coefficient of -2 and metabolite E a coefficient of 1. A stoichiometric coefficient of 0 is assigned when a metabolite does not participate in a given reaction. The stoichiometric matrix will then be utilised in linear programming calculations, which will be explained later in this chapter. It is therefore important to ensure that the reactions incorporated into the model are accurate and properly mass balanced to avoid metabolic inconsistencies.

Gene	Protein	Reaction		Metabolites
tx1	TX1	Ex1		A: 1
x	X	R1		A: -1 B: -1 C: 1 D: 1
y	Y	R2		C: -2 E: 1
tx2	TX2	Ex2		E: -1

Figure 1.11 Simple metabolic model

This figure shows a simple representation of a metabolic model composed of four reactions; two of which are boundary transport reactions that bring in or out metabolites and two are internal reactions (visually represented in the third column). The last column shows the individual metabolites involved in each reaction and their corresponding stoichiometric coefficient. A negative coefficient means that the metabolite is consumed while a positive one represents the production of a given metabolite. These coefficients are then used to generate a stoichiometric matrix shown in the figure below.

$$
\begin{array}{c c}
 & \begin{matrix} \textbf{Ex1} & \textbf{R1} & \textbf{R2} & \textbf{Ex2} \end{matrix} \\
\begin{matrix} \textbf{A} \\ \textbf{B} \\ \textbf{C} \\ \textbf{D} \\ \textbf{E} \end{matrix} &
\begin{bmatrix}
1 & -1 & 0 & 0 \\
0 & -1 & 0 & 0 \\
0 & 1 & -2 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & -1
\end{bmatrix}
\end{array}
$$

Figure 1.12 Stoichiometric matrix

Apart from the stoichiometric equation, each reaction is given lower and upper flux boundaries, v_{LB} and v_{UB}, respectively. These boundaries limit the rate of metabolite consumption through a reaction, which is also known as the reaction flux. The reaction flux, conventionally represented by the variable v, is often expressed in millimole of metabolite consumed/produced per gram dry weight of the organism per hour. A reaction with a positive flux is said to be going forward, i.e., the reactants are converted to products. On the other hand, a reaction with a negative flux is said to be going in reverse, i.e., the products are converted back to their corresponding reactants. Therefore, these boundaries can also dictate the directionality of a given reaction. A reversible reaction can be represented with a negative lower boundary and a positive upper boundary, while an irreversible reaction is given a zero lower boundary and a positive upper boundary. These flux boundaries may be based on published literature and experimental values. A vector of all the fluxes for all the reactions in the model is denoted by the variable $\vec{v}$. An example of a vector of fluxes in relation to the sample model is shown in Equation 1.1.

$$\vec{v} = \begin{bmatrix} v_{Ex1} \\ v_{R1} \\ v_{R2} \\ v_{Ex2} \end{bmatrix}$$

Equation 1.1

where the reaction fluxes are between the assigned lower (LB) and upper (UB) boundaries:

$$v_{Ex1_{LB}} \leq v_{Ex1} \leq v_{Ex1_{UB}}$$
$$v_{R1_{LB}} \leq v_{R1} \leq v_{R1_{UB}}$$
$$v_{R2_{LB}} \leq v_{R2} \leq v_{R2_{UB}}$$
$$v_{Ex2_{LB}} \leq v_{Ex2} \leq v_{Ex2_{UB}}$$

The product of S and $\vec{v}$ is therefore equivalent to a vector of size n that represents the rates of change in metabolite concentration over time ($\frac{\overrightarrow{dC}}{dt}$) per gram dry weight of organism for all metabolites in the model as shown in Equation 1.2.

$$S \cdot \vec{v} = \frac{\overrightarrow{dC}}{dt}$$

Equation 1.2

where:

S = the stoichiometric matrix with m number of columns (reactions) and n number of rows (metabolites)

$\vec{v}$ = a vector of length m representing the fluxes for all the reactions in the model

$\frac{\overrightarrow{dC}}{dt}$ = a vector of length n representing the change in metabolite concentration over time for all metabolites in the model

Note that the product of a matrix and a vector is calculated by taking the sum of the products of the stoichiometric coefficient and the corresponding reaction flux, resulting in a vector with a length equal to the total number of rows, which in this case, is equal to the total number of metabolites (Equation 1.3).

$$S \cdot \vec{v} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \cdot \begin{bmatrix} v_{Ex1} \\ v_{R1} \\ v_{R2} \\ v_{Ex2} \end{bmatrix}$$

Equation 1.3

$$= \begin{bmatrix} 1 \times v_{Ex1} & -1 \times v_{R1} & +0 \times v_{R2} & +0 \times v_{Ex2} \\ 0 \times v_{Ex1} & -1 \times v_{R1} & +0 \times v_{R2} & +0 \times v_{Ex2} \\ 0 \times v_{Ex1} & +1 \times v_{R1} & -2 \times v_{R2} & +0 \times v_{Ex2} \\ 0 \times v_{Ex1} & +1 \times v_{R1} & +0 \times v_{R2} & +0 \times v_{Ex2} \\ 0 \times v_{Ex1} & +0 \times v_{R1} & +1 \times v_{R2} & -1 \times v_{Ex2} \end{bmatrix}$$

Finally, a genome scale metabolic model is assigned an objective, often times to represent the growth of the organism (Orth *et al*, 2010). This objective is also referred to as the biomass objective or the biomass reaction (B). This objective is represented by a reaction that "consumes" the necessary components of the organism's biomass. These components include nucleic acids, proteins, lipids and carbohydrates. In order to generate an accurate biomass reaction, extensive data curation is required (Monk *et al*, 2014). In the absence of published biomass data, other models have relied on data from other related organisms (Chavali *et al*, 2008; Plata *et al*, 2010; Huthmacher *et al*, 2010). The balanced stoichiometric reaction equations, the flux boundaries and the biomass objective provide constraints into the model, which is the core concept of constraint-based modelling. Constraint-based modelling works under the steady state assumption, which means that the metabolite concentrations in the model are no longer changing (Equation 1.4).

$$\frac{\overrightarrow{dC}}{dt} = 0$$

Equation 1.4

Combining Equation 1.2 and Equation 1.4, gives Equation 1.5, which states that the product of the stoichiometric matrix and the vector of reaction fluxes is equal to zero.

$$S \cdot \vec{v} = 0$$

Equation 1.5

The steady state assumption allows the use of the model without the need for enzyme kinetics data. Instead, the accuracy of the reactions in the stoichiometric matrix, which includes

the biomass reaction, and the reaction flux boundaries are what matters in deriving a solution from the metabolic model. Through constraint-based modelling, multiple flux distributions throughout the whole metabolic network at steady state can be calculated with the aid of linear programming. With the previous sample model in mind, and assigning $Ex2$ as the objective function, flux values for all the other three reactions that will maximise the $Ex2$ reaction flux can be calculated. Here, Equation 1.3 is equated to zero to satisfy the steady state equation (Equation 1.5), and together with the constraints set for all the reaction fluxes (Equation 1.1), the equation becomes the linear programming problem. The reaction flux values (v) are solved such that the value of $Ex2$ (i.e., the objective function) is maximised and the values of the individual reaction flux fall within the given lower and upper bound flux constraints (Makhorin, 2008). Note that it is possible to assign multiple objective functions and the flux through the associated reactions are maximised. This is called flux balance analysis (FBA). Through FBA, genome scale metabolic models can be utilised to predict genes and reactions that are essential to the growth of the modelled organism. Reaction knockouts can be simulated in the model by forcing a zero flux over the said reaction; while gene knockouts can be simulated by giving a zero flux on all reactions associated with the given gene. If the simulated knockout results in a zero flux on the biomass reaction, then the gene/reaction is predicted to be essential (Chavali *et al*, 2012b; Orth *et al*, 2010). By comparing the biomass reaction flux before and after a simulated knockout, one can also predict any growth reduction as a result of the knockout (Chavali *et al*, 2008; Plata *et al*, 2010).

1.2.4   *Existing malaria metabolic models*

Several metabolic models on pathogens have been developed towards the identification of potential drug targets (Chavali *et al*, 2012a, 2012b; Forth, 2012; Huthmacher *et al*, 2010; Oberhardt *et al*, 2009; Plata *et al*, 2010; Zhang & Hua, 2015). The models of Plata *et al* (2010) and Huthmacher *et al* (2010) were the earliest model reconstructions for *P. falciparum*. Taking advantage of the availability of published experimental data, these models were able to produce accurate predictions. These models have identified a number of targets, some of which have already been validated experimentally in the past and some were tested and validated *in vitro*. These two models were also utilised alongside gene expression data and managed to generate predictions on metabolite exchanges that matched experimental data. The group of Fang *et al* (2014) modified the Plata model, and with the incorporation of gene expression data, was able to model stage specific biomass, biomass component and metabolite production. Subsequently, the same group added the red blood cell metabolic model onto the existing model to represent the intraerythrocytic parasite (Wallqvist *et al*, 2016). Using published metabolite data on

infected and uninfected RBCs, external flux constraints were incorporated into the model. The resulting model was then able to predict external metabolite concentrations comparable to experimental values. On the other hand, a highly curated *P. falciparum* model was developed by Thomas Forth as part of his PhD project at the University of Leeds in 2012 (Forth, 2012). The Forth model, having only 247 reactions, may be considered small compared to the Plata and Huthmacher models with 1,001 and 1,376 reactions, respectively. However, the Forth model utilised experimentally measured biomass components (i.e., DNA, RNA and protein) extracted from *P. falciparum* culture in developing the biomass reaction, as opposed to the Plata model which based its biomass function on a previously published yeast metabolic model (Duarte *et al*, 2004). In addition, initial analysis of the Huthmacher model revealed multiple reactions that are not connected to the network. There were also a number of dead-end metabolites (i.e., metabolites that are either produced but not consumed, or consumed but not produced; this will be explained in detail in Chapter 2) that are present in the Huthmacher model (Forth, 2012). Given the limitations of these three main malaria metabolic models, namely the Forth, Plata and Huthmacher models, in terms of network size or input of accurate metabolomics and biomass data, it is important to develop a consensus model from these three existing model that will be able to provide more accurate predictions towards the identification of novel drug targets.

1.3   Research objectives

The main objective of this research was to develop a highly curated and experimentally validated *P. falciparum* 3D7 metabolic network model. This was achieved through the following specific objectives:

1. To combine the existing metabolic network model by Forth (2012) with other published models (Plata *et al*, 2010; Huthmacher *et al*, 2010)

   - This involved the comparison of the three existing models, conversion of the models using a common species and reaction ID system, and the evaluation of individual and groups of reactions (i.e., pathways) for their accuracy and correctness before inclusion into the final/combined model. This also entailed the evaluation of the merged model by looking at specific model parameters such as number of live reactions and dead-end metabolites, and correcting dead-ends by gap filling. Important reaction and metabolite data and identifiers were also added to improve the model (Chapter 2).

2. To generate experimental flux data to be incorporated into the final model, as well as *in vitro* growth assays in altered environmental conditions to be used to validate model predictions

- This involved the maintenance of continuous *Plasmodium falciparum* 3D7 *in vitro* cultures as well as the conduct of glucose and amino acid flux measurements, which were incorporated into the model. Perturbations on the *in vitro* glucose influx were also performed using mannose as a competitive inhibitor of glucose transport into the parasite to compare against model predictions (Chapter 3).

3. To identify novel drug targets and validate these targets *in vitro*

   - This involved running gene knockout simulations and comparing the predictions with published/experimentally validated essential genes. The list of novel targets were narrowed down and candidate drug targets were tested and validated through *in vitro* experiments using repurposed FDA-approved drugs (Chapter 4)

Chapter 2   Merging the three models

2.1   Introduction

Systems Biology Markup Language (SBML) is a formatting language used for representing *in silico* models such as pathways for cell signalling, gene expression and metabolic reactions. SBML enumerates a standardised set of guidelines on the format and contents of a model, and thus it provides a "common intermediate format" that can be read by software tools. The standards were developed by a community of experts in the relevant fields of research and software development, and are constantly being updated and improved. Significant SBML updates are marked as levels while minor ones are called versions. The main release is numbered as release 1 and subsequent releases may arise to address errors in the current release. SBML level 1 was introduced in 2001 and the latest release in 2010 is SBML level 3 version 1 release 1 (Hucka *et al*, 2003; Chaouiya *et al*, 2015). At the time of writing, SBML level 3 version 1 release 2 is being prepared while a preliminary release of SBML level 3 version 2 (not final) is available at the SBML website (www.sbml.org). Under SBML standards, the final model file is written in eXtensible Markup Language (XML), which is widely utilised as a standard language in bioinformatics (Achard *et al*, 2001).

A genome scale metabolic model in SBML format has a number of components that are necessary to describe the model and the metabolic reactions within it. A model is given a mandatory identifier (ID) and a name, which is an optional attribute in SBML level 3 version 1. Within the model, one can provide annotations to further describe the model such as the authors' names and affiliations (Chaouiya *et al*, 2015). This is then followed by unit definitions, mainly composed of the definition for flux in metabolic models. Flux is commonly expressed as millimole per gram dry weight per hour (Orth *et al*, 2010, 2011; Plata *et al*, 2010; Chavali *et al*, 2008); however, in some cases, flux can be expressed as mole per second (Swainston *et al*, 2016). The unit definition is then followed by the list of compartments, which describes the different compartments and their assigned attributes. Compartments represent closed spaces in the model that contain the different metabolites and in genome scale metabolic models these compartments represent cellular organelles that hold different metabolites. The external environment is also represented in the model as a separate compartment. In SBML, only the compartment ID is obligatory, while the compartment name, size and unit are optional (Chaouiya *et al*, 2015).

In an SBML model, species represent metabolites that participate in reactions and are contained in specific compartments. Species IDs and compartment (i.e., the location of the species) are mandatory attributes but the name is optional. In addition, "constant" and

"boundaryCondition" are mandatory Boolean attributes for species, and they are assigned the values "true" or "false." The "constant" attribute dictates whether the amount of species can be changed (as a result of consumption or production via a given reaction) or the quantity remains the same regardless of involvement in a given reaction. On the other hand, the "boundaryCondition" indicates whether a species is a boundary metabolite or not. Boundary metabolites are metabolites that are allowed to move out to and/or enter from the external environment (Chaouiya *et al*, 2015; Hucka *et al*, 2003). Although not mandatory, additional identifiers are recommended to easily and accurately identify the metabolites being referred to in the model (Ravikrishnan & Raman, 2015). Identifiers include database IDs such as those from Chemical Entities of Biological Interest (ChEBI) (Hastings *et al*, 2013), KEGG (Kanehisa *et al*, 2016) and PubChem (Bolton *et al*, 2008) databases. Additionally, structure-based IUPAC International Chemical Identifier (InChI) keys (Heller *et al*, 2015), Simplified Molecular-Input Line-Entry System (SMILES) (McNamara & Stearne, 2010) and molecular formulas can be utilised for a more accurate identification of metabolites in the model.

Reactions in the model describe what happens to participating species. In a metabolic model, species can either be converted to a different species or transported from one compartment to another. Based on SBML standards, a reaction must be defined by the list of reactants and products, and the reversibility at the very least. Metabolites as part of either the reactants or products are given stoichiometric coefficients as local attributes to further describe the reaction. A quantification of the reaction is also important, especially when utilising FBA (Chaouiya *et al*, 2015). Local reaction parameters that describe constraints on the reaction flux are therefore incorporated into the model. Essential parameters include the lower and upper bound flux constraints, and the objective coefficient. The lower and upper bound constraints serve as limits to the allowable flux for a given reaction (Orth *et al*, 2010), while the objective coefficient determines whether a reaction is part of the objective function of the model (Ravikrishnan & Raman, 2015; Schellenberger *et al*, 2010). Given available gene annotation data, the connection between the genes, the protein/s that are translated from the said genes and the metabolic reaction/s that arise from the said protein is represented by what is called as the gene-protein-reaction relationship (GPR). This GPR is incorporated into the model as a reaction attribute "gene association." For reactions associated with multiple genes, the gene association attribute is represented using Boolean operators "AND" and "OR" to signify whether all or any of the genes are required for the reaction to occur (Reed *et al*, 2003). This attribute is utilised when simulating gene knockouts (Ebrahim *et al*, 2013; Gevorgyan *et al*, 2011; Rocha *et al*, 2010). Additional attributes such as EC number and subsystem classification/pathway involvement can

be incorporated to further identify or describe the reaction (Forth, 2012; Duarte *et al*, 2004; Chavali *et al*, 2008; Plata *et al*, 2010).

The utilisation of different formats for identifiers in the model can result in inconsistencies and difficulty when comparing and combining models. As there is no single ontological format that is recommended for metabolic models, many have recommended or utilised additional identifiers similar to those mentioned above (e.g., InChI keys, SMILES) (Herrgård *et al*, 2008; Thiele *et al*, 2013; Ravikrishnan & Raman, 2015). The *E. coli* models: iJO1366 (Orth *et al*, 2011) and iAF1260 (Feist *et al*, 2010), human Recon 2.2 (Swainston *et al*, 2016), yeast iND750 (Duarte *et al*, 2004) models, as well as the Plata malaria model (Plata *et al*, 2010) used mainly identifiers from BRENDA or BIGG (Schellenberger *et al*, 2010). On the other hand, the Forth (Forth, 2012) and the Huthmacher (Huthmacher *et al*, 2010) malaria metabolic models used mainly KEGG (Kanehisa *et al*, 2016) identifiers for species and reactions. Other identifier ontological systems include Reactome (Croft *et al*, 2014), SEED (Aziz *et al*, 2012) and UniPathway (Morgat *et al*, 2012).

It is important for genome scale models to accurately represent metabolic reactions in a given organism (Monk *et al*, 2014; Herrgård *et al*, 2008). It is necessary to ensure that the reactions in the model do not violate the law of conservation of mass, and therefore reactions should be stoichiometrically balanced. Unbalanced reactions should be corrected or removed as even a single unbalanced reaction may result in inaccurate simulation results (Latendresse *et al*, 2012). Minimising the number of dead-end or orphan metabolites is also essential in improving a model. Dead-end metabolites are metabolites that are either produced but not consumed by any reaction or consumed but not produced by any reaction (Reed *et al*, 2003). These metabolites can create gaps in the metabolic network and can prevent flux through associated reactions. Reactions in the model that are not capable of carrying any flux are called dead or blocked reactions, as opposed to live reactions which are reactions that can carry a flux (Ravikrishnan & Raman, 2015). In assessing model quality, utilising model size (i.e., total number of genes or reactions incorporated in the model) as a parameter is not enough. It is recommended that the number of live reactions be considered as an additional parameter for evaluating model quality as these live reactions represent the section of the model that can actually be utilised to generate data and hypotheses (Monk *et al*, 2014).

COnstraint-Based Reconstruction and Analysis (COBRA) is a method that utilises genome scale metabolic models for the conduct of analysis and simulations. This method, which is packaged as the COBRA Toolbox, has different features for evaluating models (e.g., checking reaction mass balance, identifying dead-end metabolites), conducting simulations through flux balance analysis (which includes single and double gene or reaction knockouts and flux

variability analysis) and can read and create models in SBML format (Schellenberger *et al*, 2011). The COBRA Toolbox can be run through MATLAB or Python (i.e., COBRApy). In this project, COBRApy was utilised as it does not require proprietary software (Ebrahim *et al*, 2013). COBRApy requires libSBML application programming interface library in order to handle models in SBML format. LibSBML allows COBRApy to read, edit and save (write) SBML files in different SBML levels and versions (Bornstein *et al*, 2008). COBRApy also requires a linear programming software and in this project the GNU Linear Programming Kit (GLPK) package, a non-proprietary software, was used (Makhorin, 2008).

This chapter will describe how a malaria metabolic model was developed by utilising three existing malaria models (Forth, 2012; Huthmacher *et al*, 2010; Plata *et al*, 2010). The methodology for reconciling the three models for proper comparison, as well as the methods for collecting reactions to be added into the final model will be described here. In order to increase model quality, the addition of accurate metabolite and reaction attributes, and the identification and correction of gaps in the model will also be discussed. Finally, an initial assessment of model quality will be shown in comparison with the assessment of other published metabolic models.

2.2   Methodology

2.2.1   *Standardising the ontological format for reaction and species IDs*

The three malaria metabolic models used in this study were those developed by Forth (2012), Huthmacher *et al* (2010) and Plata *et al* (2010). The model files of Huthmacher (12918_2009_509_MOESM1_ESM.xml) and Plata (msb201060-sup-0004.xml) were retrieved from the supplemental files of the publication that described the said models, while that of Forth (Sup1 - ForthT_2013_pfalciparum_247reactions_pfa _iTF247_L2V4.xml) was obtained directly from the author himself (email communication). Python was used to gather data from the three malaria metabolic models and to compare the different ontological formats. Database files of IDs (chem_xref.tsv for species/metabolite IDs and reac_xref.tsv for reaction IDs) in different ontological formats alongside their corresponding MNXref ID were obtained from metanetx.org, an online support for genome-scale metabolic models that utilises the MNXref ontology as a means to reconcile all other ontological formats (Bernard *et al*, 2014). With this database, MNXref ontology was used as the format to bridge the mapping of one ID system to another. In order to evaluate the suitability of an ontological format as the format to use in the combined/final model, species and reaction IDs from the three models were mapped to their corresponding MNXref IDs and the resulting MNXref IDs were then mapped to the other

ontological formats (Figure 2.1). This provided a limitation where only original IDs with corresponding MNXref IDs can be mapped to another ontological format. IDs that cannot be mapped to MNXref and thus cannot be mapped to the final ontological format, such as A5 in the figure, are considered unique. The total number of unique IDs and the total number of IDs mapped to the other format were recorded for each model and for each ontological format.
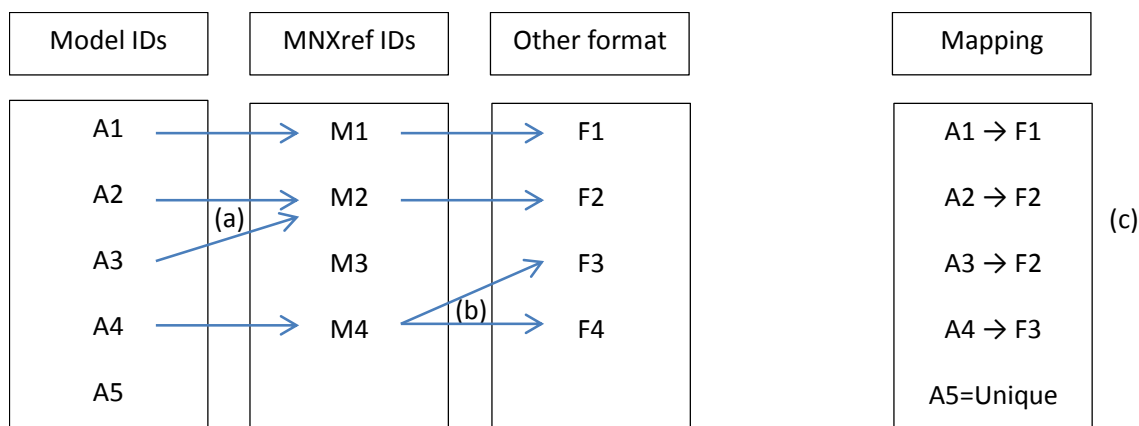
Figure 2.1   Method of mapping one ID ontological format to another
This figure illustrates the method of mapping IDs in their original format in the model to another. When two IDs in the model correspond to one MNXref ID, both were mapped to the same ID (a). On the other hand when mapping an MNXref ID to another format, when an MNXref ID corresponded to two or more IDs in the other format, the first ID in the alphabetical sequence was initially assigned (b). In the end (c), original IDs in the model were mapped to the other format.

A mapping ratio was computed for each ontological format (for species and reaction IDs) and was used to compare the different formats. The mapping ratio was computed using the formula:

$$Mapping\ ratio\ =\ \frac{total\ number\ of\ IDs\ mapped\ to\ a\ specific\ ID\ ontology}{total\ number\ of\ IDs\ mapped\ to\ MNXref\ format}$$ **Equation 2.1**

The format with the highest combined mapping ratios for species and reactions was chosen as the ontological format for the final model. Species and reaction IDs in the three models were converted to the final format to facilitate proper comparison and merging. IDs that cannot be converted to the said format were given IDs in a modified MNXref ID format. Since the character lengths of original MNXref IDs are not consistent, the modified MNXref ID format utilises the numerical portion of the original MNXref ID (to facilitate ease in tracing back IDs to its original form) that has been standardised to 9 alphanumeric characters. For example in the case of the species S-methyl-5'-thioinosine with an MNXref ID of MNXM2620, the modified ID is xpd002620. The first three characters indicate whether the ID refers to a species ("xpd") or a

reaction ("xxn") and the numerical characters from the original ID were converted into six numerical characters through the addition of zeros at the beginning (from "2620" to "002620"). It is important to note that numerical portion of all MNXref species or reaction IDs do not begin with a zero except for H(+) (i.e., MNXM01) which has its corresponding ID in all other formats. In such cases where original IDs do not have MNXref IDs, the original IDs were retained. Manual evaluation was done for IDs that were mapped to more than one MNXref ID to avoid any inconsistencies in ID conversion. The conversion of species and reaction IDs will be discussed in greater detail in Section 2.3.2.

2.2.2   *Collecting gene, enzyme and metabolite data*

Additional attributes were added to the species in the model. These include the molecular formula, charge and a few identifiers such as InChI keys (Heller & McNaught, 2009), Canonical SMILES (McNamara & Stearne, 2010) and PubChem IDs (Bolton *et al*, 2008). PubChem IDs can either be compound ID (CID) or substance ID (SID). These IDs can be used to retrieve metabolite data from PubChem. Most species data were collected using PubChemPy (Swain, 2013). PubChemPy is a Python-based package that allows compound search in PubChem using the original species ID or species name. In cases where data for a given species could not be retrieved using PubChemPy (e.g., no molecular formula), data were manually obtained from the KEGG database (Kanehisa *et al*, 2016). Molecular formula for haemoglobin was based on protein sequence data obtained from the Universal Protein Resource (UniProt) (Magrane & UniProt Consortium, 2011). It was ensured that all species have proper molecular formula in order to identify unbalanced reactions.

Attributes for reactions, whenever absent, were collected from several sources. In the case of the Huthmacher model, where reactions (mostly in KEGG ID format) do not have reaction names, EC numbers and gene association data, the reaction ID was used to obtain the EC number and the corresponding reaction name from KEGG, and using the EC numbers, the gene association was obtained from *Plasmodium falciparum* 3D7 data from the PlasmoDB database (Aurrecoechea *et al*, 2009). If the reaction ID is not in KEGG, the reaction equation (i.e., reactants and products) was used to find the corresponding reaction and reaction attributes in KEGG. As for the Forth and Plata models, where most reactions have EC numbers and gene association data (in the old ID format), reaction attributes were double checked against KEGG or PlasmoDB. Reaction attributes that did not match the KEGG or PlasmoDB data were manually evaluated and updated as needed. Using the reaction EC number, pathway information was obtained from the MPMP online resource (Ginsburg & Abdel-Haleem, 2016).

2.2.3   Identifying unique reactions from source models

The Forth model served as the minimal model to which reactions from the other models were added. This model was chosen as it was well curated and initial evaluations showed that it had the lowest percentage of dead reactions compared with the other two models. The Huthmacher and the Plata models served as the source models, from which reactions were obtained to add to the minimal model. Reactions from the source models were collected and added into the minimal model. Reactions from the source models must satisfy the following criteria before being added into the minimal model:

1.   The reaction has enzyme classification and gene association data.

2.   The reaction is not in the minimal model.

3.   At least one species in the reaction is in the minimal model.

To ensure that the reaction from the source model is not in the minimal model, reactions were compared using the reaction equations, instead of reaction IDs. The original reactants and products of a given reaction were converted into separate lists of species and these lists were compared against the lists of reactants and products of another reaction. All reactions were initially considered as reversible, thus, the reverse reaction equation was also taken into consideration when comparing reactions between models. It is important to note that some reactions utilised protonated or unprotonated species, which resulted in the presence or absence of proton/s in the balanced reaction equation. In this case, two identical reactions may be identified as dissimilar due to the presence or absence of proton/s in the equation. To avoid this error, two reactions with reaction equations that differ only by one or more protons were considered as similar reactions (Bernard *et al*, 2014).

The third criterion in adding a reaction into the minimal model involved having a common species that is present in both the reaction to be added and the minimal model. This prevented the addition of a reaction or a group of reactions that is completely separated from the original network. Furthermore, commonly participating species, i.e., species that are highly involved in a number of reactions, small molecules as well as nucleotides (e.g., H_2O, H^+, CO_2 and ATP) were excluded and were not considered as common "linking" species. This gave emphasis on more important species to link new reactions with the reactions in the minimal model.

Reactions from the source models that satisfy the three criteria were removed from the list of source model reactions and were added into the minimal model. As new reactions were being added to the minimal model, new species from these reactions were also added. Thus it was possible that reactions that initially satisfied the first two criteria but not the third could now have a common species in the growing minimal model. The remaining reactions from the source models were repeatedly assessed until no new reaction could be added into the model.

The reactions in the Plata model were first subjected into this reaction assessment loop followed by the Huthmacher reactions. After which, the remaining reactions from the Plata and Huthmacher models were again alternately reassessed until no new reactions could be added into the minimal model. The resulting combined model at this point will be referred to as the expanded model. Figure 2.2 shows the algorithm for the identification of reactions from the source model to be added to the minimal model.

Table 2.1   List of commonly participating species, small molecules and nucleotides that were excluded from the list of species that may link reactions from the source model into the minimal model

Name	SEED ID	Name	SEED ID
Acetyl-CoA	cpd00022	dUTP	cpd00358
ADP	cpd00008	FAD	cpd00015
AMP	cpd00018	FADH2	cpd00982
ATP	cpd00002	GDP	cpd00031
cAMP	cpd00446	Glycerol	cpd00100
Carbonic acid	cpd00242	GMP	cpd00126
CDP	cpd00096	GTP	cpd00038
cGMP	cpd00697	H+	cpd00067
CMP	cpd00046	H2O	cpd00001
CO2	cpd00011	H2O2	cpd00025
CoA	cpd00010	HCO3-	cpd00242
CTP	cpd00052	HO-	cpd00001
dADP	cpd00177	IMP	cpd00114
dAMP	cpd00294	NAD+	cpd00003
dATP	cpd00115	NADH	cpd00004
dCDP	cpd00533	NADP+	cpd00006
dCMP	cpd00206	NADPH	cpd00005
dCTP	cpd00356	NH3	cpd00013
dGDP	cpd00295	O2-	cpd00532
dGMP	cpd00296	Orthophosphate	cpd00009
dGTP	cpd00241	Oxygen	cpd00007
Diphosphate	cpd00012	UDP	cpd00014
dTDP	cpd00297	UMP	cpd00091
dTMP	cpd00298	Urea	cpd00073
dTTP	cpd00357	UTP	cpd00062
dUDP	cpd00978	XMP	cpd00497
dUMP	cpd00299		
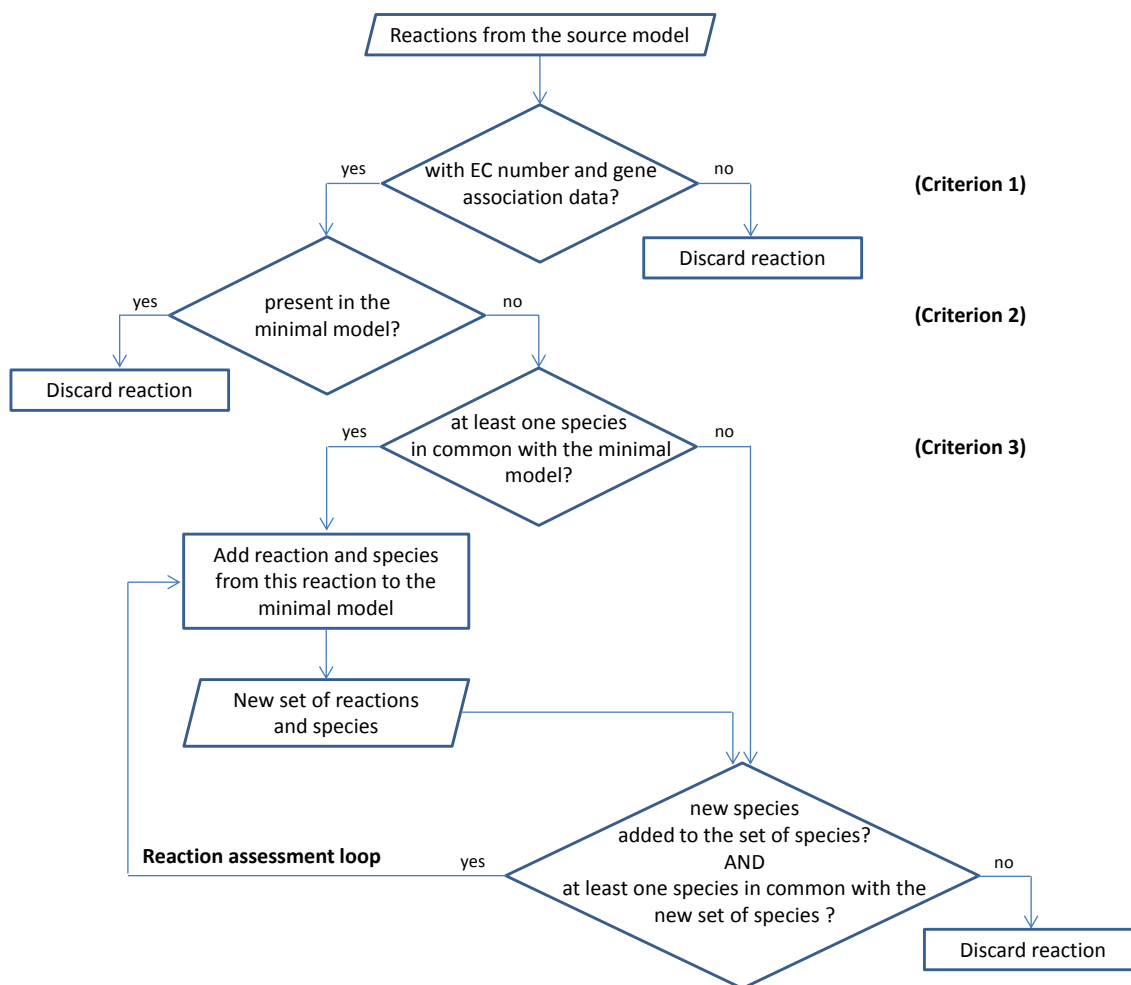
Figure 2.2   Algorithm for identifying reactions from the source models to be added into the minimal model

This figure describes how reactions from the source models were collected and added into the minimal model. Three criteria must be satisfied for a source model reaction to be added into the minimal model: Criterion 1 requires that the reaction must have an enzyme classification number and gene association data (i.e., gene ID). Criterion 2 requires that the source model reaction should not be present in the minimal model. Criterion 3 requires that the source model reaction should have at least one species in common with the species in the minimal model. As new reactions were added into the growing minimal model, new species associated with the added reactions were also added. The reaction assessment loop (Figure 2.3) allows addition of reactions that may link to previously added source model reactions. The loop terminated when no new reaction could be added into the growing minimal model.

Figure 2.3 is a visualisation of the reaction assessment loop for identifying and adding reactions from the source model to the minimal model using a simple example. In the figure, the minimal model is represented by two reactions, R1 and R2, while the source model is represented by three reactions, R3, R4 and R5. Assuming that all reactions in the example have gene and enzyme data (criterion 1), an analysis of the source model is done to identify reactions that are not in the minimal model (criterion 2) and that may be connected to the source model

through a common metabolite (criterion 3). During the first iteration, only R3 satisfies the three criteria, and is added to the minimal model resulting in a growing model. Reaction R3 is then removed from the source model and another assessment can be done. In the second iteration, both R4 and R5 satisfy the three criteria and are then added into the growing model, giving the expanded model.
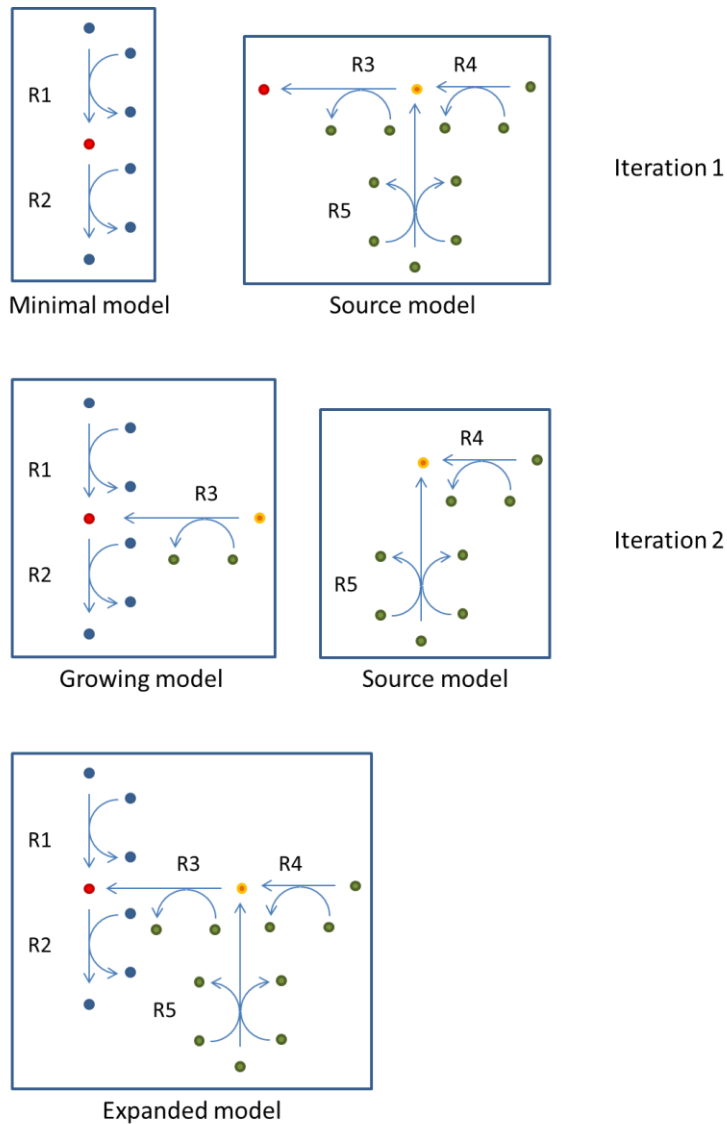
Figure 2.3   Reaction assessment loop
This is an example of the iterative process to identify and add reactions from the source model to the minimal model. An initial comparison of the reactions in the minimal and source models is done and reactions that satisfy the three criteria are added into the minimal model and removed from the source model (in this case, R3). The addition of reactions into the minimal model, which results in the addition of new species, may therefore allow new reactions in the source model to satisfy the three criteria thus additional iterations are done until no new reaction can be added into the growing model.

2.2.4   Correcting reaction direction

Data on reactions associated with *P. falciparum* 3D7 genes were downloaded from the PlasmoDB (Metabolic Pathways) database (Aurrecoechea *et al*, 2009). For PlasmoDB genes with associated metabolic reaction/s, information on EC number/s, reactants, products, pathway source and reaction direction (i.e., reversible or irreversible) were collected from the database. Reactions in the expanded model were compared against the PlasmoDB data and were corrected and updated as needed. Occasionally, a given reaction may have multiple reaction directions depending on the pathway it is involved in. In these cases, reactions and their corresponding data were evaluated manually. BRENDA (Placzek *et al*, 2017) and MPMP (Ginsburg & Abdel-Haleem, 2016) were utilised to determine the reaction direction. When reaction direction data is not available, the reaction direction was written as reversible in the model.

The flux for each reaction in the model was given lower and upper bound constraints depending on the reaction reversibility. Reversible reactions were given a default lower and upper bound constraints of -500.0 and 500.0 mmol/gDW/hr, respectively. The use of these default values for unconstrained lower and upper bound (i.e., -500 and +500) were adopted from the Forth model (Forth, 2012). Utilising other default values used in other models, e.g., 100 (Chavali *et al*, 2008) or 999,999 (Feist *et al*, 2010), did not make any difference in the values obtained from model simulations. Irreversible reactions were assigned 0.0 and 500.0 mmol/gDW/hr as lower and upper bound constraints, respectively. Reactions that were identified as reverse (reactants ← products) based on existing data were written in the opposite direction, putting the products in the reactants side and vice versa (products → reactants). This allowed uniform reaction flux constraints where the lower bound constraint is always less than the upper bound constraint. This standardisation of reaction direction was necessary as having an upper bound constraint that is less than the lower bound constraint can result in an error when conducting model simulations in some applications/programs.

2.2.5   Addressing dead-end metabolites

COBRApy (Ebrahim *et al*, 2013) was used to generate a list of nodes from the reactions in the expanded model for visualisation and analysis using Cytoscape 3.2.1 (Kohl *et al*, 2011). Cytoscape requires as an input file a list that is composed of two columns, namely, the source and target nodes at minimum. Each row contains a pair of nodes where the node in the source column is connected by a directed arrow, also known as an edge, to the node in the target column. For each reaction, every reactant was assigned as a source node and the reaction as the target node, while the reaction was assigned as a source node for every product, which was the

target node. If the reaction is reversible, then every product was also assigned as a source node with the reaction as the target node. Additionally, the reactants were assigned as target nodes for the reaction as the source node (Figure 2.4).

Reaction	Equation	
Hexokinase	Glucose + ATP → ADP + Glucose-6-phosphate	(A)
Phosphoglucose isomerase	Glucose-6-phosphate ←→ Fructose-6-phosphate	

Source nodes	Target nodes	
Glucose	Hexokinase	
ATP	Hexokinase	
Hexokinase	ADP	
Hexokinase	Glucose-6-phosphate	(B)
Glucose-6-phosphate	Phosphoglucose isomerase	
Phosphoglucose isomerase	Fructose-6-phosphate	
Fructose-6-phosphate	Phosphoglucose isomerase	
Phosphoglucose isomerase	Glucose-6-phosphate	

(C)

Figure 2.4   Visualisation of reactions using Cytoscape
This diagram shows the process of generating a Cytoscape visualisation of the expanded model, partly represented by (A) as a list of reactions and their corresponding reaction equation. Note that in the example, hexokinase is shown as an irreversible reaction while phosphoglucose isomerase is represented as a reversible reaction. A nodes list (B) is generated from the list of reactions, taking into account the reaction direction and reversibility, and is used as the input for Cytoscape. The resulting Cytoscape network (C) represents the "flow" of reactants and products (in blue) through the reactions (in red). Reversible reactions are represented in this figure as double headed arrows.

By generating a Cytoscape directed network, the total number of incoming and outgoing edges can be calculated for each node in order to identify dead-end metabolites. These edges represent the reactions that produce ("InDegree") or consume ("OutDegree") a given metabolite. Furthermore, a non-directed network (i.e., a network that disregards directionality of reactions) can be used to identify the number of reactions connected to each metabolite (i.e, simply known as "Degree"). Metabolite nodes with an InDegree or OutDegree equal to 0, regardless of Degree value, were considered dead-end metabolites. Additionally, metabolites with a Degree equal to 1, regardless of InDegree and OutDegree values (i.e., limited to either 0 or 1), were also considered as dead-end metabolites. Table 2.2 presents a summary of InDegree, OutDegree and Degree values for the identification of dead-end metabolites. The blue nodes represent the metabolite being evaluated and the edges represent reactions that consume or produce the given metabolite. A list of dead-end metabolites and their corresponding reactions was generated and manually reviewed. Data from KEGG (Kanehisa *et al*, 2016) and MPMP (Ginsburg & Abdel-Haleem, 2016) were used to fill in gaps. Reactions with dead-end metabolites that were not in the MPMP database were removed from the model. Metabolites that appear as dead-ends in MPMP were assigned a boundary reaction or a reaction that can transport the metabolite to another compartment. To each of the gap filling reactions, an additional annotation signifying that the reaction was added to correct dead-end metabolites was written as a reaction attribute labelled as "OTHER_NOTES." This allows future users to easily identify these reactions.

Table 2.2   Identification of dead-end metabolites using Cytoscape*

Visualisation	Description	InDegree	OutDegree	Degree
	Consumed but not produced	0	n	n
	Produced but not consumed	n	0	n
	End of a reversible or irreversible pathway	≤ 1	≤ 1	1

*n represents any positive integer

2.2.6 Deriving the biomass equation

Previous experimental data on the proportions of proteins, DNA and RNA in terms of weight of macromolecules per dry weight of parasite (Forth, 2012) were incorporated into the derivation of the biomass equation. This entailed growing *P. falciparum* 3D7 strains *in vitro* in Roswell Park Memorial Institute (RPMI) 1640 growth medium (Life Technologies) supplemented with 5% (w/v) Albumax I (Gibco), 2% (w/v) sodium bicarbonate (Sigma), 0.01% (w/v) hypoxanthine (Sigma) and 0.1% (v/v) gentamicin at 5% haematocrit (O+ blood obtained from St. James's University Hospital). Cultures were grown in a 37°C incubator at 1% oxygen, 3% carbon dioxide and 96% nitrogen gas mixture. Parasites were synchronised using 5% sorbitol and were allowed to grow to late schizont stage before parasite isolation. A total of 1,080 ml of synchronised culture was utilised for biomass measurements. RBCs were collected and lysed using 0.15% saponin. Parasites were isolated by high-speed centrifugation. Parasite dry weight was measured prior to quantification of biomass components. Protein content of the isolated parasites was measured using Bradford Assay (with bovine serum albumin as standard) and the percentage of protein by weight of total biomass (dry weight) was calculated. DNA and RNA content were measured using a nanodrop spectrophotometer. Total carbohydrate and lipid compositions were estimated using a 27:15 carbohydrate to lipid ratio of the remaining biomass proportion, as adopted from the biomass function of *L. major* (Chavali *et al*, 2008).

Stoichiometric coefficients for reactants and products to make 1 gDW of parasite in the biomass reaction were calculated using Equation 2.2 (using a single amino acid, as an example). This simplified equation was adopted from that used by Chavali *et al* (2008) in calculating the stoichiometric coefficients of the *L. major* biomass reaction.

$$c_{amino\ acid} = \frac{m_{protein} \times p_{amino\ acid} \times 1\ gDW_{parasite} \times 10^3\ \frac{mmol}{mol}}{MW_{protein}} \qquad \textbf{Equation 2.2}$$

where:

$c_{amino\ acid}$ = stoichiometric coefficient
$m_{protein}$ = gram of protein per gram of parasite dry weight
$p_{amino\ acid}$ = percentage of amino acid in parasite proteome
$MW_{protein}$ = weighted average molecular mass of parasite protein (gram/mole)

The weighted average molecular mass for each biomass component (protein, DNA, RNA, carbohydrates and lipids) was calculated using the published proportions or ratios of individual subcomponents and the corresponding molecular mass. Individual amino acid proportions were estimated based on the relative abundance of each amino acid in *P. falciparum* (Chanda *et al*, 2005). Individual deoxynucleotide proportions were estimated using published data on the G+C

content of the *P. falciparum* 3D7 genome while RNA nucleotide proportions were estimated by taking the weighted average of the G+C content of exons and introns (Gardner *et al*, 2002). The proportions of the carbohydrate subcomponents in the original model were retained. These subcomponents were limited to GDP-mannose for carbohydrate accumulation, adapted from published *Leishmania major* data (Ralton *et al*, 2003), and GDP-fucose for glycosylation (Forth, 2012). GDP-mannose to GDP-fucose ratio was revised using data from published experimental quantifications of sugar nucleotides in *P. falciparum* (Sanz *et al*, 2013). The lipid component in the biomass function of the original model was also updated. For simplicity, the lipid component was not divided into different proportions of fatty acid species. Instead, the components were subdivided into three: fatty acids, desaturated fatty acids and C_2H_4 elongation unit. The proportions of these three were adopted from the original model which were also based on experimentally published data (Forth, 2012; Mi-Ichi *et al*, 2007). The weighted average molecular mass for fatty acids and desaturated fatty acids were recalculated and updated using published experimental data on the proportion of different fatty acid species in *P. falciparum* 3D7 strains (Botté *et al*, 2013), while the molecular weight of the elongation unit, 28.05 g/mol, was calculated based on the chemical formula.

A multiplier (10^3) was added to the equation to convert the unit of the stoichiometric coefficient from moles to mmoles. Additional metabolites (ATP, GTP, AMP, GDP, orthophosphate and water) were added to the biomass reaction to account for amino acid elongation. Two high energy bonds are consumed in the synthesis of an amino acyl-tRNA (Equation 2.3) for activation and transfer steps and hydrolysis of pyrophosphate, while two GTP molecules are required for 1) positioning of amino acyl-tRNA into the A site of the ribosome; and 2) release of the elongation factor and formation of peptide bond (Equation 2.4) (Stryer, 1988). Therefore, one mole of ATP, two moles of GTP and three moles of H_2O were added for every mole of amino acid to account for amino acid elongation. One mole of H_2O for every mole of GDP-fucose and GDP-mannose for the hydrolysis of these components was also added into the biomass equation. A molecular formula for the parasite biomass was estimated by obtaining the sum of all atoms in the molecules included in the biomass components.

Amino acid + ATP + tRNA + 2 H_2O → Amino acyl-tRNA + AMP + 2 P_i **Equation 2.3**

2 GTP + 2 H_2O → 2 GDP + 2 P_i **Equation 2.4**
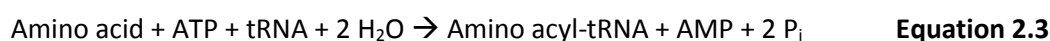
To ensure that the biomass reaction is mass balanced, all stoichiometric coefficients were multiplied by 1000 to convert these coefficients into integers since the smallest coefficient was 0.003. The molecular formula for the biomass metabolite was then derived from the sum of all

atoms of all metabolites that contribute to the parasite's biomass, excluding those associated with amino acid elongation. COBRApy was then used to check for reaction mass balance. Note that the original coefficient (using Equation 2.2) was retained as the coefficients in the model.

The final model was created using COBRApy and was saved an XML file in SBML version 3 level 1 format, using the flux balance constraint package (Chaouiya *et al*, 2015). The source codes for reading and writing SBML files were slightly modified to accommodate the addition of other attributes to metabolites (e.g., InChI keys, SMILES).

2.2.7   *Ensuring reaction mass balance*

Reaction mass balance was assessed using the check_mass_balance() function in COBRApy to ensure that the reactions follow the law of conservation of mass. This function evaluates the reaction stoichiometry by looking at the molecular formula and the stoichiometric coefficients of the reactants and products. If a reaction is not mass balanced, the function will identify unbalanced atoms and a corresponding number of atoms in excess. This output is in the form of a Python list as shown in the example below:

['rxn00001', {'C': 0.0, 'H': -2.0, 'O': 1.0, 'N': 0.0, 'P': 0.0}]

The output list contains two items: the first one being the reaction ID ('rxn00001') and the second is a dictionary of the atoms involved in the reaction. A dictionary is a Python object type that consists of a series of paired objects (i.e., key and value). In the case of the dictionary output for the said function, the keys are atoms while the values are the total number of atoms in excess. A negative stoichiometric coefficient indicates an excess in the reactant side, while a positive coefficient is an excess in the product side. The example shown above means that the reaction with a reaction ID of 'rxn00001' has two extra hydrogen atoms in the reactant side, and an extra oxygen atom in the product side. On the other hand, if a reaction is balanced, the function will return an empty list, represented by square brackets ([ ]). In some cases, metabolite formulas have an R-group (i.e., Markush structure) resulting in unbalanced reactions. Using data from KEGG and PubChem, formulas for metabolites with these generic R-groups were standardised, and reactions involving these metabolites were evaluated manually to ensure mass balance. In the final model, there are 68 metabolites that contain R-groups in their molecular formulas, and in most of which, the R-group represents a generic acyl-group. Only seven metabolites have an R-group in their molecular formula that does not represent an acyl-group. These metabolites include apocytochrome, ferri-/ferrocytochrome, thioredoxin and carboxylase-carrier protein which are recycled cofactors. All other unbalanced reactions were corrected using reaction data from KEGG.

2.2.8   Identifying dead-end metabolites and live reactions for model evaluation

A single biomass solution (i.e., calculated maximised objective function value) may be achieved through different combinations of reaction fluxes. Flux variability analysis (FVA) looks into the flux range of all reactions in the model that can result in the maximised biomass output. FVA therefore can also identify reactions that are not capable of carrying any flux (i.e., dead reactions). FVA was performed using the biomass reaction as the objective function through COBRApy (Ebrahim *et al*, 2013), and this returned a minimum and maximum flux for each reaction. Reactions that could not carry a flux, where the minimum and maximum are both equal to zero, were considered dead reactions. Otherwise, reactions that could carry a flux were considered as live reactions. It is important to note that COBRApy runs using Python and Python utilises floating-point math in its calculations. Floating-point math tends to give out inexact zeros, where the values that should be equal to zero are returned as very small values that are close to zero (Lutz, 2013). Thus, this was taken into consideration by setting values between -1 x 10^{-10} and 1 x 10^{-10} as zero. MetExplore (Cottret *et al*, 2010) is an online resource that can be used to create, evaluate and run simulations using metabolic models. The XML version of the model was uploaded into the MetExplore website and was analysed for dead reactions and dead-end metabolites in the model. Public models available in MetExplore, as well as the Plata (Plata *et al*, 2010) and Huthmacher (Huthmacher *et al*, 2010) models were also evaluated for dead-end metabolites and dead reactions and compared with the final model.

2.3   Results

2.3.1   Characteristics of the three models

The original malaria metabolic models of Forth (Forth, 2012), Huthmacher (Huthmacher *et al*, 2010) and Plata (Plata *et al*, 2010) were all written in SBML (version 2 level 1) and are in XML file format. Although data on the network characteristics were available from the corresponding publications, the actual model files were evaluated to obtain the data shown in Table 2.3. It is important to note that for the Forth model, a total of 247 modifiers were included in the list of species to represent enzymes (with or without EC numbers). These were not included in the total number of species in the table below and in further comparisons with the species in the two other models. Common to all three models are four compartments, the apicoplast, cytosol, mitochondria and the extracellular compartment. In addition to these four compartments, the Forth model also has a compartment representing the food vacuole, while the Huthmacher model has an additional four compartments representing the endoplasmic reticulum, food vacuole, Golgi and nucleus.

Table 2.3   Network characteristics of the three models

Model	Compartments	Genes	Species	Reactions
Forth	5	143	267	247
Huthmacher	8	579	1673	1376
Plata	4	366	1025	1001

The species in the three models included the following species attributes: species ID, name, compartment location and boundary condition. None of the models included metabolite identifiers such as SMILES (McNamara & Stearne, 2010) or InChI keys (Heller *et al*, 2015). As for chemical formula, although the Plata model incorporated the metabolite formula into most of the species names, none of the models have included chemical formula as a species attribute. Having a formula incorporated as a species attribute is important in assessing reaction mass balance. Water and ATP were the two most commonly participating metabolites in both the Forth and the Huthmacher model, while H^+ and water were for the Plata model (Figure 2.5). It is worth mentioning that more than 50% of reactions in the Plata model involved H^+. The Plata model utilised different databases including KEGG (Kanehisa *et al*, 2008) and PlasmoCyc (a part of BioCyc) (Yeh *et al*, 2004) in model reconstruction. KEGG for example presents many of its species in the protonated form, while BioCyc has many of its metabolites in the unprotonated form. Figure 2.6 shows the difference in the chemical formula of ATP between KEGG and BioCyc as an example. Although the proportion of reactions obtained from different databases were not mentioned in the publication (Plata *et al*, 2010), it is possible that many of the reactions in the Plata model involved unprotonated forms of the metabolites thus requiring the addition of protons in the reaction to balance the equation.
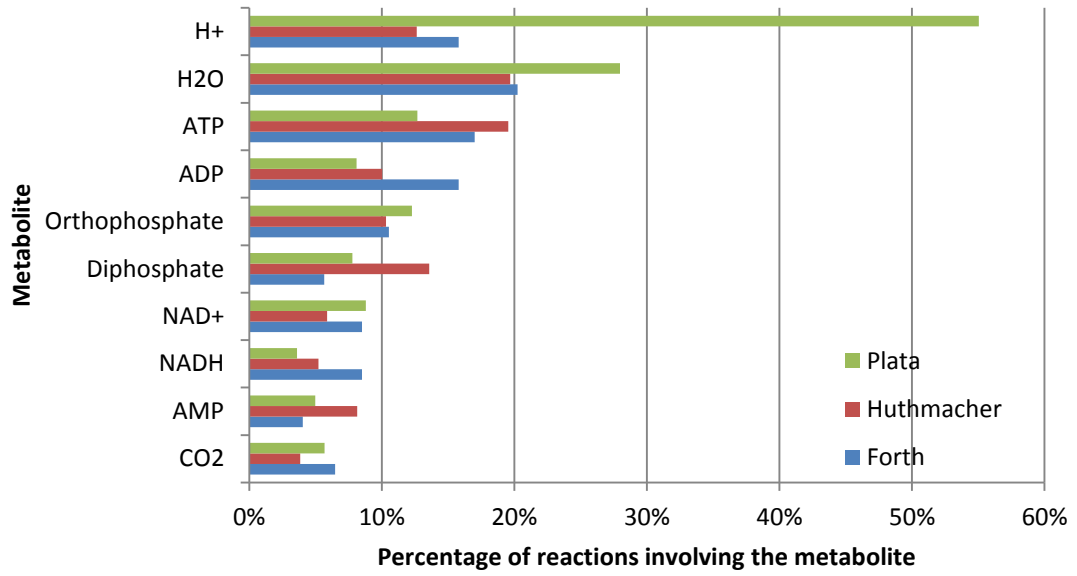
Figure 2.5   Top commonly participating metabolites
This figure shows the percentage of reactions that involve the top ten most commonly participating metabolites in the three models.

$$C_{10}H_{16}N_5O_{13}P_3 \qquad\qquad C_{10}H_{12}N_5O_{13}P_3$$

KEGG   BioCyc
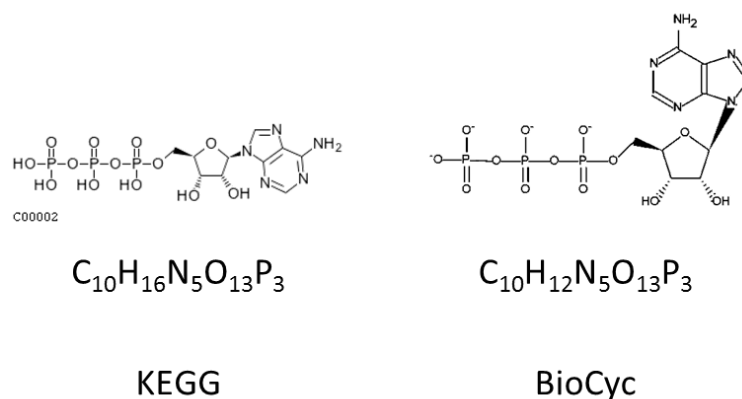
Figure 2.6   Adenosine triphosphate as presented in KEGG and BioCyc
This is a comparison of the molecular structure and formula of ATP as represented in the KEGG (Kanehisa *et al*, 2016) and BioCyc (Caspi *et al*, 2014) databases. It is important to keep these differences in mind when comparing reaction equations as similar reactions may have different reaction equations.

Whenever available, attributes were included in the Forth model reactions. These attributes include gene association in the old *Plasmodium* gene ID format, protein class (EC number), subsystem number, which corresponds to KEGG pathway map (Kanehisa *et al*, 2016), SHARKhunt e-value (Pinney *et al*, 2005) and BRENDA hit (i.e., data evidence from BRENDA). Lower and upper bound flux constraints, as well as the reaction reversibility (i.e., true or false) for each reaction were also included in the model. A default value of ±500 mmol/gDW/hr was used in this model; however, some experimentally obtained flux values were used as upper and lower boundary constraints for their respective boundary reactions. None of the reactions in the Huthmacher model had specific reaction names. Instead, the reaction ID was also assigned as the reaction name. The reaction reversibility was also included in each reaction; however, lower and upper boundary constraints were not included in the model file. None of the reactions in this model had gene, enzyme or subsystem data. Apart from the reaction ID, name and reversibility, reactions in the Plata model also included the EC number, subsystem (pathway) data, reaction equation and gene association (also in the old *Plasmodium* gene ID format) when available. Flux values were set at default value of ±999999.0 mmol/gDW/hr. A majority of the Forth reactions (199, 80.6%), and all of those in the Huthmacher model are reversible, while a majority (589, 58.8%) of the reactions in the Plata model are irreversible.

Common to all three models, most of the reactions were cytosolic while transport reactions accounted for the second most common. Unique to the Huthmacher model were the three compartments, the endoplasmic reticulum and the nucleus which housed 3.5% of all reactions in the model, and the Golgi which did not contain any metabolic reaction. Instead, metabolites were merely transported to and from this compartment. As for the Plata model,

there were an additional 113 (11.3%) reactions that facilitate transport of metabolites from the extracellular matrix out into the boundary. These reactions are shown as part of the reactions in the extracellular space in Figure 2.7.
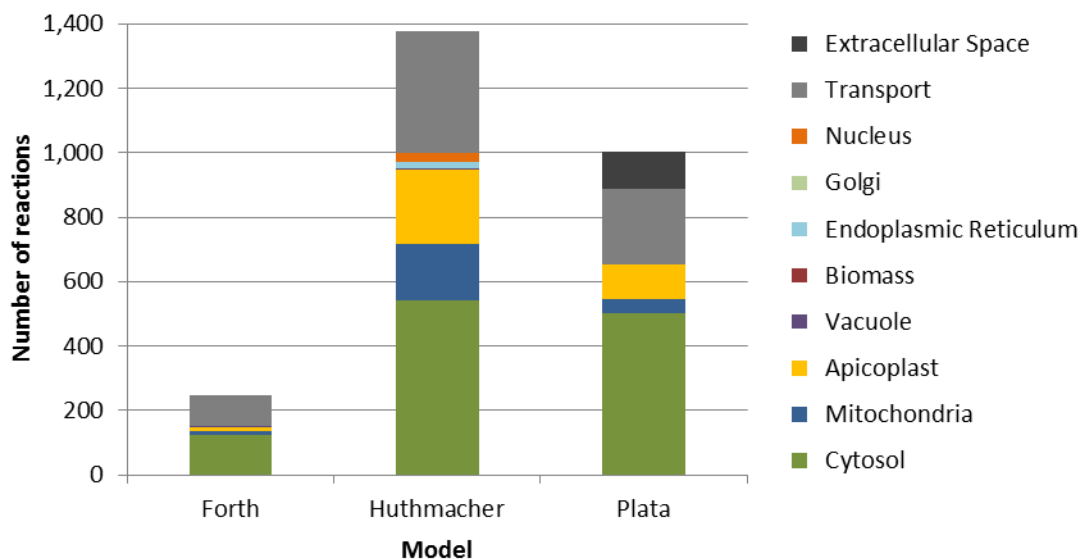
Figure 2.7 **Number of reactions by compartment in the three models**
Shown here are the total counts of the reactions in their respective compartments as well as transport reactions in the three models. In light grey is the total number of transport reactions that bring metabolites from one intracellular compartment to another or to the extracellular space. Unique to the Plata model are reactions that transport metabolites from the extracellular space to the boundary (shown in dark grey).

2.3.2   *Standardising the ontological formats for reaction and species IDs*

The Forth model represented its species mainly using KEGG IDs, with the addition of four metabolites. Metabolite IDs DesatFA_c and FAC2H4unit_a were given to represent a desaturated fatty acid and C_2H_4 unit elongation of a fatty acid, respectively. These metabolites were used as part of a simplified fatty acid desaturation and elongation reactions. Haemozoin, both as a boundary and vacuole metabolite, were assigned the IDs Hemozoin_b and Hemozoin_v, respectively (Forth, 2012). The Huthmacher model utilised the KEGG ID system, except for a total of 82 species, which were given unique identifiers to represent metabolites such as glycoproteins, generic mRNA, haem and haemozoin (Huthmacher *et al*, 2010). There were a few IDs (12) that follow the KEGG format; however they were not part of the KEGG compound database (e.g., C00660 representing D-glucose 1,6-bisphosphate). The Plata model utilised the BIGG ontology system in most of the species ID, except for five IDs which do not have a corresponding BIGG ID (e.g., haemoglobin, 5'-methylthioinosine) (Plata *et al*, 2010).

A majority of the reactions (149, 60.3%) in the Forth model were represented using KEGG IDs except for 95 (38.5%) transport reactions and three (1.2%) non-transport reactions.

Transport reactions were given IDs composed of the metabolite name, the compartments between which the metabolite is transported, and a suffix of "s" signifying that it is a transport reaction. For example, the reaction ID "NADPH_mtoc_s" represents a reaction that transports NADPH between the mitochondria and the cytosol, while "O2_btoc_s" represents boundary reaction that transports molecular oxygen between the external compartment and the cytosol. Three non-transport reactions were given unique reaction IDs: "SFAE_a," "OHGbDigV_v" and "MalariaBiomass_plus16maint_s" representing a simplified fatty acid elongation reaction, oxyhaemoglobin digestion and the biomass reaction, respectively. Similar to the Forth model, the Huthmacher model used KEGG IDs for a majority (802, 58.3%) of its reaction, while transport reactions (378, 27.5%) were given unique IDs (i.e., uppercase "T" followed by a number). There were 196 (14.2%) non-transport reactions with unique IDs that are not in KEGG. Most of the Plata reactions were in the BIGG format, except for four (0.4%).

A total of 11 ontological formats for species IDs and nine for reaction IDs were included in the metanetx.org database files as summarised in Figure 2.8. The Chemical Entities of Biological Interest (ChEBI) ontology had the most number of species IDs compared with all other formats (Degtyarenko *et al*, 2008), while Reactome (Croft *et al*, 2014) had the most number of reaction IDs. Analysis of the database files revealed that a single MNXref ID may correspond to more than one ID of a different format. In some cases, the duplication is a result of IDs corresponding to the same species/reaction, while in other cases, the IDs may refer to specific types or nomenclature formats of a given compound (Table 2.4). The occurrence of these duplications presented as a problem in terms of the accuracy of the conversion from one ontological format to another. However, for the purposes of identifying a suitable format for the final/combined model, the occurrence of these duplications was disregarded. Nevertheless, in the conversion of the models into the final format, manual double-checking of the reactions and species ID was done to guarantee accuracy of ID conversion to the final format.
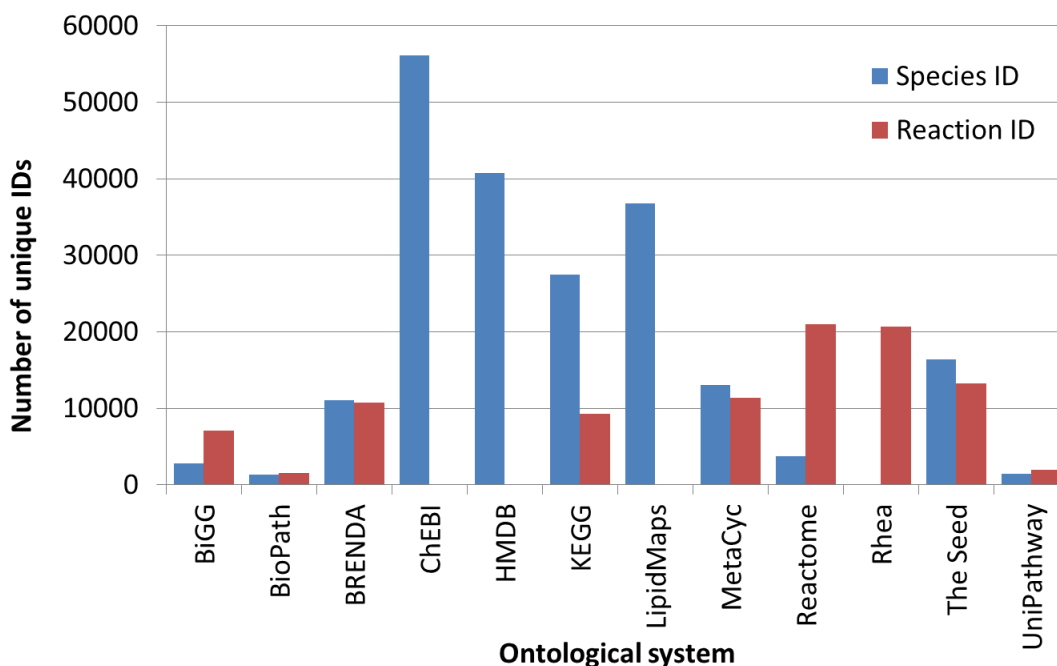
Figure 2.8   Total number of species and reaction IDs by ontological system
This figure shows a summary of the total number of species and reaction IDs by ontological system obtained from metanetx.org.

Table 2.4   Examples of MNXref IDs corresponding to more than one ID in an alternative ontological format (e.g., KEGG)

MNXref ID	KEGG ID	Compound name/s
MNXM1004	C02591	Sucrose 6'-phosphate; Sucrose 6F-phosphate; Sucrose 6'-phosphate
MNXM1004	C16688	Sucrose 6-phosphate; 6-O-Phosphonosucrose; Sucrose 6-phosphate; beta-D-Fructofuranosyl-6-O-phosphono-alpha-D-glucopyranoside; 6-Phosphosucrose
MNXM101154	C07514	Amphetamine; Amfetamin (TN); Amfetamine (INN)
MNXM101154	D07445	Amphetamine; Amfetamin (TN); Amfetamine (INN)

The calculation of the mapping ratio (Equation 2.1) facilitated the identification of a suitable ontological format that has the most number of IDs that correspond to the original IDs in the three models (Figure 2.9). The SEED format had the highest number of IDs corresponding to the original IDs in the models. The SEED is an annotation environment that allows experts from a variety of fields to provide information on subsystems or protein families within a certain genome (Overbeek *et al*, 2005). This system utilises the Model SEED biochemistry database that contains non-redundant species and reactions from the KEGG database in addition to curated species and reactions from a number of published metabolic models (Aziz *et al*, 2012). IDs in the SEED format starts with either "cmp" (for compounds or species) or "rxn" (for reactions)

followed by 5 numeric characters. The consistency of the number of characters in the ID, and the use of simple alphanumeric characters allow easier search and evaluation of specific IDs within a model file. This is in contrast to ontological formats that use IDs with varying character lengths such as BIGG (Schellenberger *et al*, 2010), BioPath (Reitz *et al*, 2004), BRENDA (Placzek *et al*, 2017), ChEBI (Degtyarenko *et al*, 2008) and MetaCyc (Caspi *et al*, 2014), where it is possible that the specific string of characters of short IDs can be found within other IDs (e.g., "amine" vs. "5-prime-phospho-beta-d-ribosylamine"). Furthermore, the use of non-alphanumeric characters such as (_), (-) and (+) in BIGG, BioPath and MetaCyc can be a source of error or confusion during encoding (e.g., "asp__L", "Acyl(n+3)-CoA", "+-bornyl-diphosphate"), and some use characters that are not allowed in defining an ID in SBML (Chaouiya *et al*, 2015).
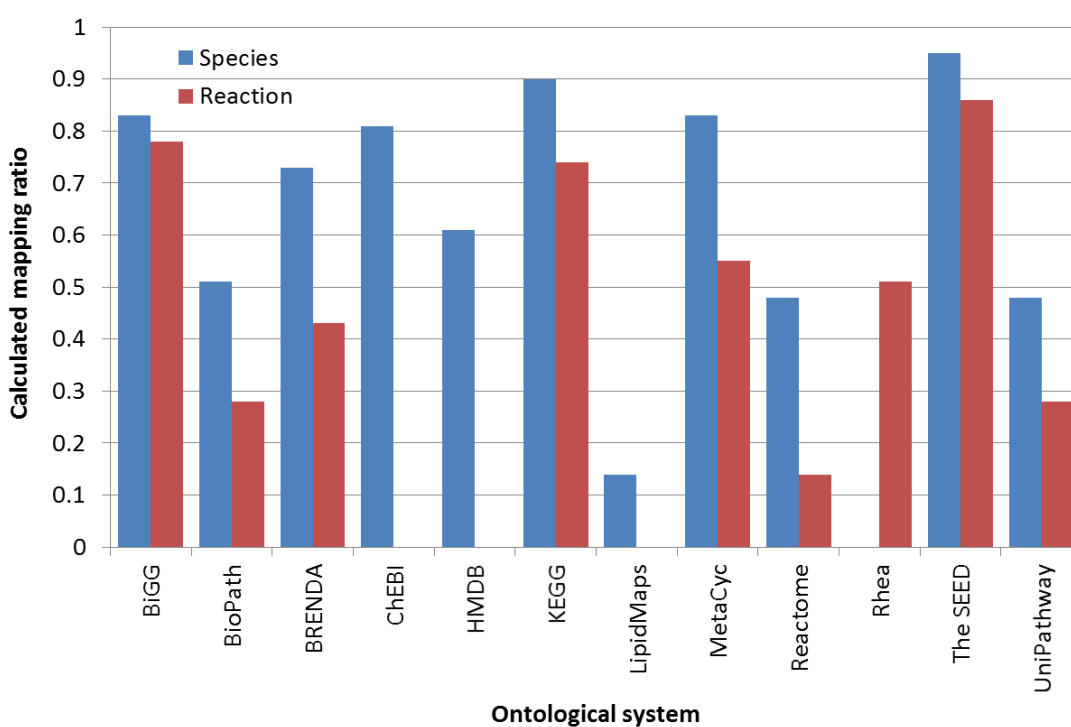
Figure 2.9   Species and reaction ID mapping ratios by ontological system
Mapping ratios were calculated as a means to represent the success in conversion of the original species and reaction IDs in the original models to a given ontological system. Shown here are the cumulative mapping ratios covering all three models. Compared with the other ontological systems, the largest ratio of the original IDs in the three models were converted to IDs under the SEED ontological system (Aziz *et al*, 2012).

Conversion of species and reaction ID initially involved automated conversion using the metanetx species and reaction database file done in Python. In cases where the original ID did not match any ID, these were manually evaluated against KEGG (Kanehisa *et al*, 2016), BIGG (Schellenberger *et al*, 2010) and SEED (Aziz *et al*, 2012) databases to ensure accurate conversion.

It was noted that some models used multiple species IDs for the same species while others utilised multiple species to represent stereoisomers, as in the case of α- and β-glucose, and α- and β-glucose 6-phosphate in the Forth model. These species were standardised for ease in the comparison of the three models. Table 2.5 shows the original IDs and names of species that were standardised alongside the revised IDs in SEED format and the standardised names.

Table 2.5 Standardised species IDs

Old ID	Original name in model	SEED ID	Standardised name
Forth model			
C00267	alpha-D-Glucose	cpd00027	D-Glucose
C00221	beta-D-Glucose	cpd00027	D-Glucose
C00668	alpha-D-Glucose 6-phosphate	cpd00079	D-Glucose 6-phosphate
C01172	beta-D-Glucose 6-phosphate	cpd00079	D-Glucose 6-phosphate
C01353	Carbonic acid	cpd00242	HCO3-
C00288	HCO3-	cpd00242	HCO3-
Huthmacher model			
U00039	2-methyl-1-hydroxybutylthiamine diphosphate	cpd14702	2-methyl-1-hydroxybutylthiamine diphosphate
C15978	2-Methyl-1-hydroxybutyl-ThPP	cpd14702	2-methyl-1-hydroxybutylthiamine diphosphate
U00038	3-methyl-1-hydroxybutylthiamine diphosphate	cpd14698	3-Methyl-1-hydroxybutyl-ThPP
C15974	3-Methyl-1-hydroxybutyl-ThPP	cpd14698	3-Methyl-1-hydroxybutyl-ThPP
U00032	Cytochrome c oxidized	cpd00109	Ferricytochrome c
C00125	Ferricytochrome c	cpd00109	Ferricytochrome c
C00126	Ferrocytochrome c	cpd00110	Ferrocytochrome c
C00524	Cytochrome c	cpd00110	Ferrocytochrome c
C00080	H+	cpd00067	H+
U00013	H+-pumped	cpd00067	H+
U00028	mRNA	cpd11462	mRNA
U00019	mRNA	cpd11462	mRNA
U00034	mRNA	cpd11462	mRNA
C04501	N-Acetyl-alpha-D-glucosamine 1-phosphate	cpd02611	N-Acetyl-D-glucosamine 1-phosphate
C04256	N-Acetyl-D-glucosamine 1-phosphate	cpd02611	N-Acetyl-D-glucosamine 1-phosphate
C15812	C15812 ([Enzyme]-S-sulfanylcysteine; Thiamine biosynthesis intermediate 3)	cpd14548	Thiamine biosynthesis intermediate 3
U00052	protein-S-sulfanylcysteine	cpd14548	Thiamine biosynthesis intermediate 3
Plata model			
dolp	Dolichol_phosphate_C15H27O4P	cpd11619	Dolichyl phosphate
dolp_L	Dolichol phosphate	cpd11619	Dolichyl phosphate
dolmanp	Dolichyl_phosphate_D_mannose_C21H38O9P	cpd12407	Dolichyl phosphate D-mannose
dolmanp_L	Dolichyl phosphate D-mannose	cpd12407	Dolichyl phosphate D-mannose
hcys_L	L_Homocysteine_C4H9NO2S	cpd00135	L-Homocysteine
hcys_l	Homocysteine	cpd00135	L-Homocysteine
pail	phosphatidylinositol	cpd11822	Phosphatidyl-1D-myo-inositol
ptd1ino	phosphatidyl_1D_myo_inositol_C4140H7644O1300P100	cpd11822	Phosphatidyl-1D-myo-inositol

56

Most of the species IDs in the Plata model were converted to SEED format, with only 1.5% of the IDs being assigned modified MNXref IDs. On the other hand, two (0.7%) and 50 (3.1%) of the species IDs in the Forth and Huthmacher models, respectively, were not converted to either SEED or revised MNXref format (Figure 2.10). Close to 60% of reaction IDs in the three models were converted to either SEED or revised MNXref format, while the rest of those that were not converted were transport or boundary reactions. There was difficulty in converting some reactions, especially in the absence of additional information (e.g., reaction name, gene and enzyme data) as in the case of the Huthmacher model. In some cases, the corresponding EC number was used as the reaction ID as in the Plata model (e.g., R_1_7_1_1_mt, R_1_7_1_3_mt to represent nitrate reductase:NADH and nitrate reductase:NADPH reactions, respectively), which were manually checked against the KEGG database (to retrieve the KEGG ID) before converting to their corresponding SEED IDs.
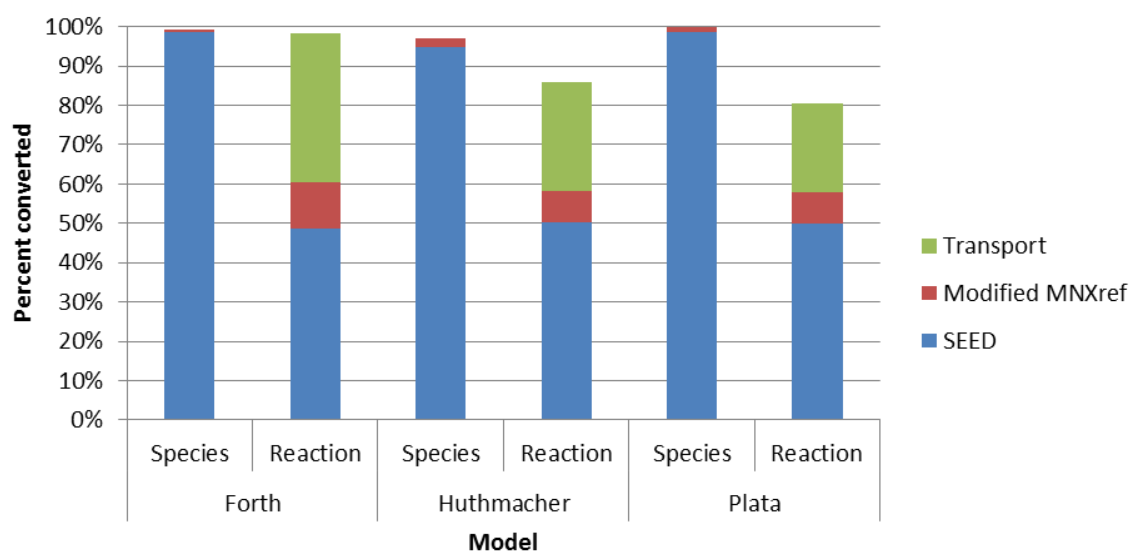
Figure 2.10 Conversion of species and reaction IDs to SEED/modified MNXref ID format
The percentages of species and reaction IDs in the original models that have been converted to either SEED or modified MNXref IDs are presented here. A large majority of the species in the three models were successfully converted to SEED or modified MNXref ID format, while only close to 60% of the reaction IDs were converted. Most of the transport reaction IDs were not converted to the SEED format.

2.3.3   Comparing the three models

After standardising the species names and IDs, duplicate species and reactions were removed giving the total number of unique species and reactions shown in Table 2.6. Conversion of the species and reaction IDs to SEED format made it possible to identify common and unique species and reactions between the three models. Since there were reaction IDs that were not converted to SEED format, comparison of reactions required inspecting the reaction equations

(with species IDs in SEED format) rather than just the reaction IDs. Also, it was important to take into consideration the presence or absence of proton/s placed on either side of the reaction to balance the equation (Bernard *et al*, 2014). Take for example glucose phosphorylation by hexokinase as represented in the three models. Although neither of the Forth and the Huthmacher models included chemical formulas for their metabolites, it can be assumed based on the reaction equation of glucose phosphorylation in these two models that the ATP, glucose 6-phosphate and ADP participating in the said reaction are protonated (Equation 2.5). The reaction equation and chemical formulas shown in Equation 2.5 were obtained from KEGG (Kanehisa *et al*, 2016). This is in contrast with the same reaction represented in the Plata model where unprotonated forms of ATP, glucose 6-phosphate and ADP were included in the reaction (Equation 2.6), consistent with the reactions and chemical formulas available in BioCyc (Caspi *et al*, 2014). Marked in blue in the equations below are the hydrogen atoms in the chemical formula that include those that protonate the phosphates in the corresponding metabolites, while marked in red are the hydrogen atoms in the chemical formulas of the unprotonated form.

Table 2.6   Total number of species and reactions before and after species standardisation

	Before species standardisation		After species standardisation	
Model	Species	Reactions	Species	Reactions
Forth	267	247	263	244
Huthmacher	1673	1376	1609	1374
Plata	1025	1001	914	969

$$C_6H_{12}O_6 \ + \ C_{10}H_{16}N_5O_{13}P_3 \ \rightarrow \ C_6H_{13}O_9P \ + \ C_{10}H_{15}N_5O_{10}P_2$$

Glucose   ATP   Glucose 6-phosphate   ADP   **Equation 2.5**

$$C_6H_{12}O_6 \ + \ C_{10}H_{12}N_5O_{13}P_3 \ \rightarrow \ C_6H_{11}O_9P \ + \ C_{10}H_{12}N_5O_{10}P_2 \ + \ H$$

Glucose   ATP   Glucose 6-phosphate   ADP H^+ **Equation 2.6**
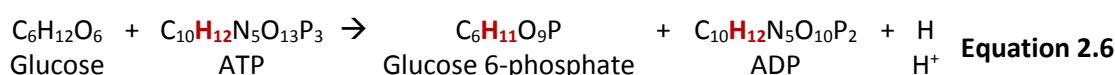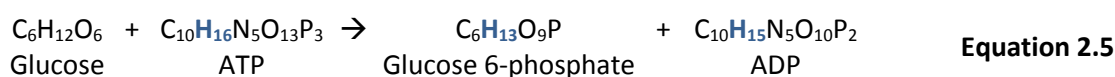
There are 198 species that are in common between the three models (Figure 2.11). Nucleotides, nucleosides, bases and intermediate metabolites involved in nucleotide metabolism comprised 23.2% of these common species, while 22.2% of the species were amino acids and their derivatives. Small molecules such as water, oxygen and carbon dioxide accounted for 13.1% of common species, while cofactors such as coenzyme A, nicotinamides and folates comprised 12.1%. Unfortunately, comparison of chemical formulas could not be done as these were not included in both the Forth and Huthmacher models. There were only 99 reactions in common between the three models. Transferases and oxidoreductases were the most common enzyme classifications among the reactions common to all three models following a similar pattern as in other models such as the yeast (Förster *et al*, 2003) and the *Leishmania major*

(Chavali *et al*, 2008). Most of the common species and reactions were located in the cytosol. More than 60% of the common reactions were involved in nucleotide and carbohydrate metabolism, representing core reactions such as purine and pyrimidine metabolism, as well as glycolysis and the pentose phosphate pathway.

Of the 99 common reactions, 26 (26.3%) were those that were represented with different stoichiometric equations in the original models as a result of utilising protonated or unprotonated metabolites. Between the Forth and the Huthmacher model, only two common reactions had dissimilar equations. This was expected as most of the reactions in the two models were obtained from the KEGG database. The difference in the equations of these two identified reactions were because of manual reaction equation balancing done in the Forth model (Forth, 2012). On the other hand, 28 common reactions between the Plata and the Forth models, and 98 common reactions between the Plata and the Huthmacher models had dissimilar reaction equations due to the use of different metabolite chemical formulas. Again, this was expected as the Plata model utilised data not just from KEGG but also from PlasmoCyc (Plata *et al*, 2010). Given the considerable number of common reactions with dissimilar reaction equations, it was important to take this into consideration when comparing reactions. Otherwise, redundant reactions would have been incorporated into the combined model.
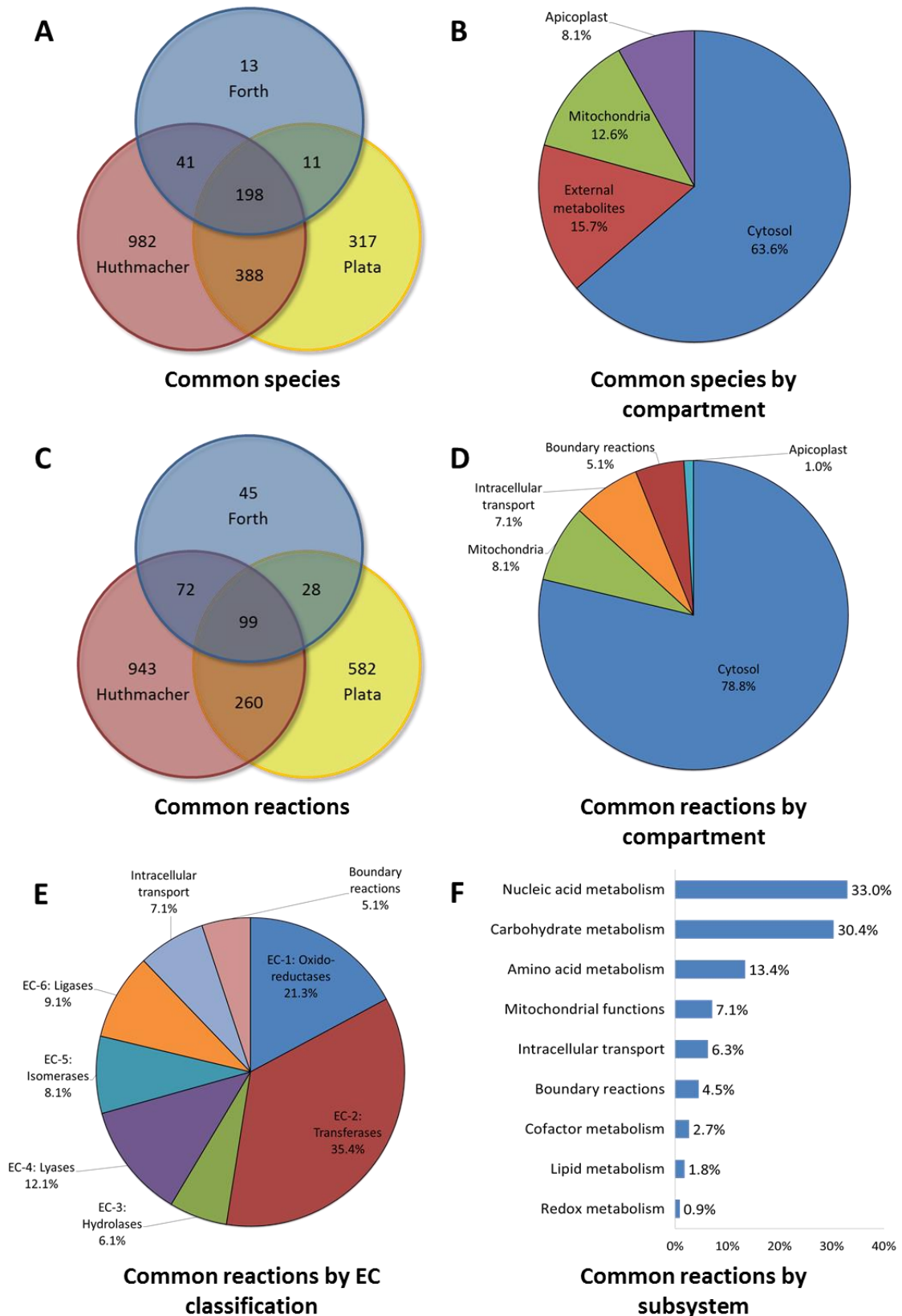
Figure 2.11 Common species and reactions between the three models
(A) A Venn diagram of common species between the three models; (B) Distribution of common species by cellular compartment; (C) Venn diagram of common reactions between the three models, and common reactions grouped by (D) compartment (including intracellular transport reactions, and boundary reactions that transport external metabolites), (E) enzyme commission (EC) classification and (F) subsystem involvement.

2.3.4   Identifying unique reactions from source models

As enumerated in the methodology section, three criteria were used to identify reactions from the source models (i.e., the Huthmacher and Plata models) to be added to the minimal model (i.e., the Forth model). In order to satisfy the first criterion, the reaction from the source model should have gene association and enzyme data. Available data on a total of 5,777 *P. falciparum* 3D7 genes were downloaded from PlasmoDB (Aurrecoechea *et al*, 2009). All gene data included updated gene IDs, gene names and lists of old gene IDs, while 1,321 of these genes that are associated with metabolic reactions had their corresponding EC numbers. In the case of the reactions in the Plata model, the gene IDs (in the old format) were compared against PlasmoDB data. A total of 616 reactions (61.5%) had gene association data that matched PlasmoDB data, while 41 reactions (4.1%) had gene IDs not in PlasmoDB. The rest of the reactions were boundary and transport reactions that did not have gene association data.

Since the Huthmacher model did not have any gene or enzyme data, reaction IDs in KEGG format were looked up in the KEGG database (Kanehisa *et al*, 2016) to obtain the corresponding EC number. For reaction IDs that were not in KEGG format, the reaction equation was used to obtain the corresponding EC number. There were 736 reactions that had associated EC numbers from KEGG. The EC numbers of these reactions were compared against those in the PlasmoDB database, yielding a total of 517 reactions with gene association data. These 517 Huthmacher reactions and 616 Plata reactions were then evaluated on the basis of the second and the third criteria, in an iterative process as shown in simple example in Figure 2.3 (Methodology section). At the end of the iterations, 201 reactions involving 199 species and 162 reactions involving 245 species were collected from the Huthmacher and Plata models, respectively, giving an expanded model with 607 reactions and 707 species.

The use of the three criteria in selecting reactions to be added to the minimal model has an important drawback. Since a majority of intracellular transport reactions from the source models did not have any gene association data, these were not selected for inclusion in the minimal model. As a result, intracellular transport reactions (in the source models) that link reactions in existing compartments in the minimal model to other compartments that are solely present in the source models, specifically in the Huthmacher model (i.e., nucleus, endoplasmic reticulum) were not added. This in turn limited the reactions into the original compartments in the minimal model and the expanded model retained the same compartments that were present in the minimal model. It is important to note that a majority of the nuclear reactions (involved in RNA synthesis) were not present in KEGG or MPMP and these were not included in the final model. On the other hand, many of the reactions in the endoplasmic reticulum such as

those involved in dolichol and aminosugar metabolism were also present in the Plata model, which were eventually added to the expanded model (as cytosolic reactions).

2.3.5   Correcting reaction direction and dead-ends

Out of the 1,321 *P. falciparum* 3D7 genes associated with metabolic reactions that are available from PlasmoDB (Aurrecoechea *et al*, 2009), 428 had metabolic pathway data, which included the gene name, product description (e.g., enzyme name) and a table of associated reactions as shown in the example in Figure 2.12. In this example, reactions that are associated with the gene PF3D7_0624000 (hexokinase) were enumerated in table form. The table included the (1) reaction ID, (2) one representative substrate that participates in the reaction, (3) EC number, (4) one representative product that participates in the reaction, (5) pathway name and ID, and (6) direction. Note that the reaction, species and pathway IDs are in KEGG format.

Gene ID: PF3D7_0624000
Gene Name or Symbol: HK
Product Description: hexokinase (HK)

Table: Metabolic Pathway Reactions

Reaction	Substrates	EC number	Products	Pathway name	Direction
R01786	C00267	2.7.1.1	C00668	Glycolysis / Gluconeogenesis (ec00010)	Reversible
R01786	C00267	2.7.1.1	C00668	Glycolysis / Gluconeogenesis (ec00010)	Irreversible
R01600	C00221	2.7.1.1	C01172	Glycolysis / Gluconeogenesis (ec00010)	Irreversible
R01786	C00267	2.7.1.1	C00668	Glycolysis / Gluconeogenesis (ec00010)	Reversible
R01786	C00267	2.7.1.1	C00668	Glycolysis / Gluconeogenesis (ec00010)	Irreversible
R01786	C00267	2.7.1.1	C00668	Galactose metabolism (ec00052)	Irreversible
R01786	C00267	2.7.1.1	C00668	Galactose metabolism (ec00052)	Reversible
R01786	C00267	2.7.1.1	C00668	Galactose metabolism (ec00052)	Reversible
R01786	C00267	2.7.1.1	C00668	Galactose metabolism (ec00052)	Irreversible
R01326	C00159	2.7.1.1	C00275	Fructose and mannose metabolism (ec00051)	Irreversible
R00867	C00095	2.7.1.1	C05345	Fructose and mannose metabolism (ec00051)	Irreversible
R01326	C00159	2.7.1.1	C00275	Fructose and mannose metabolism (ec00051)	Irreversible

Figure 2.12   An example of metabolic pathway data from PlasmoDB
Data on metabolic pathway related to *P. falciparum* 3D7 genes were downloaded from PlasmoDB. Genes with available data are provided with the information shown in this figure such as EC numbers, reaction reversibility and pathway involvement.

Model reactions with gene ID and EC numbers were then compared against the PlasmoDB metabolic pathway data. In some cases, the comparison was straightforward especially when a given PlasmoDB gene is only associated with a single reaction. However, since a majority of the genes with metabolic pathway data from PlasmoDB had two or more associated reactions (268 out of 428, 62.6%), a significant number of reactions in the model had to be evaluated manually. These included model reactions that matched with multiple reaction directions as well as those with conflicting EC number information. Reaction reversibility was amended for 60 (9.8%) of the model reactions (i.e., change from reversible to irreversible or vice versa), while nine (1.5%)

irreversible reactions required reversal of the reactants and products. Among those that required manual evaluation and correction, 39 reactions (6.4%) had multiple associated reactions while 154 (25.4%) had conflicting EC numbers. The rest of the reactions did not require any correction because either they do not have gene/reaction data or the reaction attributes matched those in PlasmoDB.

Using Cytoscape, 241 dead-end metabolites, involving 171 reactions were identified in the expanded model. Gaps were filled using data from KEGG (Kanehisa *et al*, 2016), BRENDA (Placzek *et al*, 2017), PlasmoDB (Aurrecoechea *et al*, 2009) and MPMP (Ginsburg & Abdel-Haleem, 2016). Reactions involving metabolites that are not present in MPMP were removed from the model. After gap filling, the model was re-evaluated for dead-end metabolites, revealing 55 dead-end metabolites. Gap filling was again done using data from the databases mentioned above. For dead metabolites that also appeared as dead-ends in MPMP, exchange reactions were added to bring the metabolite in/out of a given compartment. A total of 11 boundary reactions were included in the model to remove dead-end metabolites, while seven intracellular transport reactions were added to move dead-end metabolites from one compartment to another. Interestingly, cytidine monophosphate (CMP), both in the cytosol and the apicoplast, was identified as a dead-end metabolite even though it was being produced and consumed by different reactions. After careful assessment, it was identified that CMP was being produced by multiple irreversible reactions in the cytosol and by a single reaction in the apicoplast. Furthermore, the only reaction that consumes CMP was the reversible intracellular transport reaction that shuttles CMP between the cytosol and the apicoplast. These reactions producing CMP as well as the intracellular transport reaction that bring CMP between the cytoplasm and the apicoplast are shown in red in Figure 2.13. Thus, CMP was considered as a dead-end metabolite. In order to correct this, an additional reaction, CMP phosphohydrolase (rxn00363_c, EC number 3.1.3.5), which hydrolyses CMP to cytidine and orthophosphate, was added to the model. Although the *Plasmodium falciparum* 3D7 has a gene that is associated with a nucleotide phosphorylase (PF3D7_1206100, EC number 3.1.3.5), the enzyme is specific to IMP as a substrate (Aurrecoechea *et al*, 2009). Therefore CMP phosphohydrolase was added into the model without any gene association data. An additional boundary transport reaction was added to the model to transport cytidine via the PfENT1 nucleoside transporter (PF3D7_1347200). Reactions added to address the issue of CMP as a dead-end metabolite are shown in yellow in Figure 2.13. Table 2.7 shows the list of reactions that were added to address dead-end metabolites in the model.
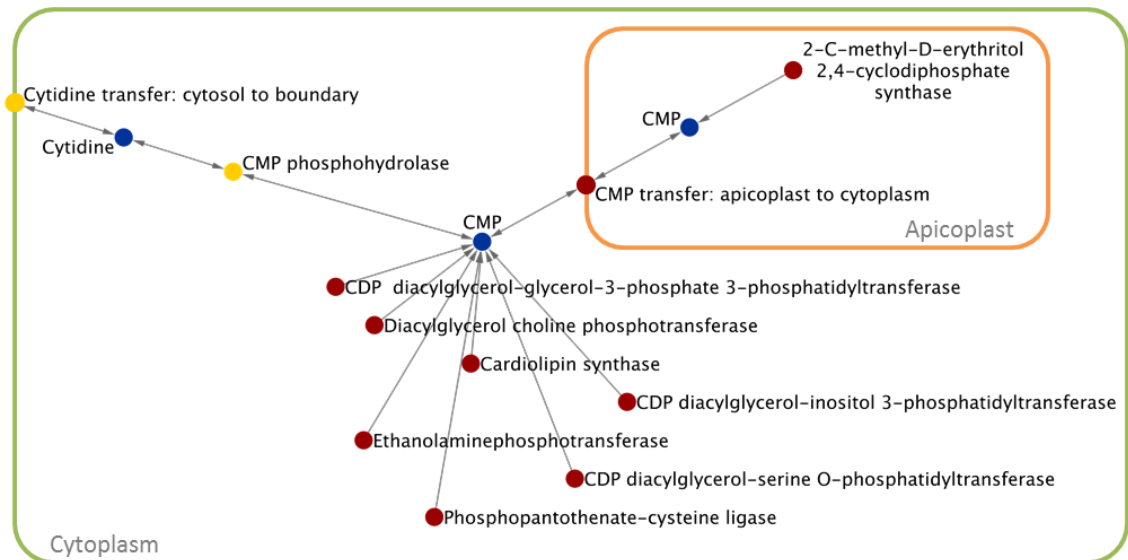
Figure 2.13 Diagram showing CMP as a dead-end metabolite
CMP in the model is produced by a number of irreversible reactions (in red) in the cytoplasm and apicoplast, while it is only consumed by a single reaction, which is the transport reaction that brings CMP between the cytosol and the apicoplast. This makes CMP a dead-end metabolite. This was addressed by the addition of the two reactions (in yellow).

Table 2.7   Reactions added into the model to correct dead-end metabolites after gap filling

Reaction ID	Reaction name	Reversibility
Biliverdin_ctob_s	Biliverdin transfer: cytosol to boundary	False
Cadaverine_ctob_s	Cadaverine transfer: cytosol to boundary	False
Cardiolipin_ctob_s	Cardiolipin transfer: cytosol to boundary	False
CO_ctob_s	CO transfer: cytosol to boundary	False
CTP_ctoa_s	CTP transfer: cytosol to apicoplast	False
Cytidine_ctob_s	Cytidine transfer: cytosol to boundary	True
Diphosphate_atoc_s	Diphosphate transfer: apicoplast to cytosol	True
Diphosphate_mtoc_s	Diphosphate transfer: mitochondria to cytosol	True
Fe2_ctom_s	Fe2+ transfer: cytosol to mitochondria	True
Glucosylceramide_ctob_s	Glucosylceramide transfer: cytosol to boundary	False
Glyceraldehyde_ctob_s	Glyceraldehyde transfer: cytosol to boundary	False
NH3_atoc_s	NH3 transfer: apicoplast to cytosol	False
NH3_mtoc_s	NH3 transfer: mitochondria to cytosol	True
Oxygen_mtoc_s	Oxygen transfer: mitochondria to cytosol	True
Phenylpyruvate_ctob_s	Phenylpyruvate transfer: cytosol to boundary	False
Propane_1_2_diol_ctob_s	Propane_1_2_diol transfer: cytosol to boundary	False
rxn00363_c	CMP phosphohydrolase	True
Saccharopine_ctob_s	Saccharopine transfer: cytosol to boundary	False
Spermine_ctob_s	Spermine transfer: cytosol to boundary	False

Reactions in the resulting model were counter-checked against the reactions in a number of pathways in MPMP (Ginsburg & Abdel-Haleem, 2016). A total of 35 pathways (Carbohydrates: 5; Nucleic acids: 2; Amino acids: 9; Lipids: 6; Co-factors: 10; TCA cycle; Apicoplast transport reactions: 1; Mitochondrial transport reactions: 1; and Plasma membrane transport reactions: 1) were reviewed. In the process, reaction reversibility was corrected as needed, gene IDs were

updated, duplicate reactions (usually those not in their proper compartments) were deleted, necessary transport reactions were added and pathway information were added as additional attributes for each reaction.

2.3.6   Deriving the biomass equation

Previous experimental data on the proportions of protein, DNA and RNA in the *P. falciparum* 3D7 late schizont biomass were obtained from the work of Thomas Forth (Forth, 2012), which were 48.0%, 6.7% and 5.9%, respectively. As in the original Forth model, the remaining percentage was then broken down based on a 27:15 carbohydrate to lipid ratio (Chavali *et al*, 2008), giving calculated proportions of 25.3% and 14.1%, respectively. These macromolecules (i.e., protein, DNA, RNA, carbohydrates and lipids) were further subdivided into individual subcomponents (e.g., amino acids for protein, nucleotides for DNA/RNA). The percentages of these individual subcomponents were based on published data on *P. falciparum* (Chanda *et al*, 2005; Gardner *et al*, 2002; Sanz *et al*, 2013; Botté *et al*, 2013). Using these published percentages, the weighted average molecular mass (MW in Equation 1.1) of the subcomponents representing the macromolecular component was also calculated. Table 2.8 shows the calculated weighted average molecular mass of the subcomponents of the macromolecules used to calculate the stoichiometric coefficients of the reactants and products in the biomass reaction equation. Figure 2.14 shows the percentages of the macromolecules and their subcomponents used in the derivation of the biomass equation. Table 2.9 shows the list of reactants and products in the biomass reaction. Note that negative stoichiometric coefficients were given to reactants since they are being consumed in the reaction, while products were given positive coefficients.

Table 2.8   Calculated weighted average molecular mass of the subcomponents representing the macromolecular components in the biomass reaction

Macromolecule	Subcomponents	MW (gram/mol)
Protein	Amino acids	117.48
DNA	Deoxynucleotides	486.77
RNA	Nucleotides	497.36
Carbohydrates	Nucleotide sugars	602.35
Lipids	Fatty acids	276.79
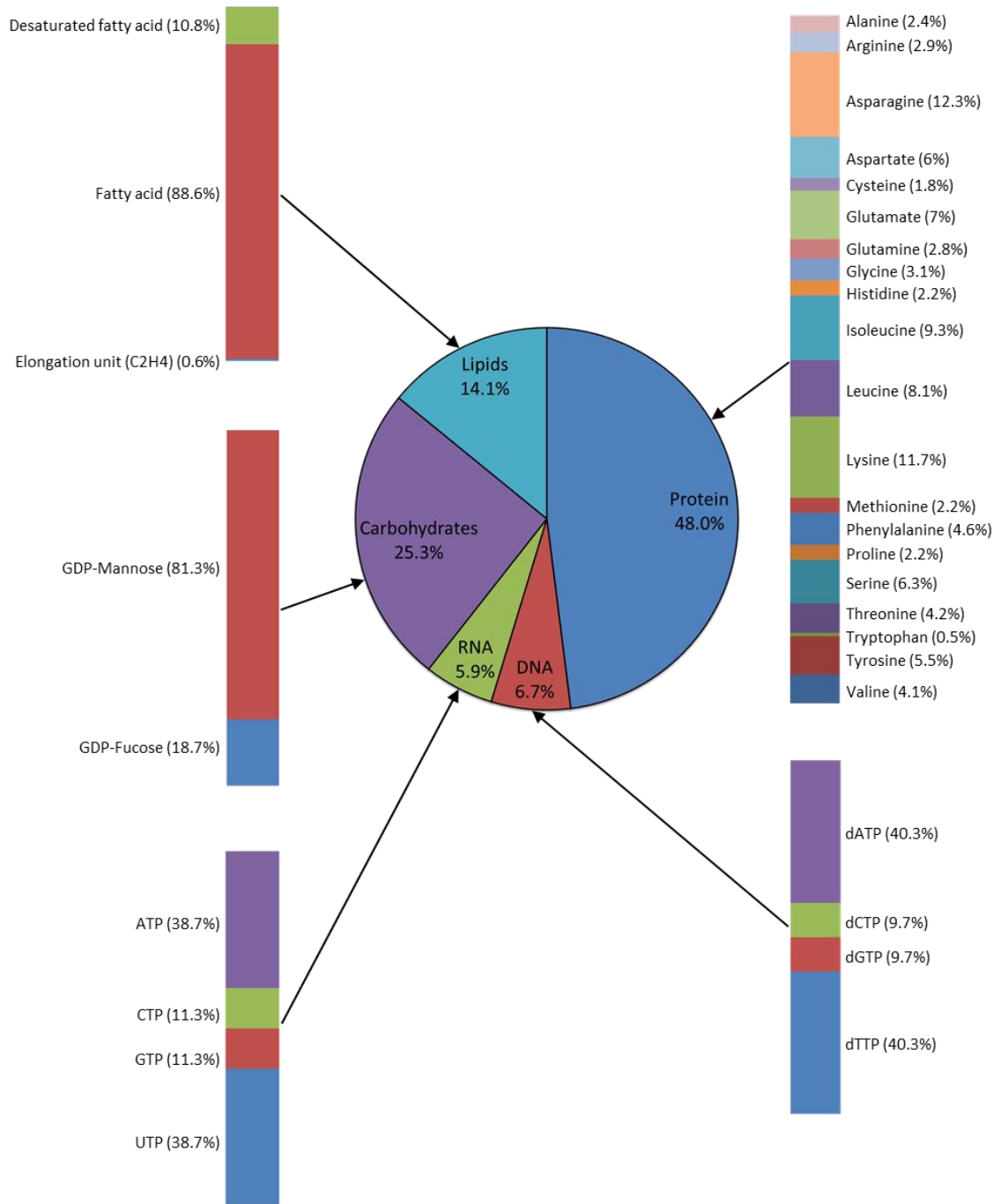
Figure 2.14 Percentages of the molecular components in the *P. falciparum* biomass
The proportions of macromolecules (centre pie chart) were taken from experimentally measured values by Forth (2012) for DNA, RNA and proteins while the rest were divided between carbohydrates and fatty acids based on data from *L. major* (Chavali *et al*, 2008). Subcomponents of the macromolecules (shown in stacked columns) were obtained from published data on *P. falciparum* (Chanda *et al*, 2005; Gardner *et al*, 2002; Sanz *et al*, 2013; Botté *et al*, 2013).

Table 2.9 **Biomass reaction metabolites with their corresponding species ID, chemical formula and stoichiometric coefficient**

Name	ID	Formula	Coefficient
L-Alanine	cpd00035_c	$C_3H_7NO_2$	-0.099
L-Arginine	cpd00051_c	$C_6H_{14}N_4O_2$	-0.12
L-Asparagine	cpd00132_c	$C_4H_8N_2O_3$	-0.503
L-Aspartate	cpd00041_c	$C_4H_7NO_4$	-0.248
L-Cysteine	cpd00084_c	$C_3H_7NO_2S$	-0.074
L-Glutamate	cpd00023_c	$C_5H_9NO_4$	-0.289
L-Glutamine	cpd00053_c	$C_5H_{10}N_2O_3$	-0.116
Glycine	cpd00033_c	$C_2H_5NO_2$	-0.128
L-Histidine	cpd00119_c	$C_6H_9N_3O_2$	-0.091
L-Isoleucine	cpd00322_c	$C_6H_{13}NO_2$	-0.384
L-Leucine	cpd00107_c	$C_6H_{13}NO_2$	-0.334
L-Lysine	cpd00039_c	$C_6H_{14}N_2O_2$	-0.479
L-Methionine	cpd00060_c	$C_5H_{11}NO_2S$	-0.091
L-Phenylalanine	cpd00066_c	$C_9H_{11}NO_2$	-0.19
L-Proline	cpd00129_c	$C_5H_9NO_2$	-0.091
L-Serine	cpd00054_c	$C_3H_7NO_3$	-0.26
L-Threonine	cpd00161_c	$C_4H_9NO_3$	-0.173
L-Tryptophan	cpd00065_c	$C_{11}H_{12}N_2O_2$	-0.021
L-Tyrosine	cpd00069_c	$C_9H_{11}NO_3$	-0.227
L-Valine	cpd00156_c	$C_5H_{11}NO_2$	-0.169
dATP	cpd00115_c	$C_{10}H_{16}N_5O_{12}P_3$	-0.055
dCTP	cpd00356_c	$C_9H_{16}N_3O_{13}P_3$	-0.013
dGTP	cpd00241_c	$C_{10}H_{16}N_5O_{13}P_3$	-0.013
dTTP	cpd00357_c	$C_{10}H_{17}N_2O_{14}P_3$	-0.055
ATP	cpd00002_c	$C_{10}H_{16}N_5O_{13}P_3$	-4.178
CTP	cpd00052_c	$C_9H_{16}N_3O_{14}P_3$	-0.013
GTP	cpd00038_c	$C_{10}H_{16}N_5O_{14}P_3$	-8.198
UTP	cpd00062_c	$C_9H_{15}N_2O_{15}P_3$	-0.046
Elongation unit	FAC2H4unit_c	C_2H_4	-0.003
Fatty acid	cpd00049_c	CHO_2R	-0.451
Desaturated fatty acid	DesatFA_c	CO_2R	-0.055
GDP-L-fucose	cpd00272_c	$C_{16}H_{25}N_5O_{15}P_2$	-0.078
GDP-mannose	cpd00083_c	$C_{16}H_{25}N_5O_{16}P_2$	-0.342
H2O	cpd00001_c	H_2O	-12.677
Malaria biomass	biomass_c		1
GDP	cpd00031_c	$C_{10}H_{15}N_5O_{11}P_2$	8.591
Orthophosphate	cpd00009_c	O_4P	12.257
AMP	cpd00018_c	$C_{10}H_{14}N_5O_7P$	4.086
H+	cpd00067_c	H	36.771

2.3.7   Final model

The name of the final model (iFT342) was based on the naming convention where the "i" indicates that it is an *in silico* model, followed by the first author's initials and the total number of genes represented in the model (Reed *et al*, 2003). The iFT342 has a total of 342 genes, 551

reactions and 560 metabolites. The model includes five compartments: the apicoplast, cytosol, mitochondria, vacuole and the external compartment. There are 106 boundary reactions that transport metabolites to and from the extracellular space, and 106 intracellular transport reactions that bring metabolites into the different compartments. Compared with the minimal model, iTF143 (Forth, 2012), the addition of reactions resulted in a notable increase in the total number of genes, reactions and metabolites in the final model as shown in Table 2.10.

Table 2.10 Comparison of metabolic model characteristics between the minimal model (iTF143) and the final model (iFT342)

	iTF143	iFT342
Genes	143	342
Reactions	247	551
Gene associated (intracellular)	141	324
Gene associated (transport)	0	15
Gene associated (exchange)	0	32
Non-gene associated (intracellular)	10	14
Non-gene associated (transport)	52	91
Non-gene associated (exchange)		
Boundary reactions	43	74
Biomass reaction	1	1
Metabolites	267	560
Pathways	19	40
Compartments	5	5
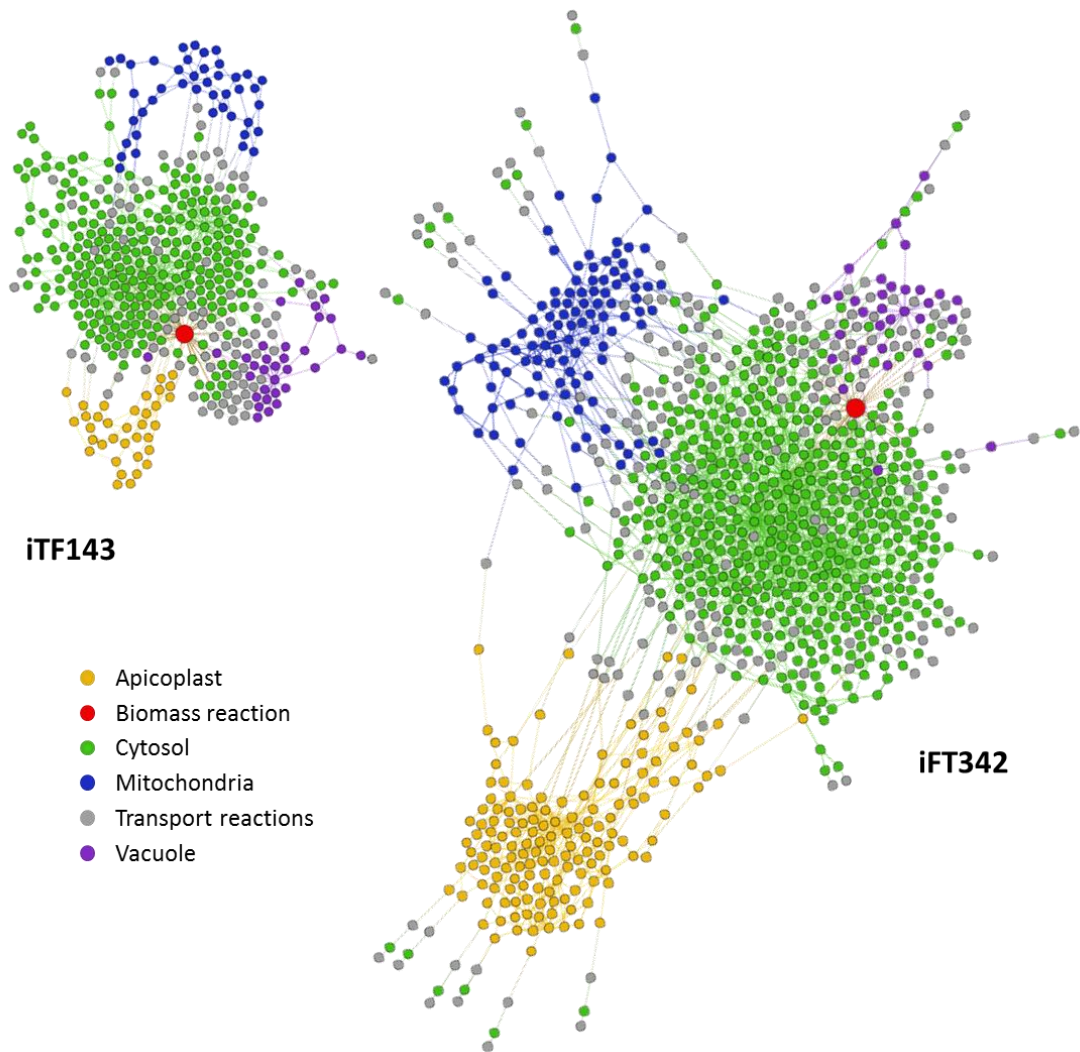
Figure 2.15 Visual representation of the minimal model (iTF143) and the final model (iFT342)
This figure provides a visual comparison between the minimal model and the final model. Shown in various colours are the species and reactions in accordance with their compartmental location. Transport reactions (which include boundary and intracellular transport reactions) are shown in grey while the objective function (i.e., the biomass reaction) is shown in red. Model networks were visualised using Gephi open source software (Bastian *et al*, 2009).

There are 106 boundary metabolites and 454 intracellular metabolites. All 560 metabolites in the model have chemical formulas, and 530 have additional metabolite attributes, i.e., PubChem ID (Bolton *et al*, 2008), InChI keys (Heller & McNaught, 2009) and Canonical SMILES (McNamara & Stearne, 2010). These additional attributes allow accurate identification of metabolites, and will contribute to the ease in comparing with other models. These characteristics are absent in the previous models. With the addition of chemical formulas, reaction stoichiometric balance was checked and corrected. All internal reactions in the model are mass balanced.

About 70% of the reactions in the model are gene associated. Most of the non-gene associated reactions are transport reactions, while only 2.5% are intracellular reactions. Most of the non-gene associated intracellular reactions were added to complete important pathways (e.g., shikimate biosynthesis, ubiquinone metabolism, inositol phosphate metabolism). Cytosolic reactions and transport reactions account for a vast majority of the reactions in the model, while reactions in the vacuole, which mostly involve haemoglobin metabolism, account for less than 1% of all the reactions. Similar to the Plata model (Plata *et al*, 2010), as well as the yeast (Förster *et al*, 2003) and the *Leishmania major* (Chavali *et al*, 2008) models, transferases are the most common enzyme class in the iFT342. These were followed by oxidoreductases, similar to the yeast and *L. major* models, but not the Plata model, where the second most common are hydrolases. Over 26% of the reactions in the model are involved in lipid metabolism, of which, more than 50% participate in fatty acid synthesis and utilisation of phospholipids. Close to 10% of the reactions in the model are involved in carbohydrate metabolism while 13% are involved in amino acid metabolism and haemoglobin digestion. Figure 2.16 shows the distribution of reactions based on the gene association, compartment, EC classification and subsystem involvement.
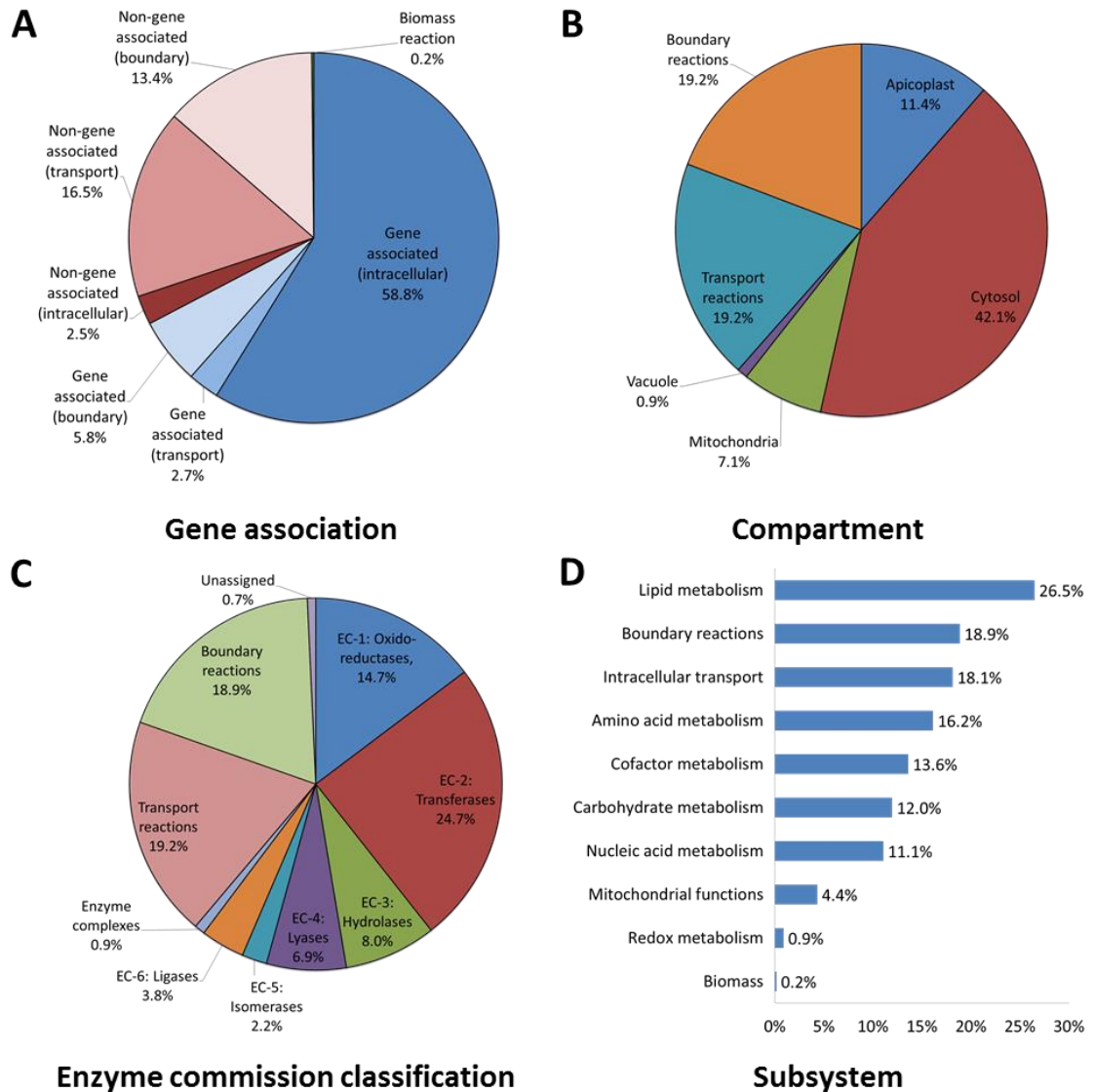
Figure 2.16 Model characteristics of iFT342
Reactions in the iFT342 model are shown here grouped by (A) gene association, (B) compartment, (C) enzyme commission classification and (D) subsystem involvement.

Although the final model size is not close to those of previously published malaria models, the number of dead reactions was minimised allowing more live reactions that can participate in model simulations. This gave our model a total number of live reactions that is comparable with the other malaria models. As shown in Table 2.11, 87.5% (482 out of the 551) of the reactions in the final model are live reactions. This percentage of live reactions is close to that in the minimal model, and is higher than in many of the highly curated metabolic models. The low percentage of dead reactions can be attributed to the absence of dead-end metabolites in the model, as these metabolites can block the flux in pathways that they are involved (Latendresse *et al*, 2012). The number of metabolites, reactions, dead-end metabolites and live reactions in the model and other published metabolic models are summarised in Table 2.11.

Table 2.11   Number and percentage of dead-end metabolites and live reactions in a number of highly-curated GSMs determined using MetExplore presented in decreasing percentages of live reactions

Model name	Organism	Metabolites	Reactions	Dead-end metabolites		Live reactions		Reference
iFT342	*P. falciparum*	560	551	0	0.0%	482	87.5%	
iTF143	*P. falciparum*	267	247	13	4.9%	211	85.4%	(Forth, 2012)
Recon2	*H. sapiens*	5063	7440	1178	23.3%	5317	71.5%	(Thiele *et al*, 2013)
iAF1260	*E. coli*	1972	2382	118	6.0%	1532	64.3%	(Feist *et al*, 2010)
iAC560	*L. major*	1165	1112	259	22.2%	714	64.2%	(Chavali *et al*, 2008)
iMM904	*S. cerevisiae*	1392	1577	198	14.2%	885	56.1%	(Herrgård *et al*, 2008)
PlasmoNet	*P. falciparum*	1623	1376	600	37.0%	534	38.8%	(Huthmacher *et al*, 2010)
iTH366	*P. falciparum*	915	1001	357	39.0%	354	35.4%	(Plata *et al*, 2010)

2.4 Discussion

At the beginning, there was difficulty reconciling the three models due to the difference in ID formats utilised. A similar issue was encountered during the development of a consensus model of *Saccharomyces cerevisiae* from two previously and independently reconstructed models (Herrgård *et al*, 2008). It was necessary for the team to initially annotate the metabolites in the two models with more specific identifiers to make model comparison possible. They compared the metabolites against data from ChEBI, KEGG and PubChem to collect InChI keys and SMILES identifiers, which were used for a more accurate comparison. Acknowledging this problem in reconciling different identifiers, there have been attempts to map different metabolite and reaction IDs of different ontological formats to aid in genome scale metabolic reconstruction (Lang *et al*, 2011; Kumar *et al*, 2012; Bernard *et al*, 2014). MNXref is one such attempt to unify different ontological formats by comparing the calculated structure of metabolites at pH 7.3. For metabolites without a given structure, exact compound names were used. As for reactions, they were compared based on the participating metabolites, but not taking into consideration the stoichiometric coefficient and the presence of water and proton (Bernard *et al*, 2014).

In the comparison of the three malaria models used in this project, MNXref data was used to reconcile metabolite names and IDs; however this was not utilised for the comparison of reactions. A similar approach to the generation of MNXref unified data was applied to the comparison of reactions in the three models to ensure accuracy. Comparison of reactions was based on the metabolites participating in the reactions. Nevertheless, reaction IDs (in SEED format) were assigned to the appropriate reaction using MNXref data. To avoid future issues regarding the exact and accurate identification of model species, it has been recommended that additional identifiers be included in describing metabolites (Ravikrishnan & Raman, 2015; Herrgård *et al*, 2008). Thus it was ensured that the final model had properly annotated reactions (i.e., updated gene IDs and EC numbers) and species (i.e., chemical formula, SMILES, InChI keys and PubChem IDs) to facilitate ease when comparing with other models.

To ensure model accuracy, reactions from the source models were compared against several highly curated and reliable databases. The PlasmoDB database was highly utilised in evaluating and correcting the annotations in the model. The database receives regular updates as well as inputs through constant communication with other databases such as GeneDB (Aurrecoechea *et al*, 2009; Logan-Klumpler *et al*, 2012). Model annotations were also highly dependent on the highly and manually curated MPMP database (Ginsburg & Abdel-Haleem, 2016; Ginsburg, 2006) which was also utilised in the development of the other malaria metabolic

models (Forth, 2012; Huthmacher *et al*, 2010; Plata *et al*, 2010). Use of accurate data is crucial to guarantee that the gene-protein-reaction relationships represented in the reaction annotations are correct. This in turn will help ensure that the data obtained from model simulations (e.g., gene knockouts) will be as accurate as possible.

Although initial model development was done in an automated fashion, manual analysis and annotations remained as the main bulk of the work to ensure model accuracy. Incorporating standardised molecular formulas for metabolites, although not mandatory in SBML level 3 version 1 (Chaouiya *et al*, 2015), was instrumental in making sure that the reactions do not violate the law of conservation of mass. This metabolite attribute should be made mandatory for metabolic models. A recent appraisal of 59 genome scale metabolic models revealed that many of these models did not include molecular formulas in the metabolic attributes, making it impossible to determine reaction mass balance (Ravikrishnan & Raman, 2015). All internal reactions in the final model were mass balanced making the model well suitable for subsequent analysis and hypotheses generation.

Model quality is sometimes assessed in terms of its completeness and comprehensiveness of gene coverage. However, the total number of genes or reactions represented in the model does not normally correspond to high model quality (Monk *et al*, 2014). The same study mentioned above reported that only 7% of the evaluated models had live reactions greater that 80% (Ravikrishnan & Raman, 2015). In the generation of the final model in this study, extensive manual curation was also necessary to correct dead-end metabolites which contributed to the generation of a model with a high percentage of live reactions (87.5%), which is an important measure of model quality. Despite having a total number of reactions and genes less than those in the Huthmacher and the Plata model, correcting dead reactions resulted in a total number of live reactions only 52 reactions fewer compared with the Huthmacher model, and 128 reactions more compared with the Plata model. In terms of percentage of live reactions, these two models only had 38.8% and 35.4%, respectively. Finally, compared with other published and highly curated genome scale metabolic models, the iFT342 has the highest percentage of live reactions. It is therefore recommended that dead-end metabolites should be addressed extensively in light of available data to increase the number of live reactions that can participate in model simulations (Latendresse *et al*, 2012). Moreover, it is also recommended that the percentage of live reactions be used as a means of assessing model quality, in congruence with the recommendations in other studies (Ravikrishnan & Raman, 2015; Monk *et al*, 2014).

Chapter 3 *In vitro* flux measurements and glucose perturbation

3.1 Introduction

Plasmodium falciparum in vitro culture using human erythrocytes was described as early as 1912 (Bass & Johns, 1912). Parasites grown in leukocyte-free erythrocytes with human serum were able to complete up to three asexual cycles; however, most of the parasites were observed to die after the first and second cycles. A landmark in the development of *in vitro P. falciparum* culture was the discovery in 1976 by William Trager and James Jensen (Trager & Jensen, 2005) that continuous growth required low oxygen conditions. They demonstrated that parasites can be indefinitely cultured in a thin layer of human erythrocytes (plasma and leukocyte-free) in RPMI media supplemented with human serum either in a continuous flow vial (i.e., media is constantly replaced at a rate of 50 ml/day using a peristaltic pump) in 5% oxygen, or in a petri dish placed in a candle jar. The latter was maintained by replacing the media daily and keeping the culture in low oxygen. The candle technique uses a lit candle placed in a chamber (that is eventually sealed air tight when the fire goes out) to consume oxygen, providing a suitable environment for parasite culture. Using these techniques, the parasites were noted to have lost their synchronicity, unlike what is normally observed in human infection (i.e., synchronised parasite growth). Asynchronous parasite cultures show a mix of all asexual stages (i.e., rings, late trophozoites and schizonts) *in vitro*. Nevertheless, the morphology of the parasites remained the same and the parasites were able to be grown continuously for more than 50 days.

A number of issues have been associated with the use of human serum in culture, including cost, availability and serum donor-to-donor variability (Willet & Canfield, 1984; Flores *et al*, 1997; Asahi & Kanazawa, 1994). It has been shown that bovine albumin, glucose, hypoxanthine and lipid-cholesterol-rich mixture were important in supporting parasite growth in the absence of human serum (Ofulla *et al*, 1993). Albumax II (Life Technologies), a lipid-rich bovine serum albumin with low immunoglobulin content, supplemented with hypoxanthine, is a popular replacement for human serum in malaria culture media (Flores *et al*, 1997). There were a number of differences observed when growing parasites in Albumax versus human serum. Reduced parasite cytoadherence, which is associated with the sequestration of red blood cells, has been shown in parasites grown in Albumax-supplemented media (Treutiger *et al*, 1999). Also, difficulty in culturing wild-type isolates in Albumax media has been reported (Taverne, 2000). Nevertheless, no significant difference in *in vitro* parasite growth was observed between cultures of *P. falciparum* strains grown in media supplemented with human serum or Albumax II and hypoxanthine (Cranmer *et al*, 1995). Albumax II has been widely utilised

especially in a number of metabolomics studies related to drug target identification (Allman *et al*, 2016; Creek *et al*, 2016; Cobbold *et al*, 2016).

The asynchronous nature of *in vitro* parasite culture makes it difficult to obtain stage specific information without performing synchronisation techniques. A commonly utilised technique for culture synchronisation is sorbitol treatment (Lambros & Vanderberg, 1979). This technique exploits the stage-dependent increase in the permeability of the new permeability pathways in infected erythrocytes. In the early stages, RBC permeability gradually increases until about 24 hours post invasion when the permeability exponentially increases (Waldecker *et al*, 2017; Lew *et al*, 2003). This increase in permeability results in an increase in osmotic fragility of the host cell. Thus in the presence of high sorbitol concentrations (5% or 274 mM), more permeable cells draw in sorbitol resulting in increased internal osmotic pressure and eventual lysis of red blood cells infected with parasites in the later stages of replication. A second synchronisation is sometimes necessary to obtain better synchronicity due to the broad window of infected red cells insensitive to the sorbitol treatment (Lambros & Vanderberg, 1979). Other synchronisation methods that take advantage of differences in temperature sensitivity, buoyancy or magnetic property of different parasite stages have also been developed (Kwiatkowski, 1989; Mons *et al*, 1985; Ahn *et al*, 2008).

As described in the first chapter, *Plasmodium* is highly dependent on glucose as an energy and carbon source. It has been reported that infected erythrocytes (in synchronised cultures) consume glucose at a much higher rate (between 86 to 186 μmol glucose per billion parasitised RBCs per day), compared with uninfected cells which consume 4.6 μmol glucose per billion RBCs per day (Jensen *et al*, 1983). In *in vitro* cultures, concentrations of non-phosphorylated sugars (including glucose) within infected RBCs have been shown to be similar to external concentrations, with the external and internal concentrations equilibrating within one hour. This suggests that the transport of these sugars is a passive rather than an active process (Kirk *et al*, 1996). With this in mind, the rate of glucose consumption by the parasite may be measured simply by monitoring the change in glucose concentration in the culture media (i.e., spent media) over a given period of time. In addition to glucose, the parasite is also dependent on amino acids from the media and from the digestion of haemoglobin. The parasite balances haemoglobin digestion with the availability of amino acids in the media (Liu *et al*, 2006) and maintains colloid osmotic pressure to prevent premature host cell lysis (Lew *et al*, 2003; Waldecker *et al*, 2017). This interplay between consuming and discarding amino acids may be observed through changes in amino acid concentration in spent media. Many genome scale metabolic model predictions are generated using FBA, which looks into the transformation of metabolites as they participate in the different metabolic reactions in the model. Experimentally

measurable inputs (i.e., influx of external metabolites through boundary reactions) and outputs (i.e., biomass production) are therefore important in generating accurate predictions using FBA (Yilmaz & Walhout, 2017). Given the importance of glucose and amino acids in parasite survival, experimentally measured rates of consumption of these metabolites will be a vital addition to a malaria metabolic model.

In this chapter, mannose was used as a competitive inhibitor of the *P. falciparum* hexose transporter, PfHT1, which transports glucose and mannose into the parasite (Woodrow *et al*, 2000). Inhibition of glucose uptake by the parasite was done to determine and quantify the effect of limiting the parasite's most important energy source on overall growth. Competitive inhibition of glucose uptake by mannose is not pronounced *in vivo* as the glucose concentration is nearly a hundred-fold higher than mannose in human serum (LeRoux *et al*, 2009). Mannose, which is primed by hexokinase to produce mannose 6-phosphate, generally participates in the production of glycosylphosphatidylinositol anchors which contribute to parasite virulence (Sanz *et al*, 2013; Olszewski & Llinás, 2011). Although it is a monosaccharide that may be utilised as an energy source in other organisms, mannose cannot sustain *in vitro* growth (Geary *et al*, 1985). Specific hexoses and amino-hexoses (e.g., fucose and glucosamine) have been shown to inhibit merozoite invasion, and thus limit overall parasite growth in culture; however, mannose demonstrated no such inhibitory effect (Weiss *et al*, 1981).

This chapter will provide the basic methods used in parasite culture, which are important in the generation of experimental data that were eventually utilised in improving and validating the model developed in the previous chapter. Experimental procedures for the collection of spent media and quantification of metabolites will also be described, and the results of metabolite flux calculations will be discussed in the light of parasite metabolism. Finally, the effects of glucose perturbation on the overall parasite growth were quantified, which will then be compared against model predictions as a means of model validation (Chapter 4).

3.2   Methodology

3.2.1   *Parasite culture*

Continuous *P. falciparum* 3D7 *in vitro* cultures were grown in filter-sterilised malaria complete media composed of Roswell Park Memorial Institute (RPMI) 1640 growth medium (with L-glutamine, HEPES and phenol red, Life Technologies) supplemented with 5% (w/v) Albumax II (Gibco), 2% (w/v) sodium bicarbonate (Sigma), 0.01% (w/v) hypoxanthine (Sigma) and 0.1% (v/v) gentamicin (10 mg/ml, Gibco) at 5% haematocrit (O+ blood was obtained from the National Blood Service of the National Health Service Blood and Transplant Unit (NHSBT) in

Seacroft, Leeds). Whole blood was washed three times with RPMI media to remove the serum and white blood cells prior to use. Washed blood was stored at 50% haematocrit in RPMI. Cultures were grown in a 37°C incubator in 25 cm^2 polystyrene non-vented tissue culture flasks (Nunclon). The culture flasks were individually gassed with 1% oxygen, 3% carbon dioxide and 96% nitrogen gas mixture using a sterile blunt needle connected to a filtered (0.22 μm pore size) gas tubing for 10 seconds prior to sealing the flask with the non-vented cap.

Initial cultures were prepared by thawing frozen stabilates and drawing out the freezing agent (0.9% NaCl, 4.8% sorbitol and 28% glycerol) from the RBCs by a series of NaCl solutions in decreasing concentration. To the thawed stabilates, 0.2x stabilate volume of 12% NaCl was added dropwise and was allowed to stand at room temperature for 3 minutes. Then 10x stabilate volume of 1.6% NaCl was added and the solution was mixed gently. The mixture was centrifuged for 5 minutes at 3000 rpm. The supernatant was removed and 10x pellet volume of 0.2% glucose/0.9% NaCl solution was added dropwise. The mixture was once again centrifuged for 5 minutes at 3000 rpm. The supernatant was removed and 0.25 ml washed RBCs (50% haematocrit) and 6 ml of malaria complete medium were added. The culture was transferred into a 25 cm^2 non-vented tissue culture flask, gassed and incubated at 37°C.

For regular maintenance, the culture media was changed daily by initially tilting the flask for about 30 minutes to allow the red blood cells to settle on the rear bottom edge of the flask. The used media was carefully removed using a transfer pipette and a sample of the blood was collected and smeared on a glass slide (Figure 3.1). New pre-warmed media was then added into the flask and the culture was gassed before being placed back into the incubator. The blood smear was air dried and fixed using 100% methanol. The fixed blood was then soaked in 10% Giemsa (VWR Chemicals) in Sorensen's buffer (10 mM NaH$_2$PO$_4$, 28 mM Na$_2$HPO$_4$, pH 7.2) for 10 minutes. The stain was then carefully rinsed off with tap water and the slide was air dried before reading by light microscopy under oil-immersion (1000x magnification).
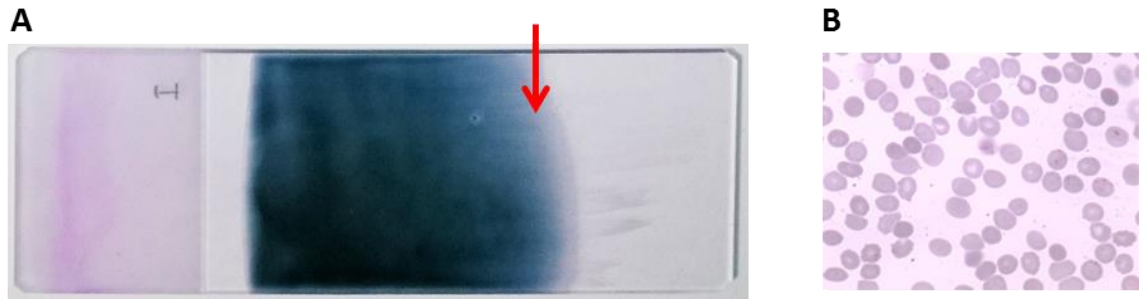
Figure 3.1   Image of a thin blood smear stained with 10% Giemsa
A blood sample is taken from the parasite culture to monitor the parasitaemia and to evaluate the status of the culture. The blood sample is smeared onto a slide (A) to give a thin layer of cells at the terminal end of the smear. In this area (red arrow), the red blood cells do not overlap (B), allowing accurate cell and parasite counting.

A Miller graticule (Figure 3.2a) was installed in the microscope eyepiece and was used as an aid for parasite counting. This graticule is normally used in counting the percentage of reticulocytes in a blood smear (Bain, 2006). The graticule, illustrated in Figure 3.2b, is composed of two squares, an outer square (in blue) and an inner square (in green), with the inner square having an area that is a tenth of the area of the outer square. When doing a parasite count, the graticule is placed in a field where the red blood cells are evenly distributed. The inner square is used to give a quick estimate of the total number of red blood cells in the outer square. This is done by multiplying the red blood cell count in the inner square by 10. The total number of infected red blood cells in the outer square (including those in the inner square) is also noted down.
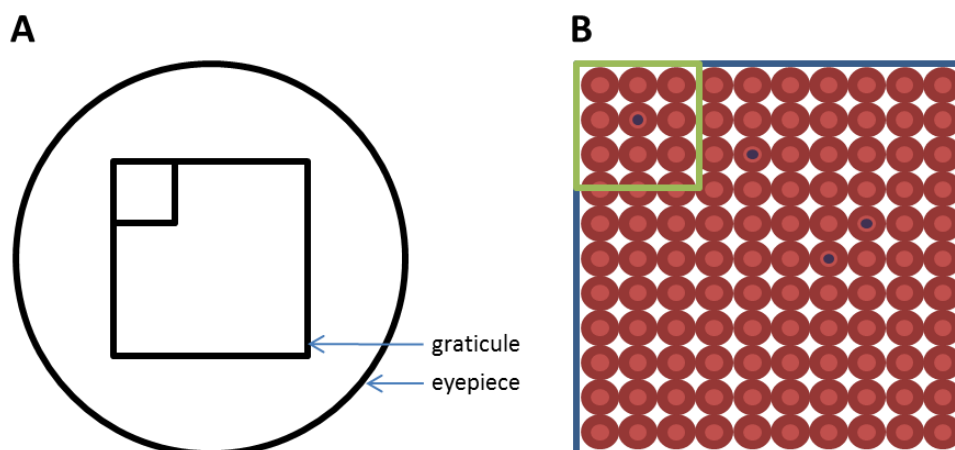
Figure 3.2   An illustration of a Miller graticule used to assist in parasite counting
The Miller graticule (A) was placed in the microscope eyepiece to help estimate the total number of red blood cell in the graticule field. The inner square (B, in green) represents a tenth of the whole area of the graticule (in blue). Multiplying the total number of red blood cells in the small square by ten gives an estimate of the total number of red bloods cells in the whole graticule field.

The parasitaemia was calculated from the total number of infected red blood cells as a percentage of the estimated total number of red blood cells counted (Equation 2.1). It is possible that a single red blood cell can be infected by multiple parasites. In this case, the infected cell is only counted as one. As these cultures were unsynchronised, the number of rings, late trophozoites and schizonts were also noted. In calculating the percentage of each parasite stage, a total of 100 parasites are counted and the number of parasites at each stage are noted and expressed as a percentage of the total number of parasites counted. When calculating for the parasitaemia, at least 1,000 red blood cells were observed and counted for daily monitoring while at least 2,000 red blood cells were examined prior to the conduct of any experiment. In addition, the haematocrit was also measured using a haemocytometer prior to any experiment. Given the measured parasitaemia and haematocrit, the desired parasitaemia and haematocrit were adjusted using Equation 3.2, Equation 3.3 and Equation 3.4. These equations give the volume of initial culture (to be subcultured), washed RBCs and new media necessary to adjust the parasitaemia and haematocrit to the desired percentages and culture volume.

$$Parasitaemia = \frac{total\ number\ of\ infected\ red\ blood\ cells}{estimated\ total\ number\ of\ red\ blood\ cells} \times 100\% \qquad \textbf{Equation 3.1}$$

$$V_i = \frac{P_f H_f V_f}{P_i H_i} \qquad \textbf{Equation 3.2}$$

$$V_r = \frac{H_f V_f}{50} \times \left(1 - \frac{P_f}{P_i}\right) \qquad \textbf{Equation 3.3}$$

$$V_m = V_f - V_i - V_r \qquad \textbf{Equation 3.4}$$

where:
V_i	=	volume of initial culture
V_f	=	final volume
P_i	=	initial parasitaemia
P_f	=	final parasitaemia
H_i	=	initial haematocrit
H_f	=	final haematocrit
V_r	=	volume of washed RBCs (50% haematocrit)
V_m	=	volume of media (malaria complete media)

The parasitaemia of the cultures were normally maintained under 4%, and the parasites were subcultured as needed. Subculturing was done by adding a calculated volume from the previous culture and diluting it with 5% haematocrit in fresh malaria complete media. As part of general parasite maintenance, a culture with a 4% parasitaemia was normally subcultured to

achieve a final parasitaemia of 0.5% (1:8 dilution) by adding 0.75 ml of the original culture to 5.25 ml of 5% washed RBCs in malaria complete media in a new 25 cm^2 tissue culture flask. For larger culture volumes, 18 and 42 ml of cultures were grown in 75 and 175 cm^2 tissue culture flasks (Nunclon), respectively. Larger flasks required longer gassing time.

Sorbitol was used to obtain synchronised cultures. The synchronisation procedure required cultures with high parasitaemia (> 4%) with a majority of the parasites in the ring stage (Lambros & Vanderberg, 1979). The culture was placed in a 50 ml Falcon tube and was centrifuged at 3000 rpm for 5 minutes. The media was removed and replaced with 0.5x media volume of filter-sterilised 5% sorbitol. The parasites were incubated in sorbitol for 10 minutes at room temperature. After incubation, pre-warmed media (5x sorbitol volume) was added to the culture to lower the sorbitol concentration before centrifugation at 3000 rpm for 5 minutes. The supernatant was removed and the pellet was resuspended with enough pre-warmed malaria culture media to dilute the haematocrit to 5%. The culture, which contained mostly ring stage parasites at this point, was transferred to a new tissue culture flask and was maintained as above. Two synchronisations were performed 40 hours apart prior to the experiment to obtain highly synchronous ring stage parasites (Figure 3.3).

Sterile materials and filter-sterilised (pore size of 0.22 μm) reagents were utilised in parasite culture. NaCl solutions for thawing parasites were autoclaved prior to use. Experimental procedures (except for slide staining) were done following proper aseptic techniques.
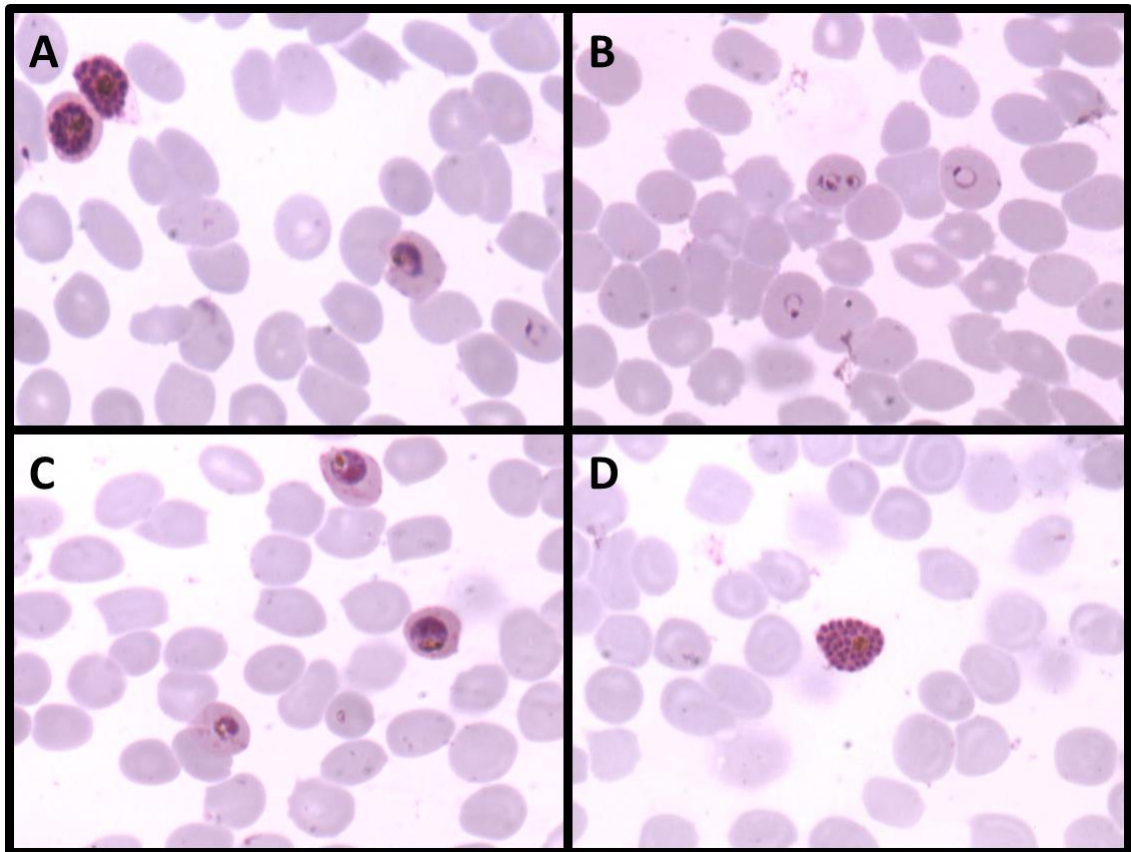
Figure 3.3 Image of a thin blood smear showing unsynchronised and synchronised *Plasmodium falciparum* **3D7 cultures**
This shows a comparison between unsynchronised (A) and synchronised (B-D) parasite cultures. Upon sorbitol synchronisation, the culture is composed mostly of ring stage parasites (B). These parasites develop to late trophozoites (C) at around 24 hours post synchronisation, and schizonts (D) around 40-42 hours post synchronisation.

3.2.2 Spent media collection

Synchronised cultures with a total volume of 18 ml at 1% parasitaemia and 5% haematocrit were placed in non-vented 75 cm^2 tissue culture flasks. Non-infected red blood cells at 5% haematocrit were used as control. One millilitre of culture was collected at time 0 (t_0), and every 6 hours until 48 hours post synchronisation (t_{48}). It was important that the culture was mixed gently until homogenous prior to sample collection. This ensured that the haematocrit was maintained at 5%. The collected sample was placed in a sterile 1.5 ml microcentrifuge tube, and centrifuged at 3000 rpm for 2 minutes. The spent media was collected and immediately stored at -80°C for later analysis. After the sample collection at 24 hours post synchronisation, the old media in the culture flask was replaced with new media (for both infected and control cultures). This was done by placing the culture in a sterile Falcon tube and spinning it at 3000 rpm for 5 minutes. The media was carefully removed and replaced with new

pre-warmed malaria complete media. At that point, the culture volume was 13 ml (after five 1 ml-sample collections), thus 12.35 ml (95% of 13 ml) of malaria complete media was added to the infected RBCs giving a final haematocrit of 5%. Three biological replicates were tested for this experiment.

3.2.3   Glucose assay

The glucose concentration in the spent media was determined using a Megazyme Glucose Assay (Glucose Oxidase-Peroxidase format) Kit. Using a 96-well clear-bottomed microtiter plate (Sarstedt), 300 µl of the glucose determination reagent (glucose oxidase, peroxidase, p-hydroxybenzoic acid and 4-aminoantipyrine, prepared as written in the assay manual of procedures) was added to 10 µl of sample and was incubated at 40°C for 20 minutes. Water (10 µl) added to 300 µl of glucose determination reagent was used as a blank, while 10 µl of 1 g/l glucose solution (provided with the assay kit) in the same amount of glucose determination reagent was used as control. The end point absorbance at 510 nm was then read using a microplate reader (SpectraMax 340PC). The concentration in g/l was calculated by dividing the difference of the absorbance of the sample (A_{sample}) and the blank (A_{blank}) by the difference of the absorbance of the control ($A_{control}$) and the blank. The unit of glucose concentration was converted to millimolar for the flux calculations by dividing the concentration in g/l by the molecular weight ($MW_{glucose}$) of glucose in g/mol, then multiplied by 1000 (Equation 3.5).

$$Glucose\ concentration\ (mM) = \frac{A_{sample} - A_{blank}}{A_{control} - A_{blank}} \times \frac{1000}{MW_{glucose}}$$ **Equation 3.5**

3.2.4   Amino acid concentration determination

Amino acid concentrations in the spent media were determined using an Ultimate 3000 High-Performance Liquid Chromatography (HPLC) system (Dionex). The HPLC protocol utilised a reverse-phase chromatography with an Acclaim 120 C18 (Dionex) 100 mm x 2.1 mm column (3 µm particle size, 120 Å pore size) stationary phase. The mobile phase was made up of an aqueous and organic gradient. Eluent A was composed of 10 mM Na_2HPO_4 (Sigma), 10 mM $Na_2HB_4O_7 \cdot 10\ H_2O$ (Sigma) and 0.5 mM NaN_3 (Sigma), adjusted to pH 8.2, and Eluent B was composed of 45% (v/v) methanol (HPLC grade, Fisher Scientific) 45% (v/v) acetonitrile (HPLC grade, Fisher Scientific) in water (Chromasolv Plus, SLS). The HPLC protocol was a modified version of that developed by Steiner *et al* (2009). The HPLC program was developed using a ramp gradient from 2% to 60% Eluent B over a period of 60 minutes. The program was then adjusted

to decrease the slope of the gradient (for better peak resolution) and a 10-minute equilibration was added at the end (Table 3.1).

Table 3.1   HPLC Program

Time (min)	Percent Eluent A	Percent Eluent B	Flow rate (ml/min)
0	98%	2%	0.200
60	55%	45%	0.200
60.5	98%	2%	0.200
70.5	98%	2%	0.200

Samples were derivatised prior to chromatography. The derivatisation procedure was done in a 1.5 ml microcentrifuge tube at room temperature by adding the following reagents in order shown below:

1.   300 μl of borate buffer (0.1 M $Na_2HB_4O_7 \cdot 10H_2O$ (Sigma), pH 10.2)

2.   15 μl of o-phthaldialdehyde (OPA) solution (75 mM OPA (SLS), 225 mM 3-mercaptopropionic acid (MPA) (Sigma) in 0.1 M borate buffer, pH 10.2)

3.   3 μl of sample (mixed 5 times using a Gilson pipette set at 300 μl), incubated for 60 seconds at room temperature

4.   6 μl 9-fluorenylmethoxycarbonyl chloride (FMOC) solution (2.5 mg/mL FMOC (Sigma) in HPLC grade acetonitrile) (mixed 5 times using a Gilson pipette set at 300 μl)

5.   42 μl phosphoric acid solution (15 μl/ml 85% phosphoric acid in Eluent A)

Derivatised samples were filtered through a 13 mm polytetrafluoroethylene (PTFE) syringe filter (0.22 μm pore size, Fisher Scientific) before placing into a 250 μl polypropylene HPLC vial. Even though derivatisation using OPA and MPA has been shown to be stable over a period of 90 minutes (Molnár-Perl, 2001), slight decrease in peak areas was observed after running the same sample again after 2 hours (during earlier experiments). Thus, to ensure minimal degradation of the derivatised compounds, the derivatisation procedure was only done within 10 minutes prior to running the sample through the HPLC and the whole run time was limited to under 80 minutes. Detection of derivatised amino acids was through UV absorbance at 338 nm at 10.0 Hz data collection rate (from 0 to 55 minutes). Amino acid peak areas were measured from the chromatograms using Chromeleon. Taking into consideration the long HPLC running time, selected spent media samples (i.e., those collected at t_6, t_{18}, t_{30}, t_{36}, t_{42} and t_{48}) were run for amino acid concentration determination. The change in concentration from t_6 to t_{18} was used to calculate amino acid flux at the mid-ring stage, from t_{30} to t_{36} for the late trophozoite stage and from t_{42} to t_{48} for the late schizont stage.

Individual amino acid standards were prepared in 0.1 N HCl. In addition, a stock solution containing all amino acids (with known concentrations) was prepared, taking into account the individual amino acid densities (when available) to ensure accurate standard concentrations (Table 3.2). Standards at different concentrations were run to generate standard curves. The amino acid peak areas from the chromatogram (y) were plotted against the standard concentration (x) using MS Excel 2010, forcing the y-intercept to zero. The concentration of a given amino acid in the spent media was then calculated by dividing its peak area by the slope of the corresponding amino acid standard curve.

Table 3.2   Amino acids and their corresponding chemical properties used in the preparation of amino acid standards (Budavari *et al*, 1989)

Amino acid	Molecular weight (g/mol)	Density (g/ml)	Solubility in water (mM)	Manufacturer/ Catalogue number
L-Alanine	89.1	1.371	1,430.0	Sigma, A7627-100G
(+)-L-Arginine HCl	210.7	1.325	Soluble	Sigma, A6969-25G
L-Asparagine	132.1	1.543	270.0	Sigma, A0884-25G
L-Aspartic acid	133.1	1.636	33.8	Sigma, A9256-100G
L-Cystine	240.3	1.655	0.4	Sigma, C7602-25G
L-Glutamic acid	147.1	1.566	58.7	Sigma, G1251-100G
L-(+)-Glutamine	146.1	Not available	329.0	Acros Organics, 10376840
Glycine	75.1	1.598	3,330.0	Sigma, G7126-100G
L-Histidine monohydrochloride monohydrate	209.6	1.412	Fairly soluble in water	Sigma, H5659-25G
Trans-4-hydroxy-L-proline	131.1	Not available	2,200.9	Sigma, H54409-25G
L-Isoleucine	131.2	1.201	314.1	Sigma, I7403-25G
L-Leucine	131.2	1.167	185.0	Sigma, L8912-25G
L-Lysine monohydrochloride	182.7	1.237	Very freely soluble	Sigma, L8662-25G
L-Methionine	149.2	1.311	Soluble in water	Sigma, M5308-25G
L-Phenylalanine	165.2	1.315	179.2	Sigma, P2126-100G
L-Proline	115.1	1.376	14,071.1	Sigma, P5607-25G
L-Serine	105.1	1.582	Soluble in water	Sigma, S4311-25G
L-Threonine	119.1	1.499	Freely soluble	Sigma, T8441-25G
L-Tryptophan	204.2	1.303	55.8	Sigma, T-0655
L-Tyrosine	181.2	1.403	2.5	Sigma, T3754-25G
L-Valine	117.2	1.267	711.9	Sigma, V0513-25G

3.2.5 Flux calculation

Spent media for both infected and uninfected (control) samples were assayed for glucose and amino acid concentrations as detailed above. The metabolite flux from time a (t_a) to time b (t_b) was calculated using the formula:

$$flux_{t_a-t_b} = \frac{v_{t_a} \times (\Delta C_i - \Delta C_u)}{gDW_{t_a} \times \Delta t}$$ **Equation 3.6**

where:

v_{t_a}	=	volume at t_a, in litres
ΔC_i	=	change in metabolite concentration in infected culture from t_a to t_b, in mM
ΔC_u	=	change in metabolite concentration in uninfected culture from t_a to t_b, in mM
gDW_{t_a}	=	gram dry weight of parasite at t_a (see Equation 3.7)
Δt	=	change in time ($t_b - t_a$), in hours

The change in metabolite concentration in uninfected culture (ΔC_u) was subtracted from the observed change in metabolite concentration in infected culture (ΔC_i) to account for the background effect of metabolism in uninfected RBCs (Yayon *et al*, 1983). The change in metabolite concentration (in infected or uninfected culture) was calculated by subtracting the concentration (in mM) at t_b from that at t_a. Thus, a positive ΔC (i.e., $C_{t_a} > C_{t_b}$) indicates a decrease in metabolite concentration in the media. Correspondingly, a positive flux signifies entry of metabolites into the cell, while a negative flux indicates movement of metabolites from the red blood cell into the surrounding media.

The parasite mass (gDW_{t_a}) was determined using the calculated number of parasites, which was based on the parasitaemia, haematocrit and the culture volume at t_a (Equation 3.7). For simplicity, two assumptions were made. (1) It was assumed that each infected RBC contained only one parasite. (2) Given a synchronised culture, it is assumed that the parasitaemia (set at 1%) remained the same throughout the 48 hour experiment, as each parasite in the culture would have merely developed from rings to trophozoites to schizonts without asexual multiplication and invasion of neighbouring red blood cells. Thorough but gentle mixing of the culture prior to sample collection ensured that the haematocrit (5%) was maintained throughout the duration of the experiment. The total parasite mass was finally calculated by multiplying the number of parasites by the mass per parasite $\left(m_p = 10.5 \times 10^{-12} \frac{gDW}{parasite}\right)$ (Forth, 2012). Table 3.3 shows the estimated amount of parasites (in gram dry weight) used to calculate for the metabolite flux in spent media at different time points. Flux calculations were done to determine

the rate of metabolite consumption/production by cultures at different stages of parasite development (i.e., mid-rings, late trophozoites and late schizonts).

$$gDW_{t_a} = k \times p \times h \times v_{t_a} \times m_p$$

<div align="right">**Equation 3.7**</div>

where:

k	=	haematocrit constant $\left(10^{13} \frac{RBCs}{litre\ of\ RBCs}\right)$
p	=	parasitaemia at t_a
h	=	haematocrit at t_a
v_{t_a}	=	culture volume at t_a, in litres
m_p	=	mass per parasite

Table 3.3   Culture volume and estimated mass of parasite in culture (gDW) at different time points during spent media collection used to calculate for metabolite flux

Hours post synchronisation	Culture volume (litre)	Estimated total gDW of parasite in culture (gram)
0	0.017	0.000893
6	0.016	0.000840
12	0.015	0.000788
18	0.014	0.000735
24	0.013	0.000683
30	0.012	0.000630
36	0.011	0.000578
42	0.010	0.000525
48	0.009	0.000473

3.2.6   In vitro glucose perturbation

To determine the effect of lowered parasite glucose consumption on growth, initial experiments were performed by growing the parasites in lower media glucose concentrations; however, it was later realised that lowering media glucose can affect not just parasite growth but also RBC metabolism. Thus it was important to reduce parasite glucose consumption without compromising glucose utilisation in the host cell. Mannose, which is cheap and widely available, was used as a competitive inhibitor for the *P. falciparum* hexose transporter 1 (PfHT1), which transports glucose present in the red blood cell into the parasite. Glucose enters the uninfected red blood cell through facilitative diffusion via GLUT1, which comprises 5% of the cell membrane (Mueckler, 1994). In infected cells, glucose enters the host cell more freely through the NPPs as well as through GLUT1 (Krishna *et al*, 2000).

Mannose has an inhibition/dissociation constant (K_i) of 1.1 mM for PfHT1, while glucose has a Michaelis constant (K_m) of 1.0 mM (Woodrow *et al*, 2000). The percent of glucose transport inhibition ($\frac{v_i}{v_u}$) can be calculated based on the media concentrations of mannose and

glucose using the Michaelis-Menten equations with (Equation 3.8) and without (Equation 3.9) a competitive inhibitor (Stryer, 1988). The two equations were combined to give Equation 3.10. For simplicity, the percent activity of PfHT1 with respect to the transport of glucose into the parasite will be called "percent PfHT1 activity" from here on.

$$v_i = \frac{V_{max}[glucose]}{K_m \left(1 + \frac{[mannose]}{K_i}\right) + [glucose]}$$

Equation 3.8

$$v_u = \frac{V_{max}[glucose]}{K_m + [glucose]}$$

Equation 3.9

$$\frac{v_i}{v_u} = \frac{K_m + [glucose]}{K_m \left(1 + \frac{[mannose]}{K_i}\right) + [glucose]} \times 100\%$$

Equation 3.10

where:

v_i	=	rate of glucose transport with the inhibitor (mannose)
v_u	=	rate of glucose transport without the inhibitor
K_m	=	Michaelis constant for glucose
K_i	=	inhibition constant for mannose
$[mannose]$	=	mannose concentration
$[glucose]$	=	glucose concentration
V_{max}	=	maximum rate of reaction

The concentrations of mannose were calculated using Equation 3.10 to result in 25% to 100% PfHT1 glucose transport in the presence of 8.0 mM glucose (Table 3.4). Concentrations of sorbitol equal to that of mannose were used in the control. Zero PfHT1 activity was simulated by growing the parasites in 0 mM glucose and 8.0 mM of mannose or sorbitol.

Table 3.4 Calculated *P. falciparum* hexose transporter 1 (PfHT1) activity at different glucose and mannose concentrations

Percent PfHT1 activity	$[glucose]$ (mM)	$[mannose]$ (mM)
0%	0.00	8.00
25%	8.00	29.70
40%	8.00	14.85
50%	8.00	9.90
60%	8.00	6.60
75%	8.00	3.30
80%	8.00	2.48
90%	8.00	1.10
100%	8.00	0.00

A lower glucose concentration (8.0 mM) compared to the glucose concentration in normal malaria complete media (11.0 mM) was used to avoid negative growth effects of high media osmolarity (Dei-Cas *et al*, 1985; Ginsburg *et al*, 1986a). Parasites were also grown in media with 11.0 mM glucose as additional control. No-glucose media was initially prepared using no-glucose RPMI (with L-glutamine and phenol red, Life Technologies) supplemented with the same components as detailed in the methodology section of this chapter, in addition to 25 mM HEPES (Sigma), which is absent in no-glucose RPMI. Appropriate amounts of filter-sterilised (0.22 μm pore size) glucose, mannose and/or sorbitol 400 mM stock solutions were added to aliquots of no-glucose media to obtain the concentrations listed in Table 3.4.

Unsynchronised parasite cultures at 0.5% parasitaemia and 3% haematocrit were grown in black-sided, clear flat-bottomed 96-well plates (Costar) in media with different glucose and mannose/sorbitol concentrations (200 μl total culture volume per well). Uninfected RBCs at 3% haematocrit were used as an additional control. The cultures were incubated in 1% oxygen, 3% carbon dioxide and 96% nitrogen in a humidified sealed chamber at 37°C. The chamber was gassed with the same gas mixture after 24 hours. After 48 hours, the relative numbers of parasites were quantified using SYBR Green following the protocol of Smilkstein *et al* (2004). This procedure involved the addition of 100 μl of 3x lysis buffer to each well. The 3x lysis buffer consisted of 0.024% (w/v) saponin (BDH), 0.24% (v/v) Triton X-100 (Sigma), 60 mM Tris (pH 7.5), 15 mM EDTA and 0.3 μl/ml SYBR Green I 1000x concentration (Thermo Fisher). SYBR Green was only added into the lysis buffer prior to use. Upon addition of the lysis buffer, the plates were covered in aluminium foil and incubated at room temperature for 45 minutes. Fluorescence was then measured using a multifunctional microplate reader (POLARstar OPTIMA, BMG LABTECH) set at 485 nm excitation and detection at 520 nm. At least three biological replicates were performed for this experiment (3 replicates for 40%, 60%, 80% and 90% PfHT1 activity; 4 replicates for 50% and 75% PfHT1 activity; and 7 replicates for 0%, 25% and 100% PfHT1 activity).

In vitro growth was presented as a percentage of the relative numbers of parasites in media without the inhibitor (100% PfHT1 activity). Background fluorescence from parasites grown in 0 mM glucose was subtracted from all other measured fluorescence (f) prior to the calculation of percent *in vitro* growth. The calculation for the percent *in vitro* growth at a given percentage of PfHT1 activity is shown in Equation 3.11. Percent *in vitro* growth (y) was plotted against the percent PfHT1 activity (x) using GraphPad Prism. The Shapiro-Wilk test was used to test for normality of data and T-test to compare the means of individual groups against the control. Differences in *in vitro* growth as a result of lowered PfHT1 activity as well as increased media osmolarity were determined. Pearson's test was also performed to examine the

correlation between percent parasite growth and percent PfHT1 activity. All statistical tests were done using IBM SPSS Statistics 20.

$$\% \text{ in vitro growth at } X\% \text{ PfHT1 activity}$$

$$= \frac{f_{X\% \text{ PfHT1 activity}} - f_{0\% \text{ PfHT1 activity}}}{f_{100\% \text{ PfHT1 activity}} - f_{0\% \text{ PfHT1 activity}}} \times 100\% \qquad \textbf{Equation 3.11}$$

where:

$f_{X\% \text{ PfHT1 activity}}$ =   fluorescence at a given percentage of PfHT1 activity
$f_{0\% \text{ PfHT1 activity}}$ =   fluorescence at 0% PfHT1 activity
$f_{100\% \text{ PfHT1 activity}}$ =   fluorescence at 100% PfHT1 activity

3.3 Results

3.3.1 Glucose concentration in spent media

Glucose concentration in spent media was monitored over a 48-hour time course as part of this study. Given that the movement of glucose into the infected red blood cell is considered as an equilibrative process (Kirk *et al*, 1996), monitoring changes in glucose concentration in the media was utilised as an alternative means of monitoring the rate of glucose consumption by the parasite. Cultures utilised for this experiment were standardised to have a parasitaemia of 1% and a haematocrit of 5%, which necessitated a change in media after 24 hours (Jensen *et al*, 1983). Although a sample of spent media at 24 hours post synchronisation was collected prior to changing the media, the addition of new media contributed to an apparent increase in metabolites measured at 30 hours post synchronisation (Table 3.5 and Figure 3.4). Therefore, flux values between 24 and 30 hours were excluded from the calculations.

Table 3.5   **Average glucose concentration (in mM) in spent media collected from infected and uninfected cultures at different time points after parasite synchronisation (n = 3)**

Hours post synchronisation	Infected RBCs		Uninfected RBCs (control)	
	Average	SEM	Average	SEM
0	9.90	0.07	10.21	0.41
6	9.71	0.16	9.93	0.46
12	9.00	0.36	9.45	0.38
18	8.15	0.39	9.19	0.23
24	7.49	0.50	8.82	0.48
30	9.14	0.41	9.45	0.23
36	8.06	0.69	9.39	0.22
42	7.10	0.90	9.10	0.40
48	6.64	0.86	8.83	0.37
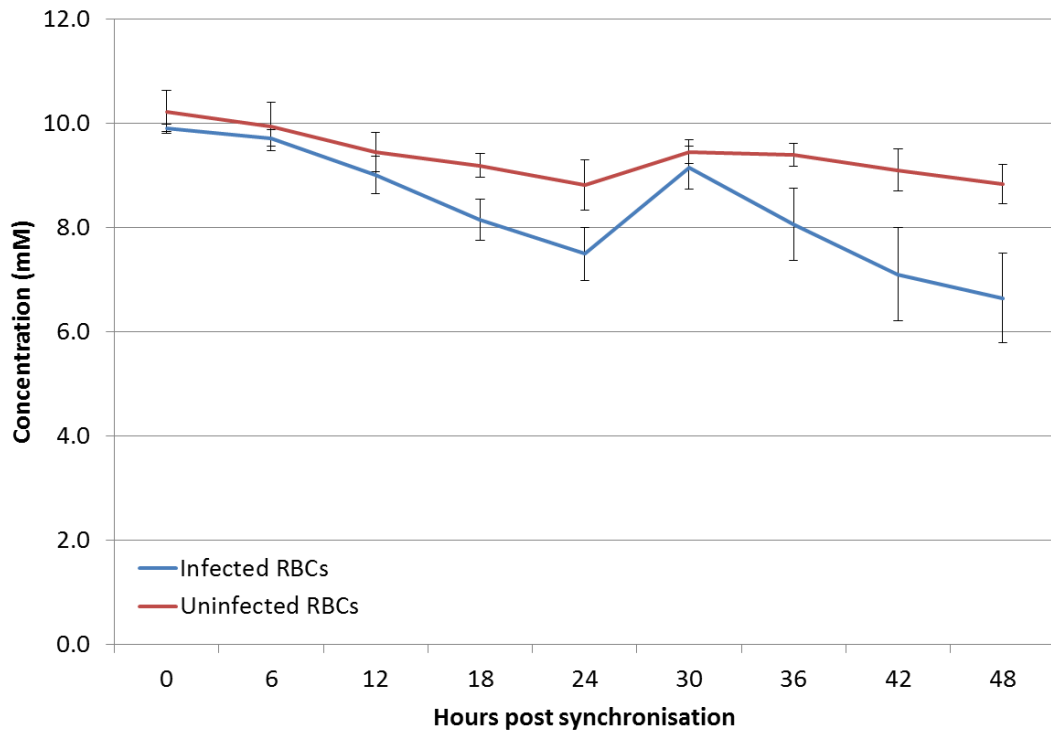
Figure 3.4   Average glucose concentration in spent media collected from infected and uninfected cultures at different time points after parasite synchronisation
Glucose concentration (± SEM) in spent media was measured at 6-hour intervals throughout the 48-hour life cycle of *P. falciparum* 3D7 in *in vitro* culture at 1% parasitaemia and 5% haematocrit. The cultures required a change in media after collecting spent media at 24 hours, causing an apparent increase in glucose concentration at 30 hours post synchronisation. (n = 3)

By simply looking at the change in concentration (Figure 3.4), it may be tempting to deduce that there was a steady rate of consumption of glucose by the parasite over time. However, it is important to remember that the total number of parasites decreased over time as a result of sample collection done at 6-hour intervals. It is therefore important to take into consideration the number of parasites at the beginning of each interval to get a more accurate picture of the rate of glucose consumption (i.e., glucose flux). In addition, the background glucose consumption of uninfected red blood cells was also taken into account in the calculation of the glucose flux in infected cells. Figure 3.5 shows the glucose flux over the 48 hour time period (excluding 24-30 hours post synchronisation). The overall average glucose flux was calculated to be 1.32 ± 0.39 mmol/gDW/hr, which is close to the value obtained by Forth (1.27 mmol/gDW/hr) using NMR-measured flux values in spent media of *P. falciparum* 3D7 *in vitro* cultures (Forth, 2012). A similar trend in glucose consumption in synchronised parasite cultures was observed in a previous study where lower glucose consumption was noted in cultures with higher percentage of ring stage parasites (Jensen *et al*, 1983). The late trophozoite stage exhibited the highest glucose consumption as expected since the trophozoite stage is the most

metabolically active of the three stages (Kirk & Lehane, 2014). Expression patterns of genes involved in glycolysis were also shown to reach maximum expression during the early trophozoite stage (Bozdech *et al*, 2003). This coincides with the peak in glucose flux observed here at 30-36 hours post synchronisation, if the delay between transcription and translation is taken into consideration. An evaluation of expression data of more than 200 genes and their associated protein in *P. falciparum* revealed a median time delay of 11 hours between peak mRNA and corresponding protein expressions (Foth *et al*, 2011). A similar pattern in the concentration of lactate dehydrogenase enzyme (PfLDH), the last step in the glycolysis pathway, was observed in synchronised *P. falciparum* 3D7 culture over the 48-hour parasite life cycle (Vivas *et al*, 2005). A decrease in PfLDH activity was observed at 18 hours, which was followed by a gradual rise until 30 hours when peak activity was observed.
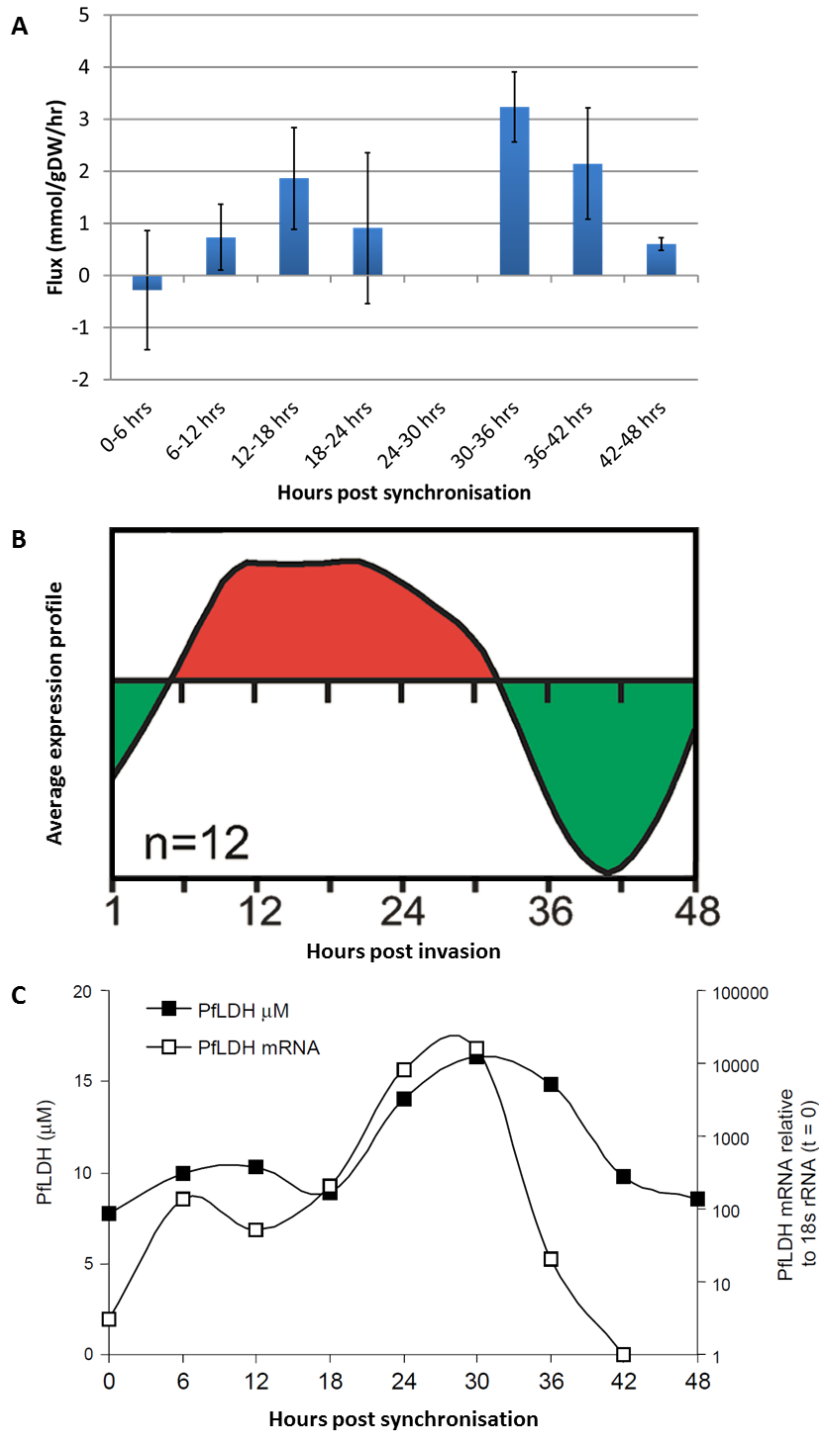
Figure 3.5   Calculated glucose flux at different intervals within the 48-hour parasite life cycle in a synchronised culture

(A) Calculated glucose flux (in mmol/gDW/hr, ± SEM) at 6-hour intervals throughout the 48-hour parasite lifecycle (n = 3) is shown to coincide with (B) the expression pattern of *P. falciparum* genes related to glycolysis (Bozdech *et al*, 2003) and (C) the expression pattern (white squares) and concentration (solid squares) of *P. falciparum* lactate dehydrogenase (PfLDH), the enzyme involved in the last step of glycolysis (Vivas *et al*, 2005). Note that the flux between 24 and 30 hours (A) was not calculated as previously explained at the beginning of the results section. (Figure 3.5B was obtained from an open access journal, while permission from the publisher was obtained for the reprint of Figure 3.5C).

3.3.2   Amino acid concentration in spent media and flux calculations

Two main reagents were used for amino acid derivatisation allowing the amino acids to be measureable by ultraviolet (UV) detection: O-phthaldialdehyde (OPA) reacts with primary amines, while 9-fluorenylmethoxycarbonyl chloride (FMOC) reacts with secondary amines (Schuster, 1988). Proline and hydroxyproline are the only secondary amines in the amino acid standard solution and the rest are considered primary amines (Kaspar *et al*, 2009). OPA and FMOC are detected at 338 nm and 263 nm wavelengths, respectively (Bartolomeo & Maisano, 2006; Schuster, 1988). The HPLC protocol, as previously mentioned in the Methodology section, was based on the protocol by Steiner *et al* (2009), which was intended for ultra-high-performance liquid chromatography (UHPLC). Compared to conventional HPLC, UHPLC can accommodate smaller stationary phase particle size and a greater pressure range, thus permitting higher flow rates and faster separation (Cielecka-Piontek *et al*, 2013). In addition, the original protocol utilised automated pre-column derivatisation, a feature that was not present in the HPLC equipment used in this study. A shift in UV wavelength from 338 nm to 263 nm towards the end of the run was required in the original protocol; however, the shift in wavelength resulted in an overall shift in baseline absorbance and failure to detect any peak at 263 nm detection. Therefore, the UV detection was set to 338 nm throughout the duration of the run, limiting the detection to the 19 primary amines.

In the generation of standard curves, a poor linear fit was generally observed with the curve for cystine compared to the rest of the amino acids which had R^2 values close to 1 (Figure 3.6). This irregularity may be attributed to condition-related susceptibility of cystine (dimeric form) to be reduced to its cysteine (monomeric form) (Alvarez-Coque *et al*, 1988), with the latter known to yield a weakly detectable complex with OPA (Birwé & Hesse, 1991). Therefore, concentration measurements for cystine were not used, giving a total number of 18 amino acids quantified in collected spent media.

A Aspartic acid

y = 2.2194x
R² = 0.9992

B Cystine

y = 16.455x
R² = 0.4486
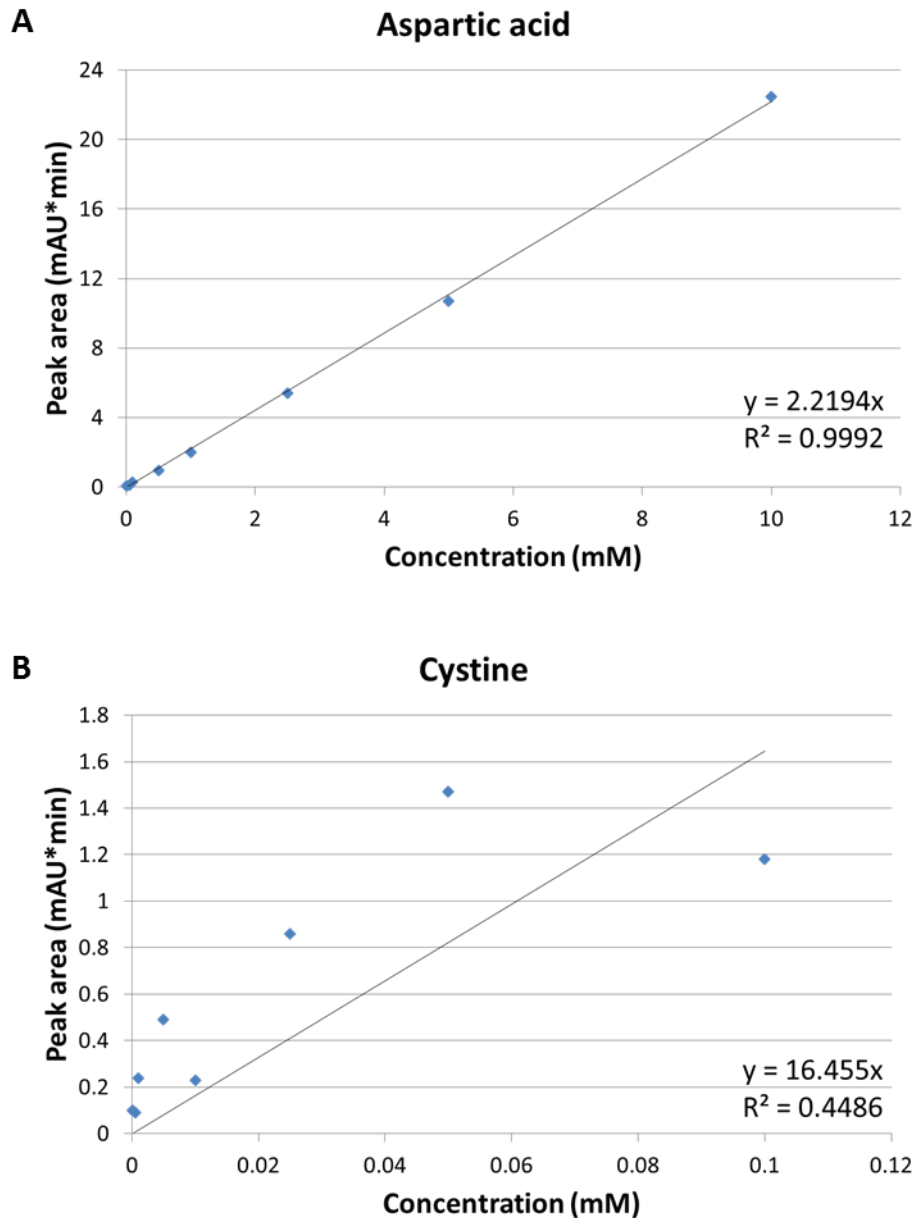
Figure 3.6   Representative standard curves showing peak areas measured from the chromatogram plotted against known standard concentrations
As an example, the standard curve for aspartic acid (A) is shown together with the trendline, linear equation and R^2 as a basis for linear fit. The cystine standard curve (B) below is showing a poor linear fit.

Fluctuations in amino acid concentrations in spent media are affected by several factors. RPMI 1640 medium used in parasite culture is composed of 20 amino acids (excluding alanine) (Moore & Woods, 1977). Through the remodelling of the RBC membrane by the parasite, amino acids (as well as other solutes) from the media can easily be transported into the infected host cell though the new permeability pathways, which are described as highly permeable, non-highly selective channels (Staines *et al*, 2006; Ginsburg & Stein, 2004; Bouyer *et al*, 2006). These channels allow the parasite to obtain the necessary amino acids that are absent or insufficient

in haemoglobin from the media. In addition, these channels permit the efflux of amino acids from haemoglobin digestion that are in excess of what the parasite requires to survive (Krugliak *et al*, 2002; Lew *et al*, 2003). This contributes to the apparent increase in the concentration of a number of amino acids in the media. The change in amino acid concentrations in the media may therefore be attributed to the balance between the relative abundance of amino acids in the media and in haemoglobin, and the amino acid requirements of the parasite (Table 3.6).

Table 3.6   Comparison of the relative abundance of amino acids in RPMI 1640 media (Life Technologies) **in mM, haemoglobin** (Magrane & UniProt Consortium, 2011) **and *P. falciparum* proteome** (Chanda *et al*, 2005) **in percentages. The red to blue colour scale represents a decreasing relative abundance of amino acid.**

Amino acid	RPMI 1640	Haemoglobin	Parasite proteome
Alanine	0.000	12.5%	2.4%
Arginine	1.149	2.1%	2.9%
Asparagine	0.379	3.5%	12.3%
Aspartic acid	0.150	5.2%	6.1%
Cystine	0.208	1.0%	1.8%
Glutamic acid	0.136	4.2%	2.8%
Glutamine	2.055	1.4%	7.1%
Glycine	0.133	6.9%	3.1%
Histidine	0.097	6.6%	2.2%
Isoleucine	0.382	0.0%	9.4%
Leucine	0.382	12.5%	8.2%
Lysine	0.274	7.6%	11.7%
Methionine	0.101	1.7%	2.2%
Phenylalanine	0.091	5.2%	4.6%
Proline	0.174	4.8%	2.2%
Serine	0.286	5.5%	6.4%
Threonine	0.168	5.5%	4.2%
Tryptophan	0.025	1.0%	0.5%
Tyrosine	0.111	2.1%	5.6%
Valine	0.171	10.7%	4.1%

Amino acid concentrations in spent media, both in infected and uninfected cultures, (Table 3.7 and Table 3.8) were used to calculate the flux to represent the three main parasite erythrocytic stages (Table 3.9, Table 3.10 and Table 3.11). The calculated glucose flux values were also included in these tables. The minimum, maximum flux values, as well as standard deviations and standard errors presented in these tables were eventually used as constraints in the model, which will be described in the next chapter. It can be appreciated here that there is a relatively high influx (positive flux values) of amino acids during the schizont stage. Conversely, a majority of amino acids during the ring and the trophozoite stages appeared to be moving out

of the infected cell. An experimental quantification of parasite protein at different stages alongside haemoglobin digestion has shown stage dependent haemoglobin digestion as well as parasite protein production (Krugliak *et al*, 2002). These investigators calculated a very low efficiency in conversion of haemoglobin to biomass by the parasite especially during the early stages. During the ring stage, only 5% of the digested haemoglobin was converted to parasite protein. This is consistent with the amino acid flux values observed in the ring stage in this study (Table 3.9) when most of the amino acids were excreted into the media. It is worth noting that there was an influx of isoleucine observed during the ring stage, consistent with the fact that the only source of isoleucine for the parasite is the media, as it is absent in haemoglobin (Magrane & UniProt Consortium, 2011).

Compared with the mid-ring stage data, calculated flux values at the trophozoite stage showed even more amino acids moving out of the infected cells, including isoleucine (Table 3.10). This pattern contradicts the fact that during the trophozoite stage (at around 24 to 30 hours post invasion), the parasite protein synthesis begins to increase significantly (Krugliak *et al*, 2002), and that isoleucine is absent in haemoglobin therefore it should be impossible for the concentration of isoleucine in the media to increase. A similar efflux of isoleucine was also observed by Forth who measured amino acid flux through the analysis of spent media in *P. falciparum in vitro* cultures (Forth, 2012). Metabolite concentrations measured over a period of 48 hours done by Olszewski *et al* (2009) also showed a rise in concentration at 32 hours post infection for 10 out of 14 amino acids (i.e., asparagine, glutamate, histidine, (iso)leucine, methionine, phenylalanine, proline, serine, tryptophan and tyrosine). Note that liquid chromatography-tandem mass spectrometry was utilised for quantifying metabolite concentrations in the media thus the total isoleucine and leucine concentrations were reported as a single value.

It is possible that the shifts in total volume of the infected cell and the gradual increase in parasite volume may result in shifts in the effective space inside the infected red blood cell surrounding the parasite. *In silico* modelling of infected cell and parasite volume was done by simulating homeostasis in infected red blood cells as a function of haemoglobin digestion, movement of important ions and parasite growth (Lew *et al*, 2003). Using the model, which was later validated experimentally (Waldecker *et al*, 2017), they have shown that the parasite and the host cell volumes were stable during the early stages but eventually increased at disproportional rates at 24 hours post invasion. Parasite volume was shown to increase at a higher rate compared to the host cell during the trophozoite stage, until about 42 hours post invasion when the parasite and the RBC volume begin to stabilise. Moreover, the effective space between the parasite and the RBC membrane decreases from around 32 to 40 hours post

invasion (Lew *et al*, 2003). It could therefore be speculated that the rapid growth of the parasite during the trophozoite stage, alongside the slower increase in host cell volume, may account for the movement of water and solutes out of the red blood cell. This phenomenon may explain the observed efflux of a majority of amino acids including isoleucine during the late trophozoite stage. On the other hand, high protein synthesis, increase in colloid pressure and eventual host cell swelling around 40 hours post invasion (Allen & Kirk, 2004; Zarchin *et al*, 1986) may account for the influx of most amino acids during the schizont stage as shown in Table 3.11.

Table 3.7   Average amino acid concentrations (in mM) in spent media collected from infected cultures at different time points (hours post synchronisation, hps). The red to blue colour scale represents a decreasing concentration of amino acid. (n = 3)

	6 hps		18 hps		30 hps		36 hps		42 hps		48 hps	
	Average	SEM	Average	SEM	Average	SEM	Average	SEM	Average	SEM	Average	SEM
Alanine	0.010	0.001	0.030	0.005	0.012	0.003	0.025	0.009	0.032	0.010	0.036	0.012
Arginine	0.650	0.063	0.484	0.149	0.578	0.079	0.496	0.099	0.472	0.150	0.367	0.127
Asparagine	0.252	0.005	0.275	0.008	0.221	0.011	0.233	0.006	0.255	0.008	0.244	0.007
Aspartic acid	0.047	0.002	0.054	0.003	0.029	0.002	0.027	0.003	0.042	0.002	0.039	0.002
Glutamic acid	0.094	0.003	0.112	0.005	0.087	0.003	0.097	0.007	0.116	0.004	0.122	0.012
Glutamine	1.291	0.042	1.266	0.061	1.129	0.068	1.109	0.042	1.148	0.063	1.040	0.026
Glycine	0.301	0.006	0.323	0.008	0.320	0.015	0.321	0.007	0.169	0.007	0.166	0.012
Histidine	0.090	0.004	0.097	0.004	0.083	0.011	0.094	0.005	0.103	0.004	0.090	0.008
Isoleucine	0.254	0.004	0.266	0.010	0.217	0.010	0.226	0.004	0.249	0.013	0.235	0.001
Leucine	0.236	0.009	0.245	0.023	0.222	0.013	0.227	0.005	0.238	0.008	0.223	0.003
Lysine	0.107	0.009	0.121	0.009	0.109	0.022	0.091	0.005	0.117	0.018	0.103	0.017
Methionine	0.061	0.001	0.068	0.003	0.052	0.003	0.058	0.003	0.062	0.003	0.058	0.000
Phenylalanine	0.061	0.000	0.070	0.005	0.056	0.003	0.061	0.005	0.067	0.002	0.064	0.003
Serine	0.191	0.005	0.212	0.007	0.171	0.007	0.179	0.007	0.192	0.006	0.185	0.005
Threonine	0.110	0.001	0.126	0.005	0.102	0.005	0.110	0.005	0.116	0.003	0.111	0.007
Tryptophan	0.064	0.000	0.067	0.002	0.070	0.004	0.069	0.005	0.021	0.002	0.020	0.000
Tyrosine	0.074	0.002	0.082	0.003	0.066	0.003	0.069	0.002	0.075	0.002	0.071	0.003
Valine	0.123	0.007	0.144	0.007	0.127	0.012	0.127	0.009	0.142	0.003	0.138	0.013

Table 3.8   Average amino acid concentrations (in mM) in spent media collected from uninfected cultures at different time points (hours post synchronisation, hps). The red to blue colour scale represents a decreasing concentration of amino acid. (n = 3)

	6 hps		18 hps		30 hps		36 hps		42 hps		48 hps	
	Average	SEM	Average	SEM	Average	SEM	Average	SEM	Average	SEM	Average	SEM
Alanine	0.007	0.001	0.013	0.002	0.004	0.000	0.004	0.000	0.007	0.001	0.012	0.004
Arginine	0.670	0.045	0.549	0.086	0.714	0.087	0.584	0.028	0.535	0.068	0.501	0.103
Asparagine	0.239	0.009	0.255	0.012	0.260	0.024	0.241	0.010	0.251	0.005	0.261	0.018
Aspartic acid	0.041	0.003	0.045	0.000	0.044	0.007	0.042	0.002	0.041	0.003	0.044	0.005
Glutamic acid	0.094	0.004	0.100	0.004	0.095	0.008	0.088	0.005	0.094	0.001	0.101	0.007
Glutamine	1.224	0.030	1.199	0.038	1.309	0.112	1.165	0.053	1.154	0.012	1.158	0.087
Glycine	0.201	0.015	0.214	0.014	0.252	0.041	0.251	0.024	0.204	0.032	0.213	0.022
Histidine	0.076	0.004	0.078	0.006	0.090	0.014	0.080	0.002	0.078	0.002	0.085	0.006
Isoleucine	0.245	0.010	0.267	0.013	0.251	0.025	0.241	0.013	0.254	0.004	0.264	0.018
Leucine	0.343	0.020	0.531	0.025	0.254	0.027	0.217	0.010	0.226	0.015	0.226	0.023
Lysine	0.118	0.018	0.131	0.010	0.105	0.020	0.100	0.012	0.110	0.017	0.120	0.009
Methionine	0.059	0.001	0.065	0.003	0.062	0.005	0.059	0.005	0.061	0.001	0.065	0.004
Phenylalanine	0.058	0.002	0.061	0.003	0.059	0.004	0.053	0.003	0.059	0.002	0.064	0.004
Serine	0.186	0.007	0.195	0.009	0.195	0.019	0.183	0.007	0.189	0.005	0.201	0.014
Threonine	0.105	0.003	0.113	0.008	0.107	0.011	0.102	0.003	0.108	0.001	0.113	0.009
Tryptophan	0.031	0.004	0.036	0.004	0.051	0.011	0.052	0.009	0.034	0.010	0.036	0.011
Tyrosine	0.071	0.003	0.075	0.004	0.074	0.006	0.067	0.003	0.074	0.002	0.055	0.027
Valine	0.218	0.037	0.216	0.018	0.127	0.009	0.124	0.013	0.132	0.005	0.135	0.007

Table 3.9   Amino acid and glucose flux (in mmol/gDW/hr) measured between 6 and 18 hours post synchronisation representing flux during the mid-ring stage. The red to blue colour scale (in the first three columns) represents decreasing flux values. (n = 3)

	Average	Minimum	Maximum	SD	SEM
Alanine	-0.023	-0.041	-0.007	0.017	0.01
Arginine	0.072	0.007	0.202	0.112	0.065
Asparagine	-0.009	-0.021	0.000	0.011	0.006
Aspartic acid	-0.005	-0.014	0.007	0.011	0.006
Glucose	1.300	0.677	2.224	0.816	0.471
Glutamic acid	-0.019	-0.028	-0.014	0.008	0.005
Glutamine	0.000	-0.026	0.017	0.023	0.013
Glycine	-0.014	-0.042	0.014	0.028	0.016
Histidine	-0.007	-0.022	0.000	0.013	0.007
Isoleucine	0.017	0.000	0.043	0.023	0.013
Leucine	0.283	0.227	0.354	0.065	0.037
Lysine	0.000	-0.070	0.057	0.064	0.037
Methionine	-0.002	-0.007	0.007	0.008	0.005
Phenylalanine	-0.010	-0.022	0.007	0.015	0.009
Serine	-0.019	-0.043	-0.007	0.021	0.012
Threonine	-0.014	-0.042	0.007	0.025	0.015
Tryptophan	0.003	0.000	0.008	0.004	0.003
Tyrosine	-0.005	-0.008	0.000	0.004	0.003
Valine	-0.036	-0.137	0.058	0.097	0.056

Table 3.10   Amino acid and glucose flux (in mmol/gDW/hr) measured between 30 and 36 hours post synchronisation representing flux during the late trophozoite stage. The red to blue colour scale (in the first three columns) represents decreasing flux values. (n = 3)

	Average	Minimum	Maximum	SD	SEM
Alanine	-0.041	-0.082	-0.014	0.036	0.021
Arginine	-0.154	-0.504	0.403	0.488	0.282
Asparagine	-0.098	-0.183	0.042	0.123	0.071
Aspartic acid	0.000	-0.029	0.029	0.029	0.017
Glucose	3.239	2.164	4.475	1.163	0.672
Glutamic acid	-0.056	-0.070	-0.028	0.024	0.014
Glutamine	-0.396	-0.844	0.327	0.632	0.365
Glycine	-0.009	-0.126	0.112	0.119	0.069
Histidine	-0.066	-0.175	0.000	0.095	0.055
Isoleucine	-0.062	-0.157	0.100	0.140	0.081
Leucine	-0.137	-0.184	-0.057	0.070	0.040
Lysine	0.038	-0.165	0.235	0.200	0.116
Methionine	-0.028	-0.042	0.000	0.024	0.014
Phenylalanine	-0.035	-0.045	-0.030	0.009	0.005
Serine	-0.062	-0.115	0.043	0.091	0.052
Threonine	-0.038	-0.085	0.014	0.050	0.029
Tryptophan	0.010	-0.031	0.031	0.036	0.021
Tyrosine	-0.031	-0.061	0.000	0.031	0.018
Valine	-0.010	-0.029	0.014	0.022	0.013

Table 3.11   Amino acid and glucose flux (in mmol/gDW/hr) measured between 42 and 48 hours post synchronisation representing flux during the schizont stage. The red to blue colour scale (in the first three columns) represents decreasing flux values. (n = 3)

	Average	Minimum	Maximum	SD	SEM
Alanine	0.000	-0.027	0.014	0.024	0.014
Arginine	0.221	-0.115	0.576	0.346	0.200
Asparagine	0.070	-0.098	0.239	0.169	0.097
Aspartic acid	0.019	-0.014	0.086	0.058	0.033
Glucose	0.605	0.463	0.846	0.210	0.121
Glutamic acid	0.005	-0.070	0.084	0.077	0.044
Glutamine	0.356	-0.362	1.085	0.724	0.418
Glycine	0.037	-0.154	0.140	0.166	0.096
Histidine	0.066	-0.022	0.109	0.076	0.044
Isoleucine	0.081	-0.071	0.213	0.143	0.083
Leucine	0.047	-0.042	0.156	0.100	0.058
Lysine	0.072	-0.261	0.278	0.291	0.168
Methionine	0.028	0.000	0.042	0.024	0.014
Phenylalanine	0.025	0.000	0.045	0.023	0.013
Serine	0.062	-0.057	0.172	0.115	0.066
Threonine	0.028	-0.071	0.127	0.099	0.057
Tryptophan	0.010	-0.016	0.031	0.024	0.014
Tyrosine	-0.046	-0.244	0.076	0.173	0.100
Valine	0.024	-0.115	0.101	0.121	0.070

3.3.3   *In vitro glucose perturbation*

Inhibition of parasite growth has been shown at high media osmolarity. Parasites grown in 50 to 100 mM of sucrose or maltose brought about a reduction in host cell volume and eventual inhibition in parasite growth (Ginsburg *et al*, 1986a). Additionally, high media osmolarity has been demonstrated to inhibit development of trophozoites to schizonts (Dei-Cas *et al*, 1985). The highest mannose concentration (29.7 mM) was calculated to give 25% PfHT1 activity in the presence of 8.0 mM glucose, while the lowest concentration (1.1 mM) was estimated to result in 90% transporter activity. To eliminate the possibility of osmolarity as a factor affecting parasite growth (as opposed to limiting PfHT1 activity), controls were grown in media supplemented with sorbitol at similar concentrations as mannose, in addition to 8.0 mM glucose. Sorbitol, which has very low affinity for PfHT1 (K_i > 50 mM), is not metabolised by *P. falciparum* (Woodrow *et al*, 2000). Given that lower glucose concentrations were used in the experiment, controls grown in 11.1 mM glucose were also included. Reduced glucose concentration and high media osmolarity (8 mM glucose plus 29.7 mM sorbitol) showed no

significant difference in parasite growth compared with control cultures (in 11.1 mM glucose) (p-value = 0.956).

Figure 3.7 shows growth of asynchronous parasite cultures in the presence of mannose at different concentrations to give different levels of PfHT1 activity. Here, mannose and glucose competitively bind to PfHT1 for transport into the parasite (Woodrow *et al*, 2000). By varying the concentrations of mannose based on its inhibition constant (K_i) and the glucose Michaelis constant (K_m), the amount of glucose being transported into the parasite was controlled. As expected, there is a significant correlation between the calculated PfHT1 activity and percentage of parasite growth (Pearson's R = 0.942, p-value < 0.001). No significant effect was seen on parasite growth above 60% inhibition of glucose uptake, and the parasites were able to sustain reduced but significantly lower growth at lower glucose uptake rates. This is not surprising as hypoglycaemia is not uncommon in severe cases of malaria where plasma glucose levels can go below 2.2 mM (World Health Organization, 2015).

Given the same calculated percentage of PfHT1 activity, reduced glucose uptake was simulated in the model and biomass flux, as a surrogate for parasite growth, was calculated and compared with the *in vitro* values as part of model validation (Chapter 4).
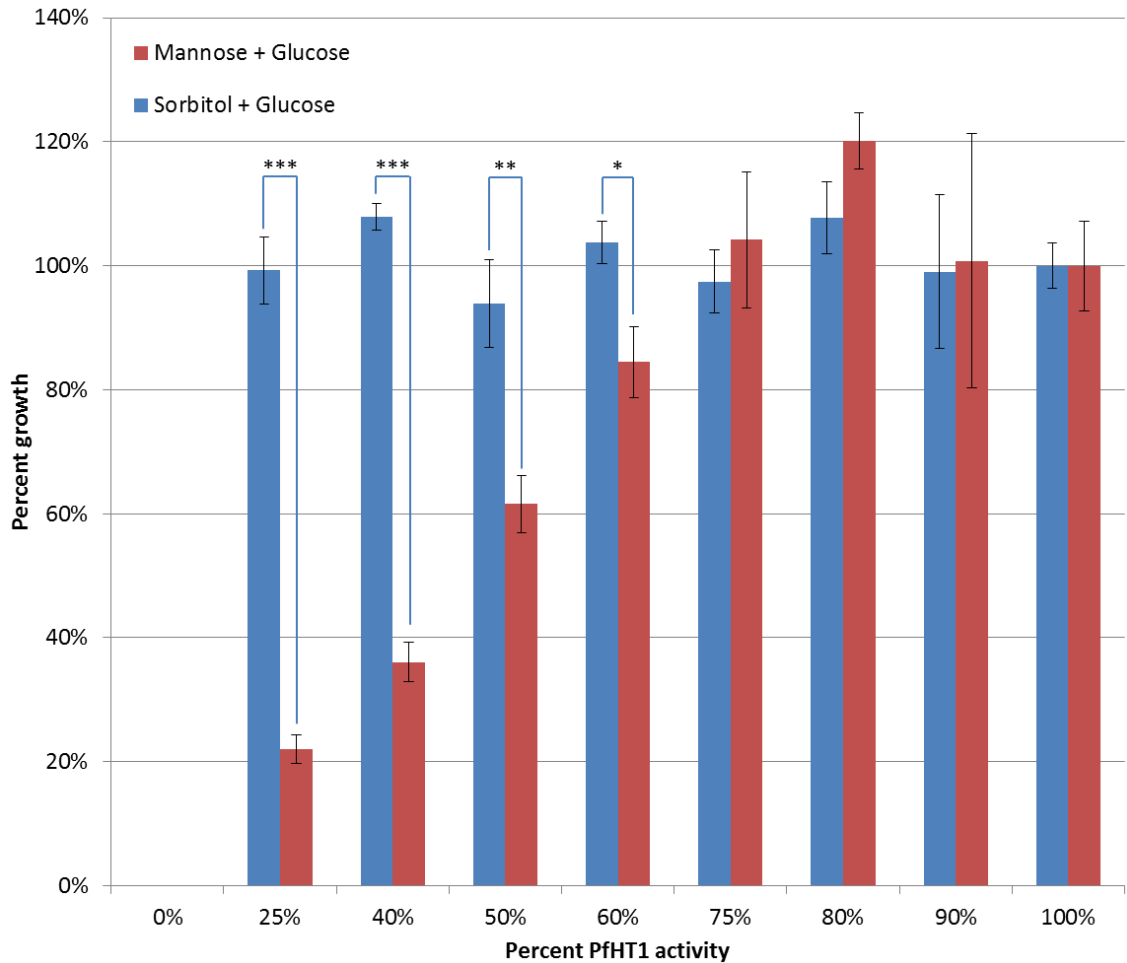
Figure 3.7   Parasite growth in the presence of mannose as a PfHT1 (glucose transporter) competitive inhibitor

This figure shows parasite growth (as a percentage of growth in the absence of an inhibitor) after 48 hours of incubation in increasing concentrations of mannose (red). Sorbitol was used as a control as it has very low affinity for PfHT1 (blue). SYBR green was used to measure the relative amount of parasites. Concentrations of mannose and glucose were calculated to achieve the desired percent inhibition of PfHT1 using K_m = 1.0 and K_i = 1.1 mM. (n ≥ 3)
*p-value = 0.0434, **p-value = 0.0087, ***p-value < 0.0001

3.4   Discussion

To further enhance the accuracy of the *in silico* metabolic model, *in vitro* flux data on 18 amino acids and glucose were measured through changes in metabolite concentrations in the media, and these values were later on incorporated as constraints in the model (Chapter 4). Spent media collection in this study utilised synchronised parasite cultures at 1% parasitaemia and 5% haematocrit, requiring the need to change the culture media every 24 hours to avoid undesirable pH conditions and accumulation of waste products (Jensen *et al*, 1983). This media change rendered the data between 24 to 30 hours unusable for the analysis of metabolite flux. Growing cultures at a lower parasitaemia and lower haematocrit may address this issue;

however, changes in metabolite concentrations might not be as detectable as in cultures with a higher parasitaemia and haematocrit. Another possible action would have been to conduct two spent media collections at 24 hours done before and after changing the media.

The trend in glucose flux observed over the life cycle of the parasite was consistent with RNA expression of genes involved in glycolysis (Bozdech *et al*, 2003) and parasite lactate dehydrogenase concentrations (Vivas *et al*, 2005). Moreover, overall average glucose flux was close to that measured using NMR quantification of glucose concentrations in spent media in *P. falciparum* 3D7 cultures (Forth, 2012). However, the data obtained had large calculated standard errors despite standardising the parasitaemia, haematocrit and culture conditions, and utilising blood within two weeks of collection from the blood bank. One possible reason for the error may be the variability in the erythrocytes utilised in the experiments. The age of the blood obtained from the NHSBT for use in parasite culture was unknown and the supply of O+ blood was only based on availability. At times, blood was obtained from other blood banks as the required blood type was not available in Leeds. Parasite growth rate and RBC invasion have been shown to decrease with increasing RBC storage time, though no significant effect on schizogony was observed (Goheen *et al*, 2016). RBC characteristics and responsive change due to storage also varies between donor to donor (DeWalski *et al*, 2015). In comparison with the rate of glucose consumption (4.58 ± 1.5 μmol per billion RBCs per 24 hours) reported by Jensen *et al* (1983), the average glucose flux in uninfected red blood cells (recalculated to obtain a similar unit as the reported value) observed in the spent media experiments was lower (2.29 ± 0.39 μmol per billion RBCs per 24 hours). This may be attributed to differences in RBC storage as storage duration has been implicated to lower glycolytic rates in red blood cells (Gevi *et al*, 2012). In future work, biopreservation of red blood cells (i.e., storage of fresh RBCs in serum and freezing media in liquid nitrogen) may be done so that blood from the same donor may be used for different experimental replicates, reducing RBC variability. A study comparing fresh and biopreserved red blood cells has shown no significant difference in parasite growth rates and merozoite invasion (Goheen *et al*, 2016).

Initially, optimisation of amino acid detection through HPLC (specifically to test the derivatisation protocol) was done using a larger column, Acclaim 120 C18 (Dionex) 150 mm x 4.6 mm column (5 μm particle size, 120 Å pore size), which was eventually replaced with a new smaller column with smaller particle size (see Methodology for column properties and dimensions). The replacement of the column significantly reduced the consumption of eluents (from 2.0 ml/min to 0.2 ml/min). This meant that a single 2-litre preparation of eluent can be used for running more samples, reducing minor variability in eluent properties (e.g., pH, concentration). A pH shift of 0.1 can result in a 10% shift in the retention time, while a 1%

difference in organic solvent concentration can result in a 5 to 15% shift in retention time (Neue, 2002). The long running time (70.5 minutes) per sample posed as a limitation and running all spent media samples would have been quite time consuming thus only selected samples were run for amino acid concentration determination. UHPLC separation using a column with smaller particle size can easily reduce running time to 10 minutes (Steiner *et al*, 2009). In addition to amino acids, *in vitro* flux measurements of other important boundary metabolites such as lactate and hypoxanthine may also contribute to enhancing the predictive capacity of the model.

A variety of 6-carbon monosaccharides, including deoxy- forms and sugar analogues, can act as ligands for the *Plasmodium falciparum* hexose transporter 1 (Woodrow *et al*, 2000). Glucose and mannose both have relatively high affinity for the transporter compared with other hexoses such as fructose and galactose (Blume *et al*, 2011). Here mannose was used against glucose as a competitive inhibitor for PfHT1. Mannose is a cheap compound that can be used in conducting inhibition experiments on PfHT1; however, a limitation of using mannose is that higher concentrations are required to achieve greater inhibitions because mannose has a K_i close to the K_m for glucose (Woodrow *et al*, 2000). Nevertheless, it has been demonstrated here that parasite transporter activity can be reduced to as low 25%, showing significant reduction in parasite growth. Higher concentrations of mannose may be used to achieve greater reduction in glucose transporter activity, but this has to be done alongside controls to ensure that the hyperosmolar culture media does not interfere with normal parasite growth. It is important to note that RPMI has an osmolality between 270 and 310 milliosmoles/kg and a specific gravity of 1.006 kg/l (Thermo Fisher) and early studies have shown no significant parasite growth changes in osmolarities of up to 328 milliosmoles/l (Dei-Cas *et al*, 1985). This gives a small margin of about 16 to 56 milliosmoles/l for the addition of mannose. Otherwise, greater PfHT1 inhibition may be achieved by using compound 3361, a glucose analogue which has a much lower K_i (0.053 mM) compared with mannose (1.1 mM) (Joët & Krishna, 2004). Other glucose analogues that have been utilised for glucose transporter studies in malaria include 2-deoxy-D-glucose and 3-O-methyl-D-glucose (Krishna *et al*, 2000; Woodrow *et al*, 2000).

Chapter 4   Model predictions and validation

4.1   Introduction

Constraint-based modelling utilises different boundaries, conditions and physical laws that limit the solution space within which the organism's phenotype can be best fitted and represented (Price *et al*, 2004; Ravikrishnan & Raman, 2015). The solution space represents a set of combinations of reaction fluxes within the allowable constraints (Orth *et al*, 2010). With any given unconstrained model, there are multiple possible combinations of reaction fluxes that can result in a feasible solution for the objective function. Take for example a simple model composed only of three reactions (i.e., R1, R2 and R3) represented as a three dimensional plot, where each axis represents the flux for each reaction (Figure 4.1). Without any constraint on all three reactions (A), the solution space is the infinite space in this three-dimensional plot and every point in this three dimensional space is a possible solution; however, putting boundaries on each axis (e.g., flux constraints, shown in blue broken lines), the solution space (in red) becomes smaller (B). With additional constraints (e.g., reaction, mass balance and thermodynamics), the solution space becomes even smaller (C). Through flux balance analysis, a given genome scale metabolic model can be used to find an optimal solution for a given objective (shown in green), such as the production of biomass (represented as R1). The green broken lines show the corresponding flux values for the reactions that result in the maximisation of the flux for the objective function.

Reactions in the model must have a stoichiometry that takes into consideration the law of conservation of mass. In addition, information on reaction thermodynamics may help identify the direction of a given reaction, i.e., reversible or irreversible, which constrain the model further. As gene expression data have become more widely available, several models have utilised these data to provide a more accurate representation of the organism being modelled (Lee *et al*, 2012; Navid, 2011). There are several ways of incorporating gene expression data into a metabolic model. One can utilise a function that turns on or off reactions associated with genes based on their gene expression, or adjust the reaction flux in proportion with the corresponding gene expression (Reed, 2012).
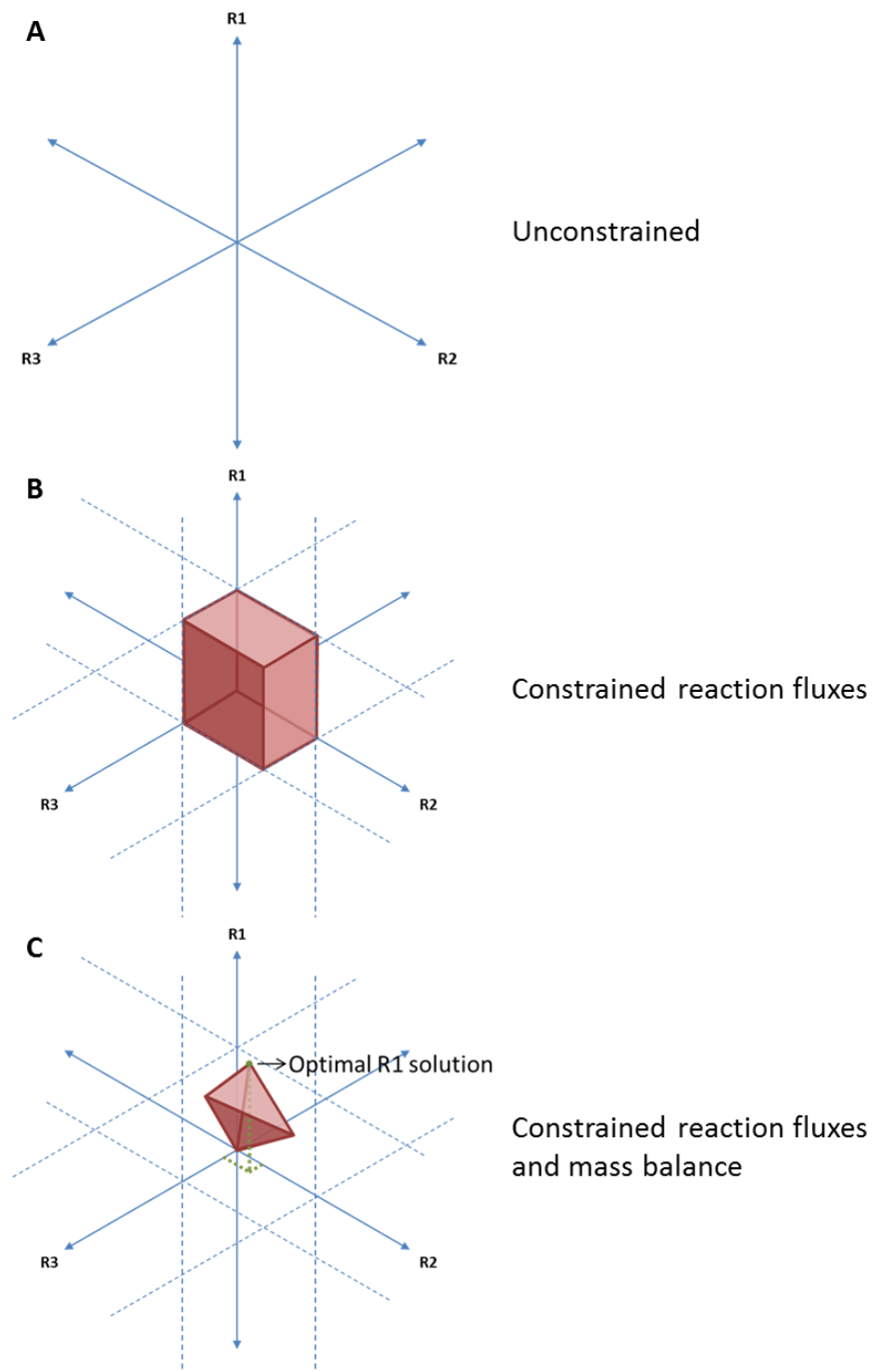
Figure 4.1   Visualisation of flux balance analysis solution space
A three reaction model is represented here by a three-dimensional plot, with the three axes representing the flux for the respective reactions (R1, R2 and R3). R1 is the reaction that represents the objective function. (A) Without constraining the reaction fluxes, there are an infinite number of possible combinations of reaction fluxes that will result in a solution. (B) By adding constraints in the reaction fluxes (blue broken lines), the solution space becomes smaller, shown in red, where each point in the solution space is a possible solution. The solution space becomes even smaller with the addition of reaction stoichiometry/mass balance and thermodynamics (C). Flux balance analysis, through linear programming, can be used to calculate the flux for each reaction that will result in maximum flux for the objective function, in this case shown as the green dot. The green broken lines show the corresponding flux values for the reactions that result in the maximisation of the objective function.

Using stage specific gene expression data, the Plata malaria model proportionally adjusted the flux of reactions based on the relative expression of the corresponding associated genes. They then compared the predicted shifts in external concentrations of 33 metabolites between stages against experimental data, showing significant correlations (Plata *et al*, 2010). On the other hand, the Huthmacher malaria model utilised five different gene expression data sets and compared the flux direction of a number of external metabolites and compared these against published experimental data. They were able to observe that predictions of model constrained using gene expression data were more in agreement with experimental data than those from the unconstrained model (Huthmacher *et al*, 2010). A similar approach of adding model constraints can be done using derived metabolite flux. The group of Wallqvist *et al* (2016) integrated the malaria metabolic model into the RBC metabolic model and incorporated strain specific gene expression data along with metabolite flux data from uninfected RBCs (experimental) and flux values in co-infected RBCs, i.e., uninfected RBCs cultured together with infected RBCs (predicted). Using these constraints, the metabolite flux values for infected RBCs were calculated through flux balance analysis, which showed 16 out of 24 metabolite fluxes having significant correlation with experimental data. These methods of comparing predictions against experimental data are part of model validation, which are often used as a measure of model quality and a means to ensure the accuracy of model predictions (Kim *et al*, 2012).

As mentioned in the first chapter, metabolic models are often utilised to identify potential drug targets through flux balance analysis (Chavali *et al*, 2008; Plata *et al*, 2010; Huthmacher *et al*, 2010; Li *et al*, 2011; Navid, 2011). Gene or reaction knockouts can be simulated and the effect on the biomass objective can be evaluated to identify whether the said knockout resulted in a lethal phenotype or limited growth. *In silico*, lethal knockouts result in zero objective flux while growth-limiting knockouts result in sub-optimal objective flux. In the case of the predictions using the Plata model, knockouts were divided into four categories: lethal, growth reducing (i.e., resultant biomass flux between 0 and 95% of wild-type biomass flux), slightly growth reducing (i.e., resultant biomass flux between 95% and 100% of wild-type biomass flux) and non-lethal (Plata *et al*, 2010). In the *Leishmania major* model, a cut off of 90% was used to define growth-limiting knockouts (Chavali *et al*, 2008). Several software can be used to simulate gene or reaction knockouts. MetExplore (Cottret *et al*, 2010) is a free online resource that allows users to upload SBML models for evaluation and simulations. SurreyFBA (Gevorgyan *et al*, 2011) and OptFlux (Rocha *et al*, 2010) are free downloadable software with user friendly graphical interface that can also perform a number of important commands for model analysis and simulations. COBRApy (Ebrahim *et al*, 2013) is a Python-based software that has similar functionalities as the previously mentioned software.

Predicted gene knockouts in comparison with experimental data have been used for model validation. Proponents of the metabolic models of *S. cerevisiae* (Duarte *et al*, 2004), *E. coli* (Latendresse *et al*, 2012), *L. major* (Chavali *et al*, 2008) and *P. falciparum* (Plata *et al*, 2010; Huthmacher *et al*, 2010) have compared their predicted essential genes/reactions with published data on experimentally validated essential genes/reactions. Furthermore, predicted targets have been experimentally validated with the use of inhibitors in *in vitro* cultures. Plata *et al* identified nicotinate mononucleotide adenylyltransferase as a potential drug target and, with the use of a small molecule inhibitor (compound 1_03), they were able to demonstrate *P. falciparum* growth inhibition (MIC_{50} = 50 μM). A few FDA-approved drugs were tested *in vitro* on *Leishmania major* to target those that were predicted by their model to be essential. They have shown that halofantrine, an antimalarial drug, was able to inhibit growth with an IC_{50} of 9.5 μM, demonstrating the potential of drug repurposing as a means of finding new anti-infectives/antiparasitics (Chavali *et al*, 2012a). They have used online databases on compound-protein interactions for the identification of potential inhibitors. One of these databases is DrugBank (Law *et al*, 2014), which contains detailed information on more than 8,000 compounds and their interactions with protein targets. The database contains FDA-approved drugs as well as those that are experimental, investigational and withdrawn. Important information on the type of drug-target interaction is also included in the database (e.g., inhibitors, cofactors, substrate, product, etc.). The DrugBank website offers drug search by name, structure, chemical formula or molecular weight, while protein targets can be searched using their name or protein sequence. When conducting a protein target search, DrugBank returns a list of aligned protein sequences along with sequence alignments with matched proteins, their corresponding alignment scores and interacting compounds. Details and external database links on the matched target and the interacting drugs are also provided.

In this chapter, experimentally measured stage specific metabolite flux values from the previous chapter were incorporated in the model developed in the second chapter. Given these constraints, *in silico* perturbations similar to those done *in vitro* in the previous chapter were simulated and the results were compared with each other. Single gene knockout simulations were also done in order to predict essential genes. These predicted essential genes were compared against published data on experimentally validated essential genes. Finally, repurposed compounds identified using the DrugBank database were tested for *in vitro* activity against selected novel targets as part of proof-of-concept model validation.

4.2 Methodology

4.2.1 In silico glucose perturbation

To mimic *in vitro* consumption of nucleoside/bases from the media, the boundary flux of adenine, adenosine, guanine, inosine and xanthine were set to zero as these are not present in Albumax II supplemented malaria complete media. This also allows the model to consume hypoxanthine for purine metabolism from the external environment. COBRApy (Ebrahim *et al*, 2013) was used to add the experimentally obtained flux values from Chapter 3 into the model developed in Chapter 2. The upper and lower bound flux for 18 amino acid boundary reactions as well as the glucose boundary reaction were constrained using the stage specific *in vitro* (1) maximum and minimum flux (i.e., range), (2) average flux ± SD and (3) average flux ± SEM, respectively. With these set constraints, the maximum biomass reaction flux or the optimal solution was calculated using FBA. For consistency, the constraint value (i.e., range, SD or SEM) that yielded feasible models (i.e., those with optimal solution) for all parasite stages was used for generating stage specific models. A stage specific model is defined as a model with boundary constraints set using the experimentally derived stage specific flux values. It is important to note that with the constraint set on glucose flux, the predicted lactate influx was contrary to what was observed *in vitro* (Jensen *et al*, 1983; Elliott *et al*, 2001). Thus, constraint on the lactate boundary reaction, which was based on previously measured lactate efflux (-1.644 mmol/gDW/hr) on *P. falciparum* 3D7 spent culture media was added into the model (Forth, 2012). Similar glucose perturbation simulation was done on the unconstrained model for comparison with the constrained models.

For each stage specific model, glucose flux was perturbed to simulate inhibition of the parasite glucose transporter (PfHT1). This was done by adjusting the upper bound flux of the glucose boundary reaction (Glucose_btoc_s), to represent 0 to 100% transporter activity. Using COBRApy, the maximum biomass solution at different percentages of glucose transporter activity was determined through FBA. These were plotted using GraphPad PRISM 7, while Pearson's correlation and the corresponding p-values between the *in vitro* and *in silico* data were calculated using SciPy (Jones *et al*, 2015). The stage specific model with the best correlation with the *in vitro* data (i.e., schizont stage model, explained in detail in the Results section) was used for subsequent simulations.

4.2.2 Comparison of flux values with gene expression data

Reaction flux values were compared against gene expression data to identify any possible correlation between the two data sets. The absolute value of the flux was used in the

comparison instead of the actual value to factor out the direction of the reaction while focussing on the rate of metabolite consumption of the reaction. Using the three stage specific models, reaction flux values that result in maximum biomass flux were calculated to represent reaction fluxes during the ring, trophozoite and schizont stages. These reaction fluxes were then mapped against their associated genes for comparison with gene expression data. In cases where a given gene is associated with multiple reactions, the average flux of all associated reactions was calculated and assigned to the said gene. Only genes in the model that were associated with reactions with non-constant flux values were considered to avoid error in the calculation of the Pearson's correlation coefficient (r-value). Gene expression data (in reads per kilobase of exon per million mapped reads) used in this comparison was obtained from the RNA-seq data of Bartfai *et al* (2010) where they measured gene expression at 5 hour intervals over 40 hours (post invasion) in *P. falciparum* 3D7 *in vitro* cultures (from T5 to T40). Gene expression at T5, T30 and T40 hours post infection were compared against ring, trophozoite and schizont stage flux, respectively. These time points were selected as they were closest to the initial time points of spent media collection for the measurement of *in vitro* metabolite flux (i.e., 6, 30, and 42 hours post synchronisation), as explained in Chapter 3.

For each gene in the model, Pearson's correlation was calculated between the gene-associated reaction flux and the gene expression data, and the average correlation for all genes was determined (actual average r-value). An empirical p-value for the average correlation was calculated in two ways:

1. For each gene, the order of the flux (i.e., ring, trophozoite and schizont flux values) were randomly shuffled and correlation with gene expression data was calculated before getting the average correlation for all genes.

2. Random gene expression values (within the minimum and maximum expression values for a given gene) were compared against the flux values in correct order and the correlation was calculated. The average correlation for all genes was then determined.

Ten thousand randomisations were done to obtain 10,000 average Pearson's correlation coefficients (random average r-values). The empirical p-value was calculated by counting the total number of random average r-values that are greater than or equal to the actual average r-value, and dividing the count by the total number of random average r-values. This gives the probability of observing an average r-value greater than or equal to the actual average r-value by chance. Two p-values will therefore be reported here for each average r-value, presented in the same order as the enumerated methods of empirical p-value calculations mentioned above.

Considering published data on correlation between peak gene expression and protein abundance (Foth *et al*, 2011) which reported a delay in peak protein abundance of 11 hours after peak gene expression, correlation between predicted reaction flux and "time-shifted" gene expression values were also tested. Predicted reaction flux values at the ring, trophozoite and schizont stages were compared against gene expression at T40 (i.e., corresponding to the gene expression roughly before the beginning of the next asexual cycle), T20 and T30 hours post infection, respectively. The average r-value and the corresponding empirical p-values were then calculated.

4.2.3   Identification of essential genes

Single gene knockouts were simulated to identify essential genes in the schizont stage model using COBRApy (Ebrahim *et al*, 2013). Gene knockouts that resulted in a zero biomass solution were considered as lethal knockouts; while those that resulted in a lowered biomass solution were considered growth-limiting knockouts. Different thresholds for defining limited growth were tested (explained in detail in section 4.3.3); and ultimately, biomass solution that is less than 95% of the optimal solution was considered as the threshold. After single gene knockout analysis, the predicted set of essential genes was matched against a set of published data on experimentally validated essential genes/reactions in *Plasmodium*. Published literature were reviewed for growth assays using compounds with known enzyme targets or gene knockouts/conditional knockouts done *in vitro* or *in vivo* on *Plasmodium* spp., while inhibition assays on purified parasite enzyme were excluded. When available, identified essential gene IDs and the EC classification number of associated reaction/s were noted together with the publication reference. Additional known targets were obtained from TDR Targets (Magariños *et al*, 2012), a database that holds information on malaria and neglected tropical diseases that are important for drug target identification. In cases where only the gene ID or the EC number was mentioned in the publication, the corresponding EC classification number or gene ID were matched using PlasmoDB data (Aurrecoechea *et al*, 2009). Also, gene IDs in the old format were updated to the new format using PlasmoDB data. A total of 128 genes make up the original gold standard list (see Appendix).

The *Plasmodium* Genetic Modification (*Plasmo*GEM) is a database that holds vector sequences that can be used to modify the genome of *Plasmodium berghei* (Schwach *et al*, 2015; Gomes *et al*, 2015). The database was developed as part of the initiative of the Malaria Programme at the Wellcome Trust Sanger Institute. In May 2016, the database was updated to include prepublication phenotypic data on more than 2,000 *P. berghei* genes (www.plasmogem.sanger.ac.uk). Updated *P. berghei* gene IDs along with phenotypic data were

downloaded from the *Plasmo*GEM database and orthologues of these genes in *P. falciparum* were identified through PlasmoDB (Aurrecoechea *et al*, 2009). For *P. falciparum* genes that are orthologous to more than one *P. berghei* gene, only those with consistent gene essentiality information were noted (e.g., all orthologous *P. berghei* genes must be essential). A total of 1,629 genes associated with lethal or slow growing phenotypes were obtained from *Plasmo*GEM, 1,572 of which were not in the original gold standard list. This gave an updated gold standard list consisting of 1,700 genes.

As part of model validation, the proportion of genes predicted through FBA that were in the gold standard list (i.e., true positive predictions) were compared against the proportion of genes in the model that were in the gold standard list; and the corresponding enrichment score and hypergeometric p-value were calculated. The enrichment score is simply the proportion of gold standard genes in the predicted essential gene set divided by the proportion of gold standard genes in the whole model. The p-value was calculated through hypergeometric testing to negate the possibility that the enrichment of true positive predictions in the set of predicted genes was a result of chance (Huang *et al*, 2009). Simply put, the hypergeometric p-value is the probability of single gene knockout analysis randomly identifying n or more true positives in the predicted essential gene set. This was done using the hypergeometric function in SciPy (Jones *et al*, 2015). These calculations were done for lethal and growth-limiting gene knockouts together and separately. Similarly, single gene knockouts were done using the unconstrained model for comparison.

The list of essential genes predicted using the Plata model was obtained from the supplemental table of single gene deletions in the corresponding publication (Plata *et al*, 2010). For consistency, only the genes in the Plata model whose knockouts resulted in < 95% biomass solution were compared with the essential gene predictions using the iFT342 model (as Plata *et al* also reported gene knockouts that resulted in biomass solution between 95% and 100%). Essential genes predicted by the Huthmacher model were obtained from the supplementary file that contained the ranked predicted essential reactions (12918_2009_509_MOESM18_ESM.txt) which listed the gene IDs associated with the essential reactions. Additionally, the authors presented in the publication predicted essential genes that coincided with their gold standard list (Huthmacher *et al*, 2010). Essential genes predicted by the Forth model were taken from Appendix VII of the final thesis manuscript (Forth, 2012). Similarly, gene knockouts that resulted in ≥ 95% biomass solution in the Forth model were excluded from the analysis. A Venn diagram was generated using the jvenn online resource (Bardou *et al*, 2014). Positive predictive value (PPV) and true positive rate (TPR) were calculated using the formulas shown in Equation 4.1 and Equation 4.2, respectively.

$$PPV = \frac{number\ of\ true\ positives}{number\ of\ predicted\ essential\ genes} \times 100\% \qquad \textbf{Equation 4.1}$$

$$TPR = \frac{number\ of\ true\ positives}{total\ true\ positives\ in\ the\ whole\ model} \times 100\% \qquad \textbf{Equation 4.2}$$

4.2.4   Validation of novel targets

Protein sequences of all 342 genes in the iFT342 model were obtained from PlasmoDB (Aurrecoechea *et al*, 2009) and were used for sequence alignment against proteins in the DrugBank database (Law *et al*, 2014). For genes in the model that aligned with proteins in the said database, information on the corresponding e-value of the alignment as well as the compounds that interact with the aligned protein were collected. These compounds include agonists, cofactors, inhibitors and substrates. Inhibitors of predicted novel essential genes (i.e., those predicted by the model to be essential but are not in the gold standard list) were selected and tested *in vitro*.

Stock solutions of cladribine (200 mM, Cambridge Bioscience), gemcitabine (20 mM, Sigma), ritodrine (20 mM, Sigma), rosiglitazone (20 mM, Cambridge Bioscience) and DSM265 (10 mM) (produced by the Chemistry Department, University of Leeds) were prepared in dimethyl sulfoxide (DMSO, Sigma). For the dose response assay, unsynchronised *P. falciparum* 3D7 (n = 3 biological replicates) at 0.5% parasitaemia and 3% haematocrit were grown in black-sided, clear flat-bottomed 96-well plates (Costar) in malaria complete media (as detailed in Chapter 3) with different concentrations (from 1.3 nM to 20 μM) of cladribine, gemcitabine and ritodrine. Rosiglitazone at similar concentrations was used as negative control. DSM265 with concentrations from 0.0128 nM to 1 μM was used as positive control. Uninfected red blood cells were also grown in malaria complete media with DMSO (at a concentration similar to those incubated in the highest drug concentration) as additional control. The cultures were incubated in 1% oxygen, 3% carbon dioxide and 96% nitrogen in a humidified sealed chamber at 37°C. The chamber was gassed with the same gas mixture after 24 hours. After 48 hours of incubation, the relative amounts of live parasites were quantified using SYBR Green as described in Chapter 3 (Smilkstein *et al*, 2004). Background fluorescence of uninfected red blood cells was subtracted prior to comparison with control cultures. Growth response was presented as a proportion of the control cultures grown in the absence of the drug. The dose response plots were done using GraphPad Prism 7. For those that showed growth inhibition, dose response was recalculated by subtracting the background fluorescence of the culture grown in the highest drug concentration (parasites left over after 48 hours of incubation) from each sample before replotting the dose

response curve separately (per replicate). The IC_{50} was calculated for each plot and the average value was reported alongside the standard error.

To further describe the effect of a compound that demonstrated inhibition in the dose response assay, highly synchronised *P. falciparum* 3D7 at 0.5% parasitaemia and 3% haematocrit were grown in malaria complete media in 25 cm^2 tissue culture flasks at 10 times the determined IC_{50} of the drug. Blood smears stained with 10% Giemsa in Sorensen's buffer were taken at 24 and 48 hours post incubation with the drug and the number of parasites at different stages (i.e., rings, late trophozoites and schizonts) was quantified by light microscopy.

4.3 Results

4.3.1 In silico glucose perturbation

The range and the average *in vitro* flux along with the standard deviation and standard error of the mean were used separately as flux constraints for the respective boundary metabolite in the model, and the maximum biomass flux was calculated using flux balance analysis (Table 4.1). In some cases, the constraints resulted in an infeasible model, and no biomass solution was obtained. Only the average ± SD as boundary metabolite flux constraints was able to yield feasible solutions for all three stages and thus these values were utilised in the subsequent simulations. Comparing the biomass flux obtained using the standard deviation of the boundary metabolite flux as constraints, the schizont model was able to generate biomass at a rate that is almost ten times that in the late trophozoite model and even more than in the ring stage model. Taking into consideration the experimental metabolite flux constraints added into the model, it can be suggested that ring and trophozoite stage constraints permit but do not maximise biomass production as much as the schizont stage constraints. Although the parasite is most metabolically active during the trophozoite stage (Kirk & Lehane, 2014), biomass production may be more evident during the schizont stage when nuclear divisions and development of daughter merozoites occur (Gerald *et al*, 2011).

Table 4.1   Maximum biomass flux (in mmol/gDW/hr) obtained using stage specific *in vitro* flux values as model constraints

Constraint	Rings	Late trophozoites	Schizonts
Range	No solution	3.10×10^{-2}	3.77×10^{-1}
SD	1.68×10^{-2}	4.17×10^{-2}	3.86×10^{-1}
SEM	No solution	No solution	No solution

Using the three stage specific models (with the average *in vitro* metabolite boundary flux values ± SD as model constraints), the maximum biomass flux values at different percentages of glucose transport flux were calculated using FBA and were compared against *in vitro* data (Figure

4.2). As expected, for each stage specific model there was increasing biomass flux with increasing glucose flux; however, at one point, further increase in glucose flux does not contribute in a further increase in biomass flux. This may be explained by looking at the contribution of some metabolites to the biomass reaction, specifically those not produced downstream to or branching out from the glycolytic pathway (i.e., non-glucose dependent biomass reactants). Constraints on these metabolites tend to restrict the biomass production independently from the constraints on the availability of glucose. The late trophozoite stage model predicted maximum biomass flux even at 25% glucose transport flux, while at least 75% glucose influx was necessary for the ring stage model to achieve maximum biomass flux. Both the ring and late trophozoite stage specific models showed abrupt shift in biomass flux from 0 to 100% at 25% and 75% glucose transport flux, respectively. In contrast, the schizont stage model displayed a more gradual increase in predicted biomass production with increasing glucose transport flux. For the schizont model, at least 80% glucose transport flux was required to give 100% biomass production. With the restrictions on the amino acid boundary flux, non-glucose dependent biomass reactants such as pyrimidines are also restricted as they are dependent on available aspartate and glutamine. Looking at the individual trends in amino acid flux with increasing influx of glucose (Figure 4.3), influx of aspartate and glutamine are always maximised to satisfy the model's need for these amino acids as well as pyrimidines; however at a certain point, the production of glutamate (which is a by-product of pyrimidine synthesis) surpasses the model's requirement and thus resulting in a shift from influx to efflux of glutamate. Alanine and leucine are the two most abundant amino acids in haemoglobin, accounting for 25% of the total amino acid composition. Alanine is not present in the media and is not very abundant in the parasite proteome, thus there is constant efflux of the said amino acid into the media. On the other hand, leucine is the fourth most abundant amino acid in the parasite proteome. At lower glucose flux, haemoglobin digestion is capable of supplying enough leucine to satisfy the biomass reaction stoichiometry, thus there is no influx of the amino acid from the media. However, overall amino acid efflux constraints limit haemoglobin digestion. So despite this amino acid being abundant in haemoglobin, leucine from the media becomes necessary with increasing glucose influx.

The late trophozoite stage model predictions showed no correlation with *in vitro* growth (r = 0.623, p-value = 0.073), while the mid-ring stage model predictions had a significant correlation (r = 0.820, p-value = 0.007). Of the three stage specific models, the late schizont stage model predictions were the best correlated with the *in vitro* data (r = 0.985, p-value = 1.36 x 10^{-6}). Additionally, the percentage of PfHT1 activity that resulted in 50% parasite growth *in vitro* (49.7%) was quite similar to what was predicted by the schizont model (50.7%), as shown in

Table 4.2. Restricting glucose flux in the unconstrained model showed similar results as the late trophozoite model, having a non-significant correlation with the *in vitro* data (r = 0.626, p-value = 0.07).
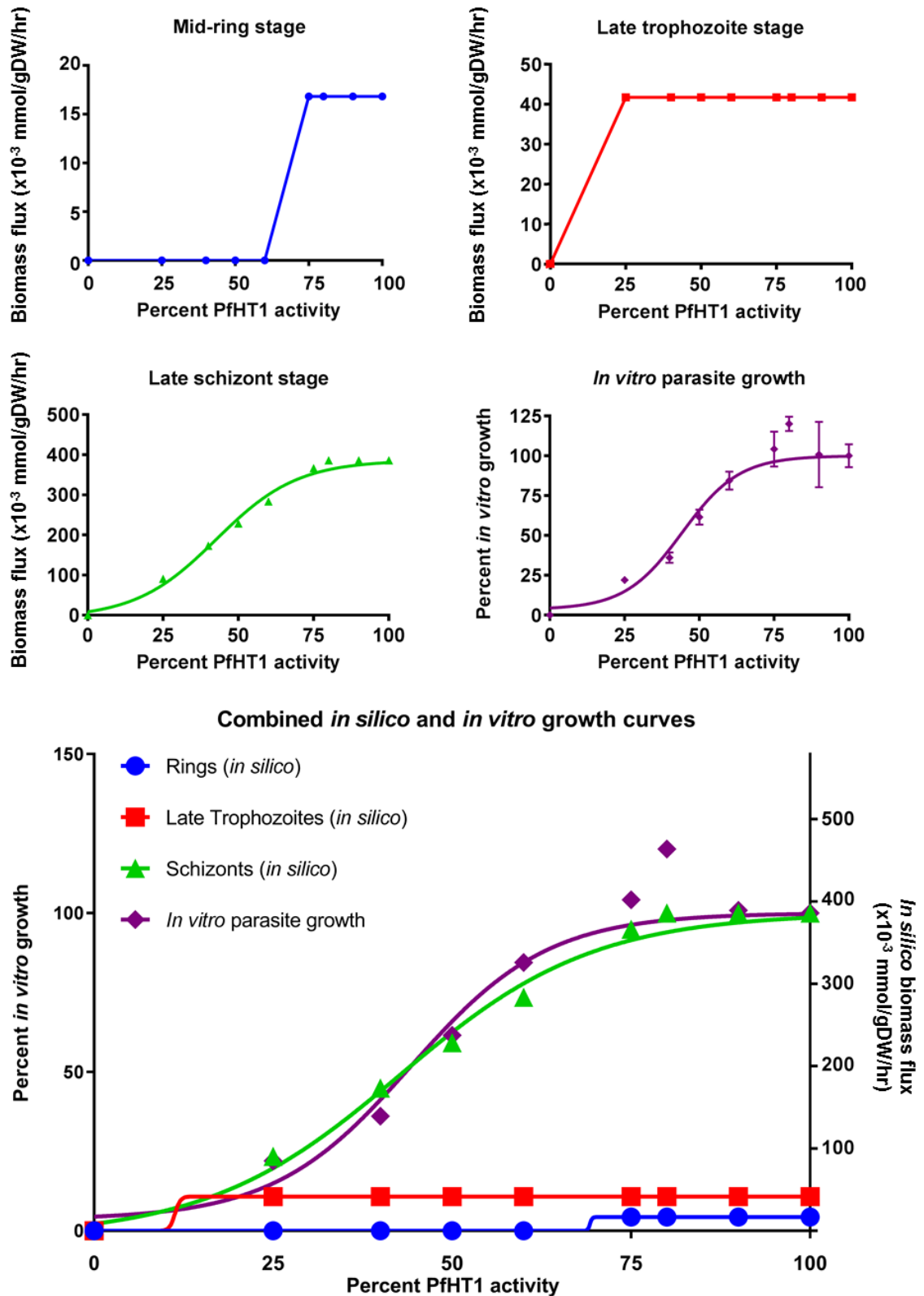
Figure 4.2   *In vitro/in silico* glucose transporter inhibition
The blue, red and green curves represent the predicted biomass flux at different percentages of glucose transporter flux in the mid-ring, late trophozoite and late schizont stage models, respectively. The purple curve (*in vitro* data) represents the relative amounts of parasite (± SEM) as a percentage of the amount of parasites in the control after 48 hours of incubation in different concentrations of mannose, which inhibits glucose transport via PfHT1 (n ≥ 3 biological replicates).
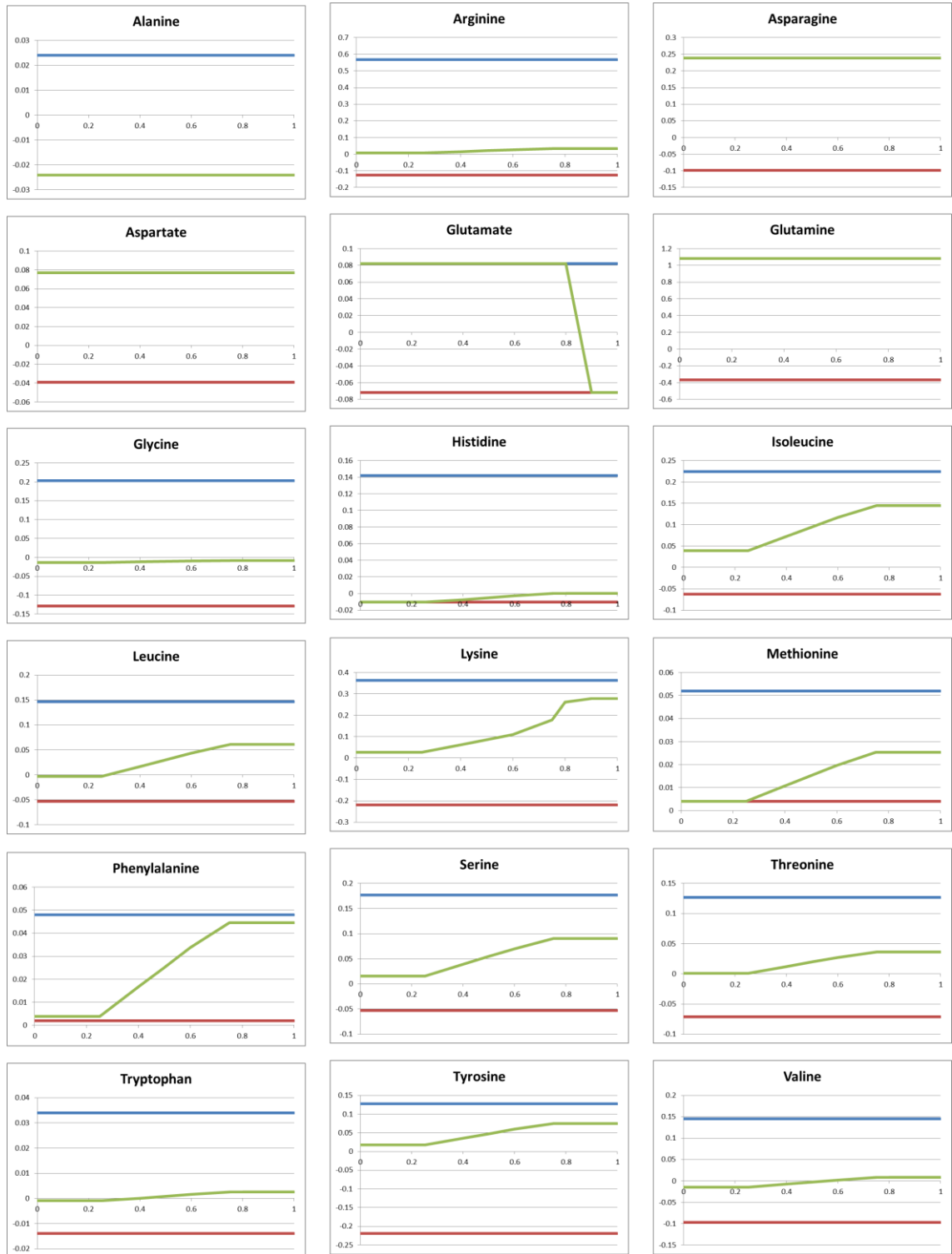
Figure 4.3   Flux of amino acid boundary reactions necessary to maximise biomass production with increasing glucose boundary flux in the schizont stage model

This figure shows the flux of 18 constrained amino acid boundary reactions at different percentages of glucose boundary flux. The y-axes represent the amino acid boundary flux (in green) that results in maximum biomass solution (in mmol/gDW/hr) while the x-axes represent the proportion of glucose boundary flux with respect to the glucose upper bound constraint. The blue and red lines represent the corresponding upper and lower bound constraints for the boundary reaction, respectively.

Table 4.2 **Comparison of *in vitro* percent growth at different levels of glucose transporter activity against stage specific *in silico* model predictions**

PfHT1 activity	*In vitro*	*In silico* (mmol/gDW/hr)		
		Mid-ring	**Late trophozoite**	**Late schizont**
0.0%	0.0%	0.00 (0.0%)	0.00 (0.0%)	0.00 (0.0%)
25.0%	22.5%	0.00 (0.0%)	4.17×10^{-2} (100.0%)	9.06×10^{-2} (23.5%)
40.0%	37.1%	0.00 (0.0%)	4.17×10^{-2} (100.0%)	1.73×10^{-1} (44.8%)
50.0%	62.7%	0.00 (0.0%)	4.17×10^{-2} (100.0%)	2.28×10^{-1} (59.1%)
60.0%	85.1%	0.00 (0.0%)	4.17×10^{-2} (100.0%)	2.84×10^{-1} (73.6%)
75.0%	104.7%	1.68×10^{-2} (100.0%)	4.17×10^{-2} (100.0%)	3.66×10^{-1} (94.8%)
80.0%	121.6%	1.68×10^{-2} (100.0%)	4.17×10^{-2} (100.0%)	3.86×10^{-1} (100.0%)
90.0%	107.1%	1.68×10^{-2} (100.0%)	4.17×10^{-2} (100.0%)	3.86×10^{-1} (100.0%)
100.0%	100.0%	1.68×10^{-2} (100.0%)	4.17×10^{-2} (100.0%)	3.86×10^{-1} (100.0%)
IC_{50}[*]	49.7%	68.8%	67.1%	50.7%
Pearson's correlation		0.82	0.62	0.99
p-value		0.007	0.07	1×10^{-6}

[*]The IC_{50} presented here is the percentage of PfHT1 activity that resulted in 50% inhibition of *in silico* or *in vitro* growth.

4.3.2 Comparison of flux values with gene expression data

Out of 342 genes in the model, only 185 were associated with reactions with non-constant flux values. Comparing the predicted associated flux values of these genes with gene expression data gave an average Pearson's correlation coefficient of 0.100 (empirical p-values = 0.03, 0.03), suggesting very weak overall correlation between gene expression and flux values. The r-values for individual reactions ranged from -0.998 to 1.000, with a median of 0.164 and a standard deviation of 0.724. Comparing the "time-shifted" gene expression values with the predicted flux gave an average r-value of -0.114 (empirical p-values = 0.99, 0.98).

4.3.3 Identification of essential genes

As mentioned earlier, predicted essential genes were divided into two groups: (1) those associated with knockouts that resulted in a lethal phenotype (biomass reaction flux = 0 mmol/gDW/hr); and (2) those whose knockouts resulted in reduced growth. Although lower thresholds produced significant enrichment of gold standard genes in the predicted set of essential genes, reduced growth was set at 95% of the maximum biomass reaction flux as lower thresholds resulted in fewer predicted essential genes and higher hypergeometric p-values for enrichment of gold standard genes in the predicted gene set (Table 4.3).

Table 4.3   Total number of predicted essential genes and true positive predictions using different thresholds for defining limited *in silico* growth (threshold x maximum biomass reaction flux)

Threshold	Total number of predicted essential genes	Number of true positives (in gold standard list)	Positive predictive value	Hypergeometric p-value
1.00	77	59	76.6%	1.69×10^{-6}
0.99	77	59	76.6%	1.69×10^{-6}
0.95	77	59	76.6%	1.69×10^{-6}
0.90	75	57	76.0%	4.59×10^{-6}
0.80	75	57	76.0%	4.59×10^{-6}
0.70	75	57	76.0%	4.59×10^{-6}
0.60	53	40	75.5%	0.0003
0.50	52	40	76.9%	0.0001

Out of 342 genes in the model, 77 genes were predicted to be essential: 45 gene knockouts were predicted to be lethal (i.e., biomass solution = 0), while 32 were growth-limiting (i.e., biomass solution < 95% optimal biomass solution). Of the genes associated with lethal and growth-limiting knockouts, 82.2% and 68.8% were in the gold standard list, respectively. Combining genes associated with lethal and growth-limiting knockouts, 76.6% were associated with experimentally validated essential genes; 1.44 times more compared with the percentage of gold standard genes in the whole model (53.2%). Significant enrichment was observed in the group of genes associated with lethal and growth-limiting knockouts, with p-values of 1.52×10^{-5} and 0.047, respectively; however, combining these two sets exhibited even more significant enrichment (p-value = 1.69×10^{-6}). Figure 4.4 shows a comparison of the total number of essential genes divided into those associated with lethal and growth-limiting knockouts alongside the percentages of true positive predictions. Table 4.4 shows the total number of predicted essential genes and the total number present in the gold standard list. The corresponding enrichment scores and hypergeometric p-values are also presented in the table.
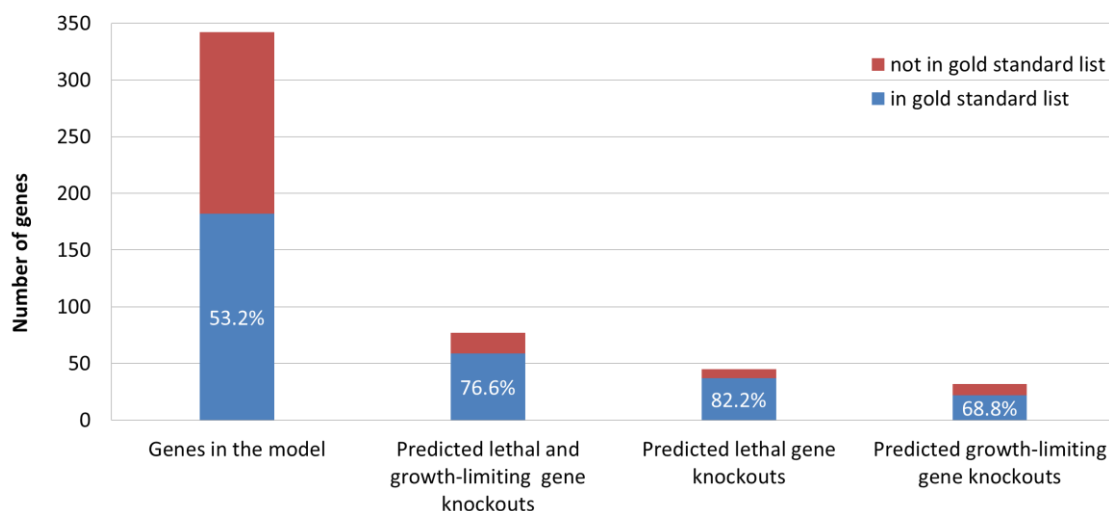
Figure 4.4 Total number of predicted essential genes
This figure shows the total number of predicted essential genes together with the percentage of true positive predictions in comparison with the total number of genes in the model. The essential genes were divided into lethal and growth-limiting depending on the predicted effect of single gene knockout using COBRApy.

Table 4.4 Total number of predicted essential genes in comparison with the set of gold standard genes

	Genes in the model	Predicted essential genes		
		Combined	**Lethal**	**Growth-limiting**
Total	342	77	45	32
in gold standard list	182 (53.2%)	59	37	22
not in gold standard list	160	18	8	10
Positive predictive value		76.6%	82.2%	68.8%
Enrichment score		1.44	1.55	1.29
p-value		1.69×10^{-6}	1.52×10^{-5}	0.047

Evaluating all 77 essential genes, 29 genes (37.7%) were associated with reactions involved in amino acid metabolism which includes haemoglobin digestion. There were 21 (27.3%) genes involved in carbohydrate metabolism and another 21 involved in nucleic acid metabolism. Notably, there was a significant increase in the proportion of genes associated with reactions involved in amino acid metabolism (p = 0.0001), nucleic acid metabolism (p = 0.0002) and carbohydrate metabolism (p = 0.007) in the list of essential genes in comparison with the overall proportions in the model. In order to identify the impact of adding boundary flux constraints in the model, single gene knockouts were also performed using the original model (iFT342) with unconstrained amino acid and glucose boundary exchanges. A total of 42 essential genes were identified, all of which were associated with a lethal phenotype and 34 (81.0%) are in the gold standard list (enrichment score = 1.52, p-value = 7.17×10^{-5}). All 42 essential genes

in the unconstrained model were in the 77 essential genes identified using the schizont stage (constrained) model. Although the unconstrained model yielded a slightly higher enrichment score, it also gave a smaller set of predicted essential genes, missing 25 (42.4%) of true positives.

In the constrained model, 7 genes involved in glycolysis were identified as essential, while only 3 were identified in the unconstrained model. This increase may be attributed to limiting the glucose and lactate flux to the values obtained experimentally. In the unconstrained model, lactate from the external environment is consumed rather than produced by the model, contributing to the production of important metabolites from glycolysis that ultimately feed into the biomass reaction (e.g., pyruvate, glyceraldehyde 3-phosphate). In comparison with nucleic acid metabolism, where boundary constraints on nucleosides/bases were not constrained, there was no difference in the set of essential genes identified by the constrained and unconstrained models. On the other hand, 22 (75.9%) out of the 29 predicted essential genes involved in amino acid metabolism were involved in haemoglobin digestion; and 17 out of the 22 genes (77.3%) were in the gold standard list. This is in stark contrast with the unconstrained model where none of the predicted essential genes were involved in haemoglobin digestion. The additional constraints in 18 amino acid boundary flux in the model have highlighted the importance of haemoglobin digestion in the model given the limited amino acid supply from the external environment. Figure 4.5 shows a comparison of the proportions of genes based on subsystem involvement between the predicted set of essential genes and the overall proportions in the model. Significant enrichment of genes involved in amino acid, carbohydrate and nucleic acid metabolism suggests the importance of these pathways in parasite growth. Table 4.5 shows the list of 18 novel essential genes (i.e., those not in the gold standard list) identified using the schizont stage model.
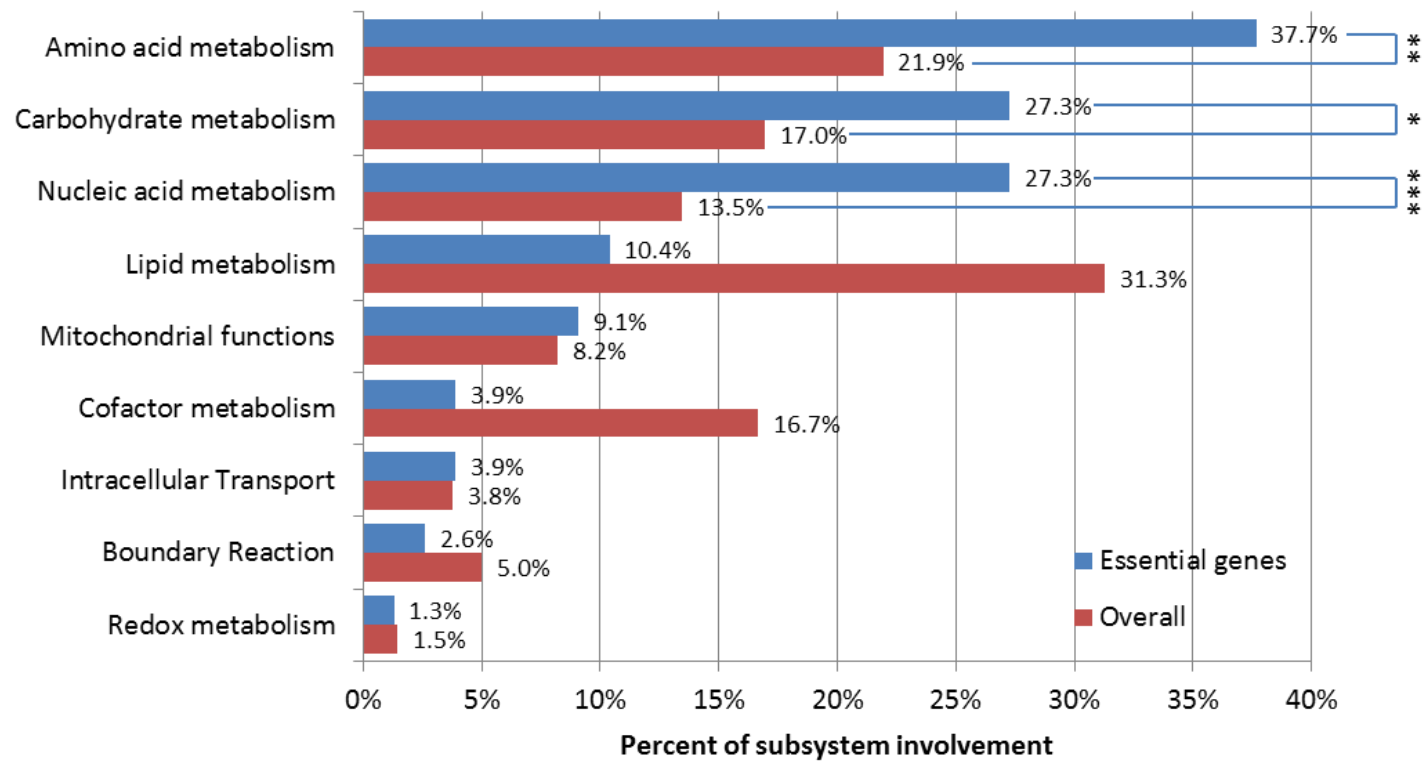
Figure 4.5   Subsystem involvement of reactions associated with predicted essential genes using the schizont stage model
This figure shows the proportion of essential genes by subsystem or pathway involvement of their associated reactions in comparison with all the genes in the model. Subsystems with significant enrichment of predicted essential genes involved are marked with asterisks. Note that depending on the associated reactions, a given gene may be involved in more than one subsystem.
*p-value = 0.0066, **p-value = 0.0002, ***p-value = 0.0001

Table 4.5   List of novel gene targets and their associated reactions as predicted by single gene knockout analysis using the schizont stage model

Gene ID	Associated reaction/s	Pathway/s	EC number	Effect of gene knockout
PF3D7_0111500	UMP-CMP kinase	Pyrimidine Metabolism	2.7.4.14	Lethal
PF3D7_0507200 PF3D7_0932300 PF3D7_1115300 PF3D7_1115400 PF3D7_1401300	OxyHaemoglobin digestion: Vacuole	Haemoglobin Digestion	3.4.11.- and 3.4.11.1 and 3.4.11.2 and 3.4.11.9 and 3.4.11.18 and 3.4.11.21 and 3.4.14.1 and 3.4.21.62 and 3.4.22.- and 3.4.23.- and 3.4.23.38 and 3.4.23.39 and 3.4.24.-	Growth-limiting
PF3D7_0813800	GDP-mannose 4,6-dehydratase	Mannose and Fructose Metabolism	4.2.1.47	Lethal
PF3D7_0815900 PF3D7_1020800 PF3D7_1446400	Dihydrolipoamide acyltransferase component E2	Pyruvate Metabolism	2.3.1.12	Lethal
	Dihydrolipoyl dehydrogenase	Glycine and Serine Metabolism, Pyruvate Metabolism	1.8.1.4	
	2-Oxoglutarate dehydrogenase complex	Mitochondrial TCA Cycle	1.2.4.2 and 1.8.1.4 and 2.3.1.61	
	Pyruvate dehydrogenase E1 component subunit alpha and beta	Pyruvate Metabolism	1.2.4.1	
PF3D7_0823900	Citrate transfer: mitochondria to cytosol; dicarboxylate/tricarboxylate carrier	Mitochondrial TCA Cycle, Pyruvate Metabolism, Intracellular Transport	-	Growth-limiting
	Dicarboxylate/tricarboxylate carrier (DTC), Malate:oxoglutarate antiporter	Mitochondrial TCA Cycle, Intracellular Transport	-	
PF3D7_0922600	L-Glutamate ammonia ligase	Glutamate Metabolism, Nitrogen Metabolism	6.3.1.2	Growth-limiting
PF3D7_0927300	Fumarate hydratase	Mitochondrial TCA Cycle	4.2.1.2	Growth-limiting

Table 4.5 List of novel gene targets and their associated reactions as predicted by single gene knockout analysis using the schizont stage model (continued)

Gene ID	Associated reaction/s	Pathway/s	EC number	Effect of gene knockout
PF3D7_1014000	GDP-L-fucose synthase	Mannose and Fructose Metabolism	1.1.1.271	Lethal
PF3D7_1017400	D-Mannose 6-phosphate 1,6-phosphomutase	Mannose and Fructose Metabolism	5.4.2.8	Lethal
PF3D7_1251300	ATP:dUMP phosphotransferase	Pyrimidine Metabolism	2.7.4.9	Lethal
	Thymidylate kinase	Pyrimidine Metabolism	2.7.4.9	
PF3D7_1368700	Mitochondrial carrier protein	Mitochondrial TCA Cycle, Intracellular Transport	-	Growth-limiting
PF3D7_1453500	NADPH:NAD+ oxidoreductase	Nicotinate and Nicotinamide Metabolism	1.6.1.1	Growth-limiting

In the Plata model, 202 out of all 366 genes (55.2%) in the model and 44 out of the 58 (75.9%) predicted essential genes (< 95% biomass solution) are in the gold standard list, giving an enrichment score of 1.38 with a hypergeometric p-value of 0.0004. Out of the 58 lethal and growth-limiting gene knockouts predicted by the Plata model, only 19 are in common with the essential genes predicted by our model. Unfortunately, gene association data was not available in the Huthmacher model file, thus the enrichment score and its accompanying hypergeometric p-value cannot be calculated. Out of the 185 genes associated with predicted essential reactions in the Huthmacher model, 115 (62.2%) are in the gold standard list and 26 are in common with our predicted gene set. The Forth model identified 79 essential genes, 55 of which (69.6%) are in the gold standard list. In comparison with the total number of genes in the Forth model that are in the gold standard list (87 out of 143, 60.8%), there is a lower but significant enrichment (1.14) of true positives in the predicted essential gene set (hypergeometric p-value = 0.01). It is not surprising that the Forth model had the highest true positive rate (63.2%) compared to iFT342 (32.4%) and the Plata models (21.8%), given the level of manual and meticulous curation done on this small model. Table 4.6 summarises the total number of essential genes predicted by the four malaria metabolic models in comparison with the gold standard list along with the corresponding enrichment scores and hypergeometric p-values. This table also includes the total number of predicted essential genes by the unconstrained model. Figure 4.6 gives a summary of the total number of common essential genes predicted by the four malaria metabolic models. There are a total of 11 genes that have been predicted by all four models to be essential, and only one, thymidylate kinase (PF3D7_1251300, 2.7.4.9) is considered novel (Table 4.7).

Table 4.6   Number of essential genes predicted by existing malaria models in comparison with the gold standard list in order of increasing hypergeometric p-value

Model	Genes				Positive predictive value	True positive rate	Enrichment score	p-value
	Total in model	Actual essential	Predicted essential	True positive				
iFT342	342	182 (53.2%)	77	59	76.6%	32.4%	1.44	1.69×10^{-6}
iFT342u*	342	182 (53.2%)	42	34	81.0%	18.7%	1.52	7.17×10^{-5}
Plata	366	202 (55.2%)	58	44	75.9%	21.8%	1.38	0.0004
Forth	143	87 (60.8%)	79	55	69.6%	63.2%	1.14	0.01
Huthmacher	579	**	185	115	62.2%	**	**	**

*Unconstrained iFT342 model

**IDs for all genes in the Huthmacher model were not available therefore comparison with the gold standard list and calculation of true positive rate, enrichment score and hypergeometric p-value were not possible
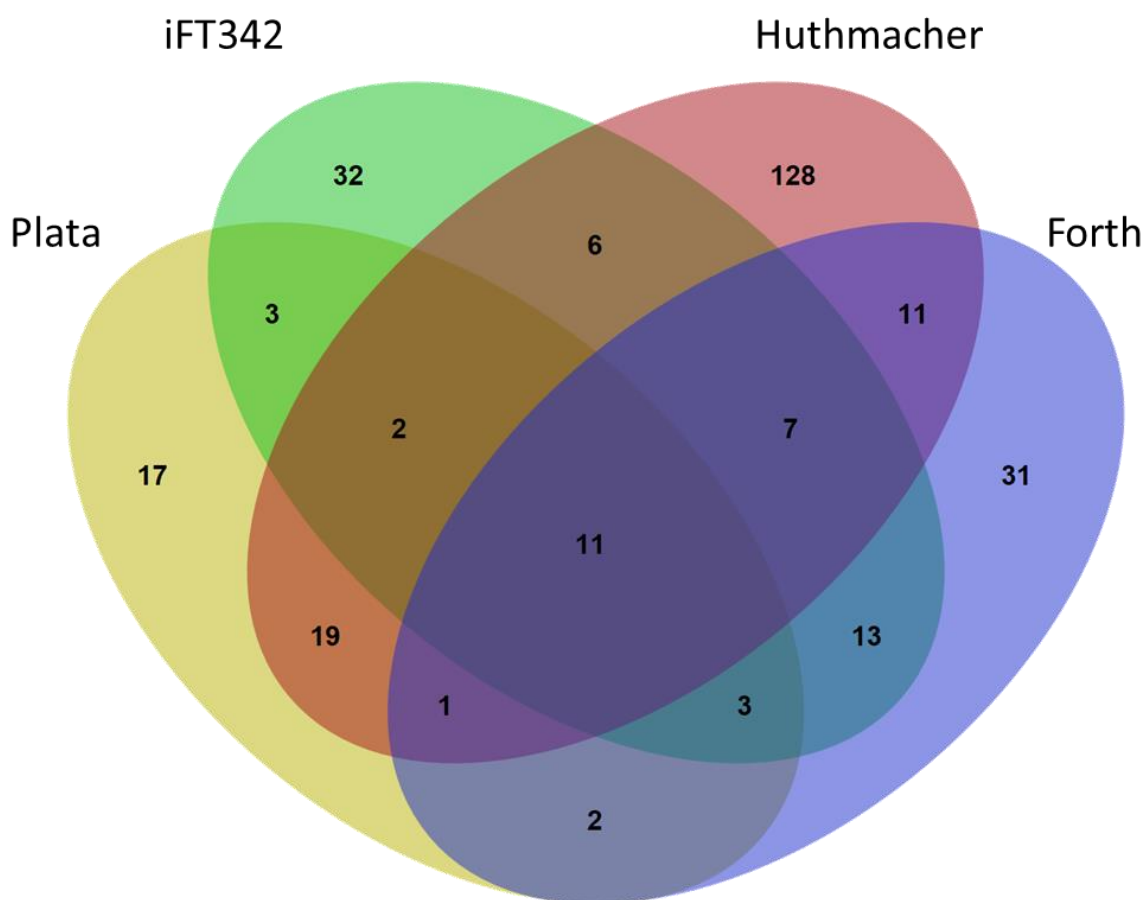
Figure 4.6   Venn diagram of essential genes predicted by the four models
This figure shows the total number of common and unique essential genes predicted by the four malaria models.

Table 4.7   Common essential genes predicted by all four malaria models

Gene ID	Name	Novel target?
PF3D7_0417200	Dihydrofolate reductase-thymidylate synthase	in gold standard
PF3D7_0512700	Orotate phosphoribosyltransferase	in gold standard
PF3D7_0603300	Dihydroorotate dehydrogenase	in gold standard
PF3D7_0923800	Thioredoxin reductase	in gold standard
PF3D7_0928900	Guanylate kinase	in gold standard
PF3D7_1015900	Enolase	in gold standard
PF3D7_1023200	Orotidine monophosphate decarboxylase	in gold standard
PF3D7_1251300	Thymidylate kinase	novel
PF3D7_1308200	Carbamoyl phosphate synthetase	in gold standard
PF3D7_1344800	Aspartate carbamoyltransferase	in gold standard
PF3D7_1472900	Dihydroorotase	in gold standard

4.3.4   Validation of novel targets

The protein sequences of all genes in the model were aligned against the sequences in the DrugBank database (Law *et al*, 2014) and compounds (e.g., inhibitors, cofactors, substrate)

that are known to interact with the aligned sequences were identified. Out of 342 genes in the model, 124 were mapped against proteins in DrugBank. It is important to note that the selection of inhibitors for identified novel targets was done prior to the incorporation of the gold standard genes obtained from *Plasmo*GEM. By utilising the original set of 128 gold standard genes, 36 out of the 77 predicted targets were identified as novel. Twenty of these "novel" targets aligned with proteins in the DrugBank database, with only 4 having known inhibitors. However, after taking into consideration the set of genes obtained from *Plasmo*GEM, only two of these targets remain novel: dihydrolipoyl dehydrogenase, apicoplast (aLipDH) (PF3D7_0815900) and UMP-CMP kinase, putative (PF3D7_0111500). Table 4.8 shows the four predicted novel gene targets that aligned with protein sequences in the DrugBank database, together with the corresponding alignment e-value and known inhibitors. Some of the identified inhibitors have already been tested against the parasite *in vitro*. Methotrexate, an anti-metabolite drug that is currently being used in autoimmune diseases and cancer, has been shown to inhibit *in vitro* growth of a number of *P. falciparum* strains including the multidrug-resistant (K1) strain (Dar *et al*, 2008). Methotrexate is known to target dihydrofolate reductase, which is also targeted by the antimalarial drug cycloguanil (Uga *et al*, 2006; Srivastava & Vaidya, 1999). In addition, it has also been identified as a potent inhibitor of the human 6-phosphogluconate dehydrogenase (6PGD). *In vitro* inhibition assay on human 6PGD has shown that ketotifen, dacarbazine and meloxicam had IC_{50} in the low nanomolar range while furosemide and methotrexate had slightly higher IC_{50} between 100 and 200 nM (Akkemik *et al*, 2010). Among the identified inhibitors of the human 6PGD in Table 4.8, ritodrine had the highest IC_{50} of 3.66 mM and was therefore selected for testing against the parasite.

Table 4.8   Novel gene targets with protein sequences that aligned with proteins in the DrugBank database with known FDA-approved inhibitors in order of increasing e-value

Gene ID	Gene name	E-value	Drug name	Drug class
PF3D7_1454700	6-phosphogluconate dehydrogenase, decarboxylating, putative	1.53E-132	Methotrexate	Anti-metabolite, anti-cancer, immunosuppressant
			Furosemide	Diuretic
			Meloxicam	Non-steroidal anti-inflammatory drug
			Dacarbazine	Anti-cancer
			Ritodrine	Uterine relaxant
			Ketotifen	Antihistamine
PF3D7_1015800	Ribonucleotide reductase small subunit, putative	3.38E-45	Cladribine	Anti-cancer
			Gallium nitrate	Treatment for hypercalcaemia
PF3D7_0815900	Dihydrolipoyl dehydrogenase, apicoplast (aLipDH)	1.24E-25	Carmustine	Anti-cancer
PF3D7_0111500	UMP-CMP kinase, putative	1.08E-24	Gemcitabine	Anti-metabolite, anti-cancer

Cladribine and gemcitabine, which are antineoplastic nucleoside analogues, were also selected for testing against the parasite *in vitro*. Cladribine is converted to its triphosphate form and inhibits ribonucleotide reductase (M2 subunit). This in turn reduces the supply of deoxynucleotide triphosphates in the cell, thus inhibiting tumour growth. The triphosphate form of cladribine also gets incorporated into the DNA during DNA elongation, affecting DNA synthesis and repair (Bonate *et al*, 2006; Robak *et al*, 2006). Similarly, gemcitabine is phosphorylated into its active metabolite, which then inhibits human ribonucleotide reductase and affects DNA synthesis and repair in neoplastic cells (Mini *et al*, 2006). Specifically, gemcitabine binds to the large subunit of the human ribonucleotide reductase which is not identical to the parasite's ribonucleotide reductase (e-value = 0.99). Moreover, gemcitabine also inhibits UMP/CMP kinase, which is important in the pyrimidine metabolism (Hsu *et al*, 2005). For proof-of-concept, three drugs (i.e., ritodrine, cladribine and gemcitabine) separately targeting three different enzymes were selected for testing in parasites *in vitro*, while carmustine, which is used for malignant gliomas, was not selected as it is the most hazardous of the four compounds (based on Sigma materials safety data sheet). Rosiglitazone, an inhibitor of acyl-CoA synthetase (PF3D7_0525100), was used as in the negative control as the enzyme was predicted to be non-essential for parasite growth. DSM265 (Phillips *et al*, 2015), which targets dihydroorotate dehydrogenase (a known essential enzyme for parasite growth) was used in the positive control. It is important to note that 6PGD is present in the red blood cell which may

interfere with the metabolism of both the parasite and the host cell; however, ribonucleotide reductase and UMP/CMP kinase are both absent in the red blood cell (Prchal & Gregg, 2005).

Although the parasite 6PGD sequence had a very significant alignment with the human counterpart (UniProt ID: P52209), growth inhibition with ritodrine was not observed *in vitro* (Figure 4.7). Similarly, there was no growth inhibition observed in the presence of cladribine even though the predicted target, the *Plasmodium* ribonucleotide reductase (PF3D7_1015800, 1.17.4.1), is quite similar to the M2 subunit of the human homologue (UniProt ID: P31350), with five out of the six metal binding sites and the single active site being conserved between the two (Magrane & UniProt Consortium, 2011). On the other hand, gemcitabine has been shown to inhibit parasite growth, with an IC_{50} of 2.62 ± 0.53 μM. Several groups have tested FDA-approved drugs for antimalarial activity (Eastman *et al*, 2013; Chong *et al*, 2006; Weisman *et al*, 2006), and these results were consistent with those of Eastman *et al* and Chong *et al*. Possible reasons for the inactivity of ritodrine and cladribine against the parasite may be because the drugs do not reach their enzyme targets, or because they are not converted into their active form. Both cladribine and gemcitabine have been shown to be transported by the *Plasmodium* equilibrative nucleoside transporter (PfENT1) (Parker *et al*, 2000) and both require phosphorylation by deoxycytidine kinase for enzyme inhibition to take place (Robak *et al*, 2006; Mini *et al*, 2006). However, differences in the affinity to nucleoside kinases as well as difference in substrate binding sites in the target enzyme may be reasons for the inactivity of the former. 6PGD is an important enzyme in the pentose phosphate pathway, specifically in the generation of redox equivalents (i.e., NADPH) and ribose 5-phosphate for DNA synthesis (Bozdech & Ginsburg, 2005). Because the second half of the pentose phosphate pathway is also capable of producing ribose 5-phosphate in the absence of 6PGD, this enzyme may be important when it comes to combatting elevated oxidative stress. Culture conditions may therefore play a role in the effectivity of ritodrine and this may be the reason why the drug exhibited 14% growth inhibition at 10 μM *in vitro* (Weisman *et al*, 2006). Looking at other drugs that have been predicted to target 6PGD (Table 4.8), methotrexate and ketotifen have been shown to inhibit *Plasmodium* growth *in vitro* and *in vivo*, respectively (Dar *et al*, 2008; Milner *et al*, 2012). However, these information must be interpreted with caution as these drugs are known to have other targets (Dar *et al*, 2008; Law *et al*, 2014).
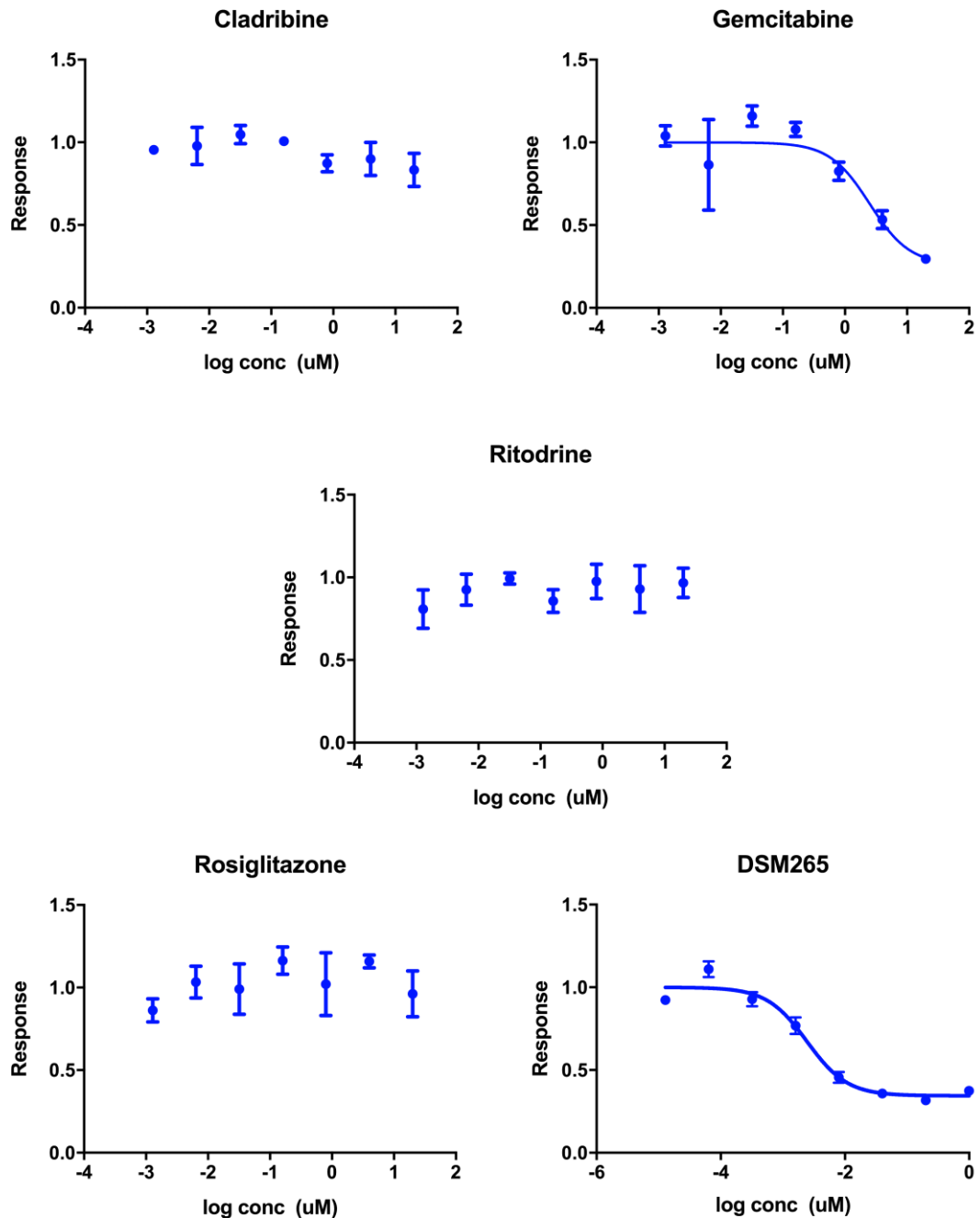
Figure 4.7 **Dose response curves of *P. falciparum* 3D7 after 48 hours of incubation in selected compounds (average ± SEM)**
Cladribine, gemcitabine and ritodrine were identified using the DrugBank database as potential inhibitors of novel targets predicted using the schizont stage model. Rosiglitazone, which was predicted to inhibit a non-essential target, was used as a negative control and DSM265, which targets DHODH, was used as a positive control. (n = 3)

The effect of gemcitabine at 10 times the determined IC_{50} on synchronised *P. falciparum in vitro* cultures was monitored over a 48 hour incubation period. Gemcitabine seemed to prevent the development of schizonts, thus resulting in a culture composed purely of trophozoites after 48 hours of incubation with the drug (Figure 4.9). Upon closer inspection of parasites, large lightly stained bodies (V) were visible in a number of trophozoites (Figure 4.10b).

In addition, haemozoin in most trophozoites appeared to be disorganised compared with normal trophozoites (Figure 4.10c-d). Gemcitabine was predicted to inhibit *P. falciparum* UMP-CMP kinase (PF3D7_0111500, 2.7.4.14), which phosphorylates UMP to UDP, the latter serving as a precursor for thymidine deoxynucleotides and cytidine nucleotides and deoxynucleotides. The parasite UMP-CMP kinase, which is 371 amino acids long, aligns with the shorter human homologue (UniProt ID: P30085, 196 amino acids long), with an e-value of 1.08×10^{-24}. Five of the six amino acid residues that serve as binding sites for a number of nucleotides are conserved between the two species, suggesting that gemcitabine may indeed be targeting the parasite UMP-CMP kinase. Limiting nucleotides important for DNA synthesis may therefore explain why parasites incubated in gemcitabine were unable to undergo schizogony.
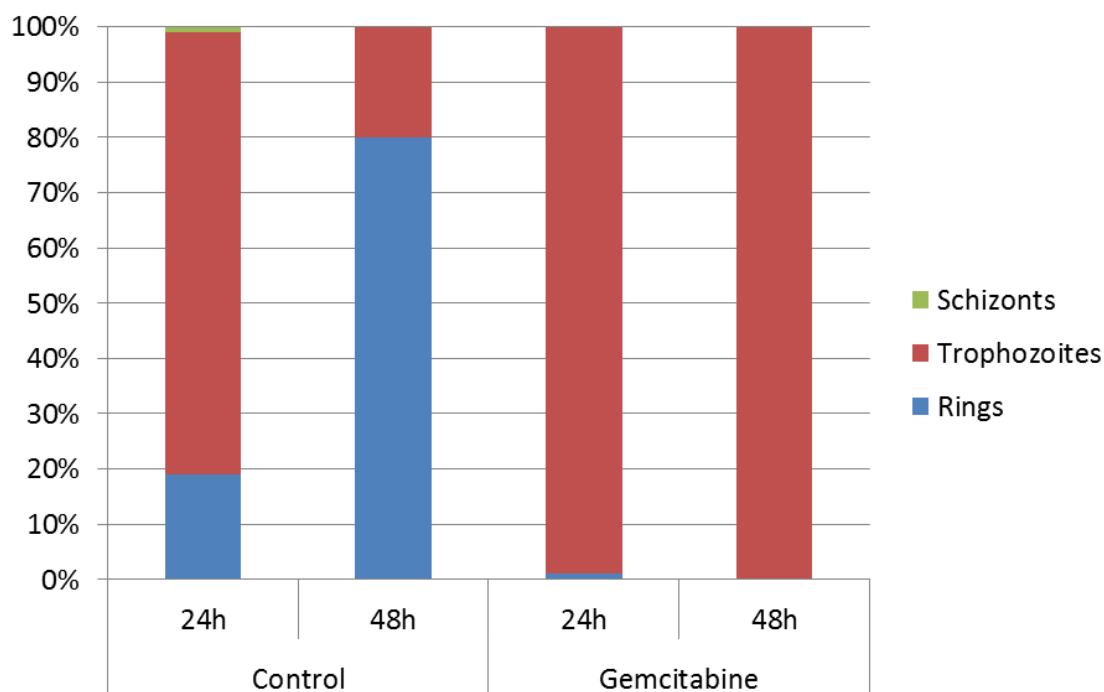
Figure 4.8   Percentage of parasite stages at 24 and 48 hours after incubation in 10 times the determined IC$_{50}$ of gemcitabine
Synchronised parasites were grown in malaria complete media with 10 times the determined IC$_{50}$ of gemcitabine and was observed at 24 and 48 hours after adding the drug. The parasites incubated in gemcitabine were unable to progress from trophozoite stage to schizont stage, and therefore were unable to replicate and produce rings after 48 hours (in contrast with the control cultures).
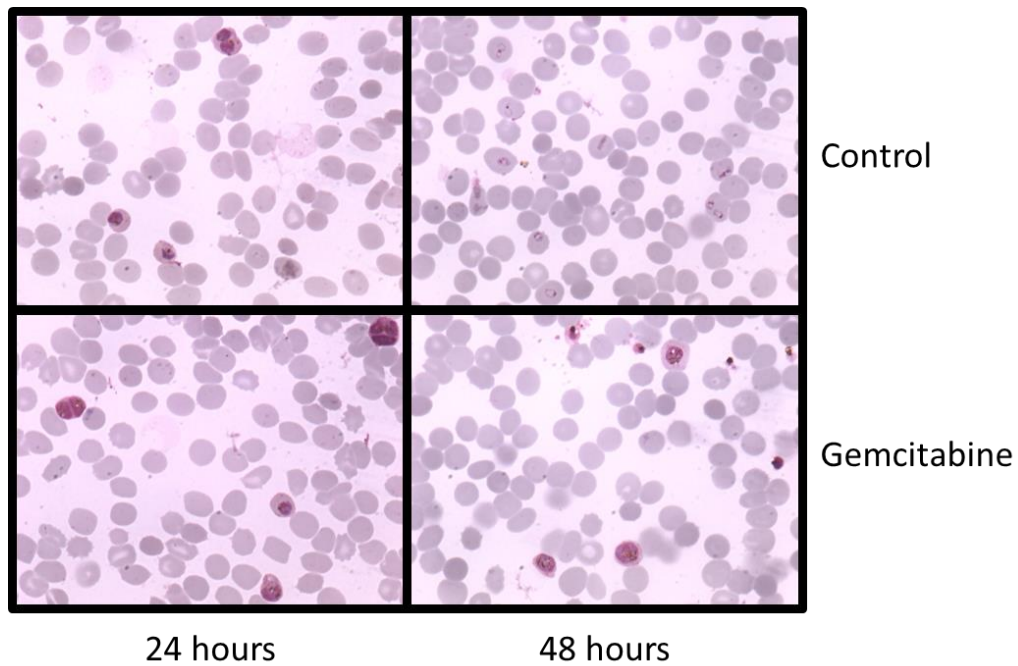
Figure 4.9   Blood smears of synchronised *P. falciparum* 3D7 parasites stained with 10% Giemsa after 24 and 48 hours of incubation in 10 times the determined IC$_{50}$ of gemcitabine
Compared with the control, parasites grown in gemcitabine were unable to develop from trophozoites to schizonts.
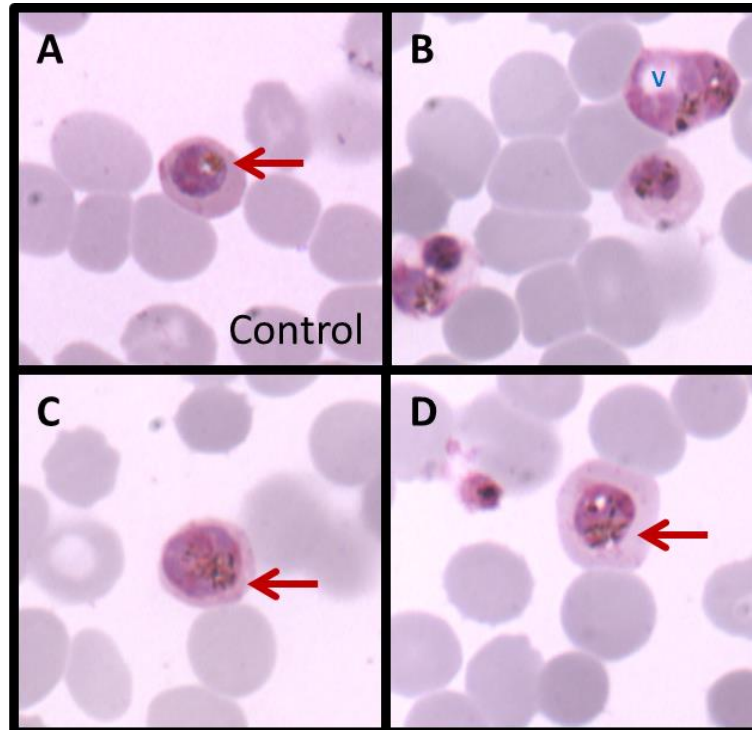
**Figure 4.10 Trophozoite stage *P. falciparum* 3D7 parasites stained with 10% Giemsa after 48 hours of incubation in 10 times the determined IC$_{50}$ of gemcitabine**
Large lightly stained bodies (V) were observed in a number of trophozoites after 48 hours of incubation in gemcitabine. Moreover, haemozoin (red arrow) in those exposed to the drug appeared to be disorganised (B-D) compared with the control (A).

4.4   Discussion

Addition of constraints is sometimes done with the aid of minimisation of metabolic adjustments (MOMA). This method finds a solution space for a constrained model that is closest to that of the unconstrained (i.e., wild-type) model (Segrè *et al*, 2002). Since some of the experimental metabolite flux constraints were able to generate feasible models, there was no need to apply MOMA as utilising experimental constraints will give better predictions compared with MOMA-adjusted constraints (Price *et al*, 2004). With the addition of experimentally obtained metabolite flux, the schizont stage model was able to model *in vitro* parasite growth in glucose restricted environment with remarkable fidelity (r = 0.985, p-value = 1.36 x 10^{-6}). Moreover, there was a significant enrichment of experimentally validated essential genes in the set of essential genes predicted by the model (p-value = 1.69 x 10^{-6}), more significant compared with the predictions of the Plata model (p-value = 0.0004). These validations therefore give us more confidence on the verity of the predictions of our model.

In contrast with other models, gene expression data was not incorporated into the iFT342. A number of malaria metabolic models have been utilised alongside gene expression data and have facilitated in the generation of accurate predictions (Huthmacher *et al*, 2010; Plata *et al*, 2010; Fang *et al*, 2014); however, the iFT342 has shown only very weak overall correlation between predicted flux and gene expression. This may be expected as there are several post-transcriptional factors that contribute to discrepancies between gene expression and overall protein activity. Although some have linked this discrepancy to the observed delay between peak gene expression and protein abundance (Le Roch *et al*, 2004; Foth *et al*, 2011), other contributing factors may include mRNA and protein half-lives. In a study that looked into the rates of decay of mRNA of 2,744 *P. falciparum* genes (Shock *et al*, 2007), it has been shown that the half-lives of different mRNAs globally change throughout the erythrocytic cycle of the parasite. Furthermore, groups of genes with similar patterns in decay rates showed significant enrichments of GO terms. This therefore suggests that good correlation may be observed between gene expression and flux in certain groups of genes in a metabolic model but not all. In lieu of gene expression data, proteomics and metabolomics data incorporated into a metabolic model have been shown to improve model accuracy (Yizhak *et al*, 2010). Also, incorporation of metabolomics data has been suggested as a model constraint as this data provides a better insight on the enzyme activity compared with transcriptomics or proteomics data (Rossell *et al*, 2006; Yizhak *et al*, 2010).

Even though the predicted essential gene set obtained using the unconstrained iFT324 model gave a higher enrichment score of 1.52, versus 1.44 in the schizont stage model, the unconstrained model predicted 45.5% fewer essential genes and missed 42.4% of the gold

standard genes predicted by the constrained model. This therefore reiterates the important contribution of adding metabolic constraints to the accuracy of model predictions (Price *et al*, 2004). With this in mind, the schizont model may further be improved through the incorporation of more experimentally measured constraints. Also, incorporation of experimentally measured lipid and carbohydrate biomass components may increase the accuracy of model predictions. In addition, it is worth noting that the Forth model showed the highest true positive rate compared with the other malaria models. This may be attributed to the manual curation that was done in developing this model. With further evaluation, analysis and curation of the iFT342 model, it is likely that the accuracy of model predictions will also improve. Cut offs for identifying growth-limiting knockouts are also important. Given the results of this study, a threshold of 95% of the maximum biomass output is recommended, as lower thresholds may miss important essential targets.

The poor correlation between the predictions of the early stage (i.e., ring and trophozoite) models with the *in vitro* data may be attributed to the limited accuracy of monitoring metabolite concentrations in the growth media as a proxy for measuring metabolite exchanges between the parasite and red blood cell, particularly during the early stages. The erythrocyte membrane has been shown to slowly increase in permeability via the NPPs during the early parasite stages (Waldecker *et al*, 2017). It is possible that the metabolite exchanges between the ring stage parasite and the internal compartment of the red blood cell (host cytosol) may not be well represented by the exchange of metabolites between the infected RBC and the external environment (media) due to the fact that the permeability during the ring stage is lesser compared with the permeability during the later stages. Furthermore, disproportionate changes in the parasite and host cell volumes (Lew *et al*, 2003) may have resulted in an apparent efflux of metabolites during the trophozoite stage, making changes in metabolite concentrations in spent media also less representative of host and parasite metabolite exchanges at this stage. However, it is also possible that the ring and trophozoite stage models may be modelling parasite metabolism more than parasite growth (i.e., biomass production), which is most exemplified in the schizont stage when mitosis occurs (Gerald *et al*, 2011). Thus assigning a more appropriate stage-specific objective function may be considered.

On the contrary, shifts in metabolite concentration in the spent media may best represent influx and efflux of these different components from the parasite during the schizont stage, when the red blood cell is most permeable and the host and parasite volumes increase at a slow but steady rate (Lew *et al*, 2003). Moreover, experimental quantification of biomass components that was used to derive the model's biomass reaction equation was obtained from late schizont stage parasites (Forth, 2012), which may explain why the schizont stage model

predictions correlated well with the *in vitro* data. In addition, asynchronous parasite cultures were utilised in generating the *in vitro* data and growth measurements were estimated using SYBR green fluorescence which detects the amount of double stranded DNA that correlates well with parasitaemia (Smilkstein *et al*, 2004). Given that DNA replication happens in the schizont stage (Bozdech *et al*, 2003), the *in vitro* growth measurements may be representing the schizont stage more than the earlier stages, which also explains the high *in vitro* and *in silico* growth correlation.

Further improvement of the model may be done through inclusion of additional metabolite boundary constraints. Published data on metabolite concentrations in spent media over a 48 hour time course done by Olszewski *et al* (2009) may be utilised and the metabolite flux recalculated for incorporation into the model; however, due to the experimental technique that they used, separate concentration measurements for leucine and isoleucine was not feasible. Nevertheless, they have provided valuable concentration measurements for 59 metabolites at 8 hour intervals over the 48 hour erythrocytic cycle of the parasite. As many of these metabolites do not have boundary transport reactions in the model, further updating and expanding the model to include transport reactions and metabolic pathways involving these metabolites may prove to be an important next step in improving model accuracy. Development of accurate gametocyte and liver stage specific models with the use of stage specific biomass quantification and metabolite measurements will be important in the identification of drug targets and the subsequent development of drugs that can reduce transmission or kill latent stage parasites.

Here, 18 novel drug targets were identified (Table 4.5) and a few of them have already being investigated. For example, oxoglutarate dehydrogenase (PF3D7_0815900, 1.2.4.2, 1.8.1.4, 2.3.1.61) has been explored as potential drug targets. Oxythiamine, a thiamine analogue, which is converted by thiamine pyrophosphokinase to oxythiamine pyrophosphate, has been shown to inhibit the enzyme complex (Chan *et al*, 2013). Several attempts using multiple techniques to knockout fumarate hydratase (PF3D7_0927300, 4.2.1.2) in *P. falciparum* did not produce any viable parasite line which may suggest that this enzyme is essential for parasite growth (Ke *et al*, 2015). Given the importance of L-glutamate ammonia ligase in *Mycobacterium*, it is also being explored as a drug target in *Leishmania* and *Plasmodium* (Singh & Siddiqi, 2017). As a proof of concept, inhibitors for selected novel targets were tested, and incubation with gemcitabine, which was predicted to target UMP-CMP kinase (PF3D7_0111500, 2.7.4.14), resulted in inhibition of *in vitro* parasite growth. Additional *in vitro* assays may be done to ascertain that gemcitabine is targeting the said enzyme. This can be done by testing gemcitabine on purified parasite UMP-CMP kinase. Once this has been confirmed, comparison of human and parasite

enzyme may be the key to developing an analogue of gemcitabine that will specifically target the parasite but not the host enzyme. It is important to note that several drugs that interact with novel targets predicted by the model still have unidentified drug action (as to whether they are inhibitors, agonists, cofactors, etc.) on the enzyme target (based on DrugBank). These drugs may be evaluated for function and identified inhibitors may be repurposed as antimalarial drugs. Additionally, other drug databases such as the STITCH database (Kuhn *et al*, 2014) may be exploited in order to find other compounds that may inhibit these predicted novel targets. In the light of prioritising drug target validation, inhibition of novel targets may be focussed primarily on those commonly predicted by all malaria models. Gene knockouts through CRISPR-Cas9, which have already been demonstrated in *P. falciparum* (Ghorbal *et al*, 2014; Wagner *et al*, 2014) may also be considered as a more definitive means of novel target validation.

Chapter 5   Discussion

5.1   General purpose of genome scale metabolic models

The main objective of developing a genome scale metabolic model is to accurately represent the metabolic processes in a given organism with the goal of utilising the model as a less costly and less time-consuming alternative to wet lab experiments (Monk *et al*, 2014). Use of metabolic models becomes even more important when it comes to understanding dangerous organisms or pathogens or when dealing with organism stages that are difficult to experimentally manipulate (Gardner & Boyle, 2017). This objective may be misinterpreted as an aim to ultimately replace wet lab experiments. Rather, model simulations ought to be used as a means to generate hypotheses for validation in the laboratory (Kell & Goodacre, 2014). More importantly, the relationship between experimental and *in silico* work should be seen as a constant cyclical movement of information where one contributes to the improvement of the other (Figure 5.1). Initially, models are reconstructed using available information on an organism's genome and gene-associated reactions and species. Subsequently, additional experimental data may be incorporated into the model as constraints, contributing to the generation of more accurate predictions (Price *et al*, 2004). Moreover, the model can be improved by considering known essential and non-essential genes or reactions that are in the model and making the necessary corrections that are strongly based on existing published data to align the model predictions with gene essentiality data. As there is a continuous influx of new information on gene and reaction annotations and essentiality, as well as data obtained from experimental validation of novel targets, the model can regularly be updated and new information can be obtained and used to further influence experimental approach. In this process, new drug targets may be identified for drug development.
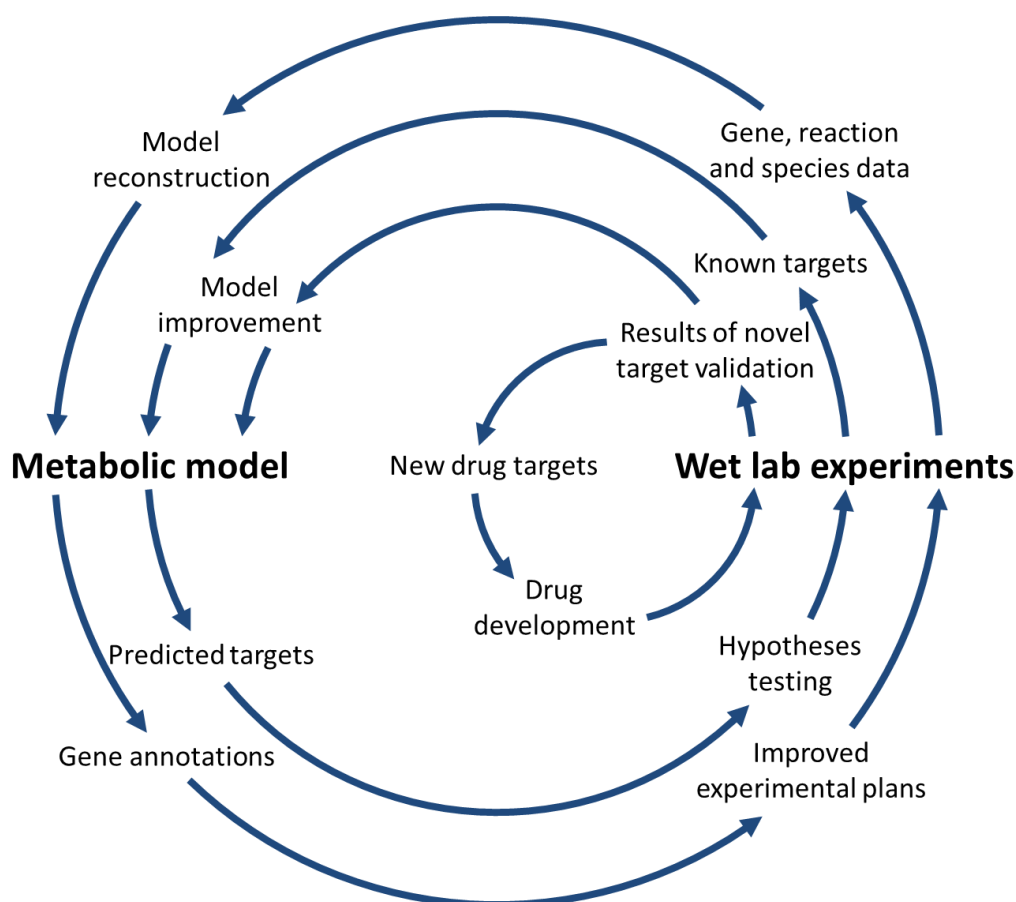
Figure 5.1   Flow of information between wet lab experiments and metabolic modelling
This diagram shows how information obtained from wet lab experiments can be used to improve genome scale metabolic models. Conversely, data and information obtained from model simulations provide hypotheses that may be utilised to prioritise experimental work and improve experimental plans.

5.2   Model reusability

A good metabolic network model may be summarised into two main characteristics: reusable and accurate. The former deals more with the format of the model while the latter focusses on the content. As with much scientific research, replicability is important. It is essential that models are made in a standardised format so that they may be utilised by the greater scientific community. The SBML format was developed for the purpose of generating a model that can be used by different software (Chaouiya *et al*, 2015). Furthermore, addition of species and reaction annotations have been recommended to facilitate model comparison (Ravikrishnan & Raman, 2015). Having species identifiers such as InChI keys, SMILES and PubChem IDs is an important characteristic of the iFT342 as none of the other existing malaria metabolic models have incorporated these identifiers. Furthermore, updating gene IDs and EC

classification numbers has also contributed to improving the quality of the model developed here.

More than simplifying model comparisons, models ought to be made in such a way that they can easily be merged with other models when necessary. This can be important in the case of developing consensus models and modelling interactions between organisms, such as the host and the parasite. This has already been done to model blood stage and liver stage malaria parasites (Huthmacher *et al*, 2010; Wallqvist *et al*, 2016; Bazzani *et al*, 2012). With the constant development and improvement of the "community consensus" human metabolic model, Recon 2.2 (Swainston *et al*, 2016), up-to-date human tissue specific models may be utilised alongside improved parasite models. In addition, there are a number of models on *Mus musculus* (Yilmaz & Walhout, 2017) that may be combined with rodent parasite models. Using gene homology and gene expression data, species specific malaria models may be developed for use with other host models. To date, there is no single ontological format that is recommended for use in genome scale metabolic models. Additional species and reaction annotations will aid in model comparison; however, this only partially addresses the issue of merging models. At least with COBRApy, reading multiple models is easy; but COBRApy relies heavily on the assigned IDs for accessing and editing species and reaction information. For modelling host-parasite interaction, one has to compare and standardise species IDs to allow proper merging of two models. As for the development of consensus models, having several models that utilise different ontological systems for reaction, species and compartment IDs, one still has to, at the very least, evaluate and compare species annotations of reactants and products in the reaction and the compartmental location to avoid unnecessary duplication of reactions.

Different groups have their own preference when it comes to the use of a specific reaction and species ontological system. As some of these groups have been using the same system for years, it would be difficult to impose a single unifying ontological system. Online resources such as metanetx.org (Bernard *et al*, 2014) offers conversion of identifiers in uploaded models into its own MNXref system, but the process is not 100% efficient and still requires manual examination of the mapped and unmapped identifiers. A similar resource may be developed that can serve as a repository for genome scale metabolic models that not only converts uploaded models into a unified ontological system, but also populates a species and reaction database with common and unique entries alongside their corresponding annotations. Species data may include the old ID, name, chemical formula, InChI keys, SMILES, PubChem IDs, links to other databases and an assigned unified ID. Reaction data may include the old reaction ID, name, EC classification number, gene association data, reactants and products, reversibility, links to other databases when available and an assigned unified ID. These species and reaction data may

then be used to evaluate and convert species and reaction IDs from uploaded models. Similar to other trustworthy databases and online resources, manual curation of database entries may be necessary. Community-driven database maintenance may be applied where different experts can contribute, annotate and amend database information in conjunction with addressing feedback from users, similar to what is being implemented in other knowledgebases like EuPathDB (Aurrecoechea *et al*, 2013), GeneDB (Logan-Klumpler *et al*, 2012) and TrypanoCyc Pathway/Genome Database (Shameer *et al*, 2015).

5.3 Model accuracy

The second characteristic of a good model is accuracy. This includes use of reliable gene, reaction and species information in model reconstruction. The accuracy of the predictions of the iFT342 model may partly be attributed to the use of the small but highly curated Forth model (Forth, 2012) and utilising online resources to ensure reactions from the other malaria models (Plata *et al*, 2010; Huthmacher *et al*, 2010) that were added into the final model have accurate annotations. Gene-protein-reaction relationships, reaction (e.g., reaction stoichiometry, reversibility) and species attributes (e.g., chemical name, formula) were verified and amended based on information obtained from PlasmoDB (Aurrecoechea *et al*, 2009), MPMP (Ginsburg & Abdel-Haleem, 2016), KEGG (Kanehisa *et al*, 2016) and BRENDA (Placzek *et al*, 2017). Increasing the genome coverage of the model and considering reactions located in compartments not present in the current model may help increase the accuracy of the model, especially when considerable attention is given to ensure the verity of the incorporated reactions and reaction attributes. Reactions from the Huthmacher model contained in the nucleus and the endoplasmic reticulum may be further evaluated before adding into the current model. Transport reactions can then be added to link these reactions to the existing metabolic network.

Keeping in mind that FBA looks into the interconversion of metabolites within the metabolic network of a given organism, the correctness of reaction stoichiometry and reversibility plays a very crucial role in ensuring the accuracy of model predictions. The reactions in the iFT342 model have been compared against the abovementioned databases, checked for reaction mass balance and corrected as needed. A similar approach of consulting multiple databases to ensure correctness of reaction reversibility was done; however, there were a number of reactions with no available information. In this case, the reaction was maintained as reversible. Since the reversibility of reactions affects the accuracy of model predictions (Swainston *et al*, 2016), the accuracy of the current model will greatly benefit from new information on reaction thermodynamics. Nevertheless, the percentage of irreversible reactions in the model is close to that of highly curated models. A comparison of the number and

percentage of irreversible reactions in selected metabolic models, including highly curated human, yeast and *E. coli* models is shown in Table 5.1.

Table 5.1   Number and percentage of irreversible reactions in selected GSMs determined using MetExplore in order of decreasing percentage of irreversible reaction

Model name	Organism	Total number of reactions	Irreversible reactions		Reference
			Number	Percentage	
iFT342	*P. falciparum*	551	355	64.4%	
iAF1260	*E. coli*	2382	1530	64.2%	(Feist *et al*, 2010)
iMM904	*S. cerevisiae*	1577	928	58.8%	(Herrgård *et al*, 2008)
iTH366	*P. falciparum*	1001	589	58.8%	(Plata *et al*, 2010)
Recon2	*H. sapiens*	7440	4112	55.3%	(Thiele *et al*, 2013)
iAC560	*L. major*	1112	482	43.3%	(Chavali *et al*, 2008)
iTF143	*P. falciparum*	247	48	19.4%	(Forth, 2012)
PlasmoNet	*P. falciparum*	1376	0	0.0%	(Huthmacher *et al*, 2010)

Even though all internal reactions in the iFT342 model are mass balanced, there is still room for improvement in terms of species with Markush structures (R-groups) in their molecular formulas. In the development of the consensus yeast model (Herrgård *et al*, 2008), some reactions with species having R-groups in their molecular formulas were removed from the model to avoid ambiguity; however, this resulted in the decrease in reactions involved in certain pathways. A better means to address this issue is through more extensive curation as in the development of the human model, reducing the total number of species with R-groups from 377 in the Recon 2.04 model to 68 in the 2.2 version (Swainston *et al*, 2016). The iFT342 model also has 68 metabolites with R-groups in their molecular formulas. It is important to note that with updating the molecular formulas of these species, reaction mass balance have to be reassessed and it may be necessary to add new reactions to accommodate the addition of these "new" metabolites.

With regard to dead-end metabolites, the model developed here had none and has the least percentage of dead reactions; however, to further improve the model, it is important to consider other forms of metabolites that may result in dead reactions. Take for example the case of the metabolite CMP in the model (as discussed in Chapter 2) where it is produced by multiple irreversible reactions (in two different compartments) but is only consumed by a single transport reaction that shuttles the metabolite between these two compartments. Even though

CMP is both produced and consumed, it has contributed to blocked reactions, therefore additional reactions were necessary to alleviate the problem. Given that 69 (12.5%) dead reactions are still present in the model, further assessment of reactions may be done to identify and correct these gaps. "Pseudo-gap" metabolites (Ponce-de-León *et al*, 2013) may be the reason for some if not all of these blocked reactions. "Pseudo-gap" metabolites are those that participate in more than one reaction (both consumed and produced) but are constantly being recycled (as in the case of a cofactor). This results in a zero net production of the said metabolites. Therefore pathways that terminate in the production or consumption of these metabolites end up with zero flux.

Compared with the other malaria metabolic models (Forth and Plata models), the iFT342 model was able to generate the most accurate gene essentiality predictions, giving the highest and most significant enrichment of true positive predictions (compared with experimentally validated essential or gold standard genes). Out of 182 experimentally validated essential genes that are present in the whole model, only 59 were identified as essential after simulating single gene knockouts. This means that the model can still be refined further by correcting the discrepancy (123 false negatives) between the gold standard genes and the list of predicted essential genes (Chavali *et al*, 2012b). By looking at the reactions associated with false negatives (i.e., falsely identified as non-essential genes), one can evaluate whether there are redundant reactions that may be removed in order to align the predictions with the gold standard list. Reactions that bypass those that are associated with gold standard genes also require re-evaluation. On the other hand, contributions of these false negatives to biomass production should also be evaluated. It is possible that many of these false negatives are important to the parasite metabolic pathways that are not involved in biomass production, such as maintaining redox balance or simply generating energy equivalents for parasite survival.

As mentioned in the previous chapters, some models have utilised gene expression data in order to generate more accurate predictions (Plata *et al*, 2010; Huthmacher *et al*, 2010; Wallqvist *et al*, 2016). A similar approach may be done using the current model to simulate different stage specific parasite metabolism. The timing of the gene expression data may be shifted to account for the lag between transcription and translation (Foth *et al*, 2011). However, use of gene expression data in metabolic models must be done with care and other post translational factors must be taken into consideration. Given that some genes may have better gene expression-protein abundance correlations than others, it may be important to consider clustering genes and utilising gene expression data on gene clusters that have good transcriptomics and metabolomics correlations. Time-shifts may also be applied on gene clusters based on published information on mRNA half-lives (Shock *et al*, 2007). As

metabolomics data may represent enzyme activity better than transcriptomics or proteomics data (Rossell *et al*, 2006; Yizhak *et al*, 2010), experimental measurements of metabolite flux of other important boundary metabolites (e.g., adenine, adenosine, guanine, inosine, hypoxanthine and xanthine) may be useful in adding constraints onto the purine salvage pathway in the model. Nevertheless, the proper utilisation of transcriptomics, proteomics and metabolomics data altogether may prove to be a valuable technique for improving the accuracy of metabolic models.

In vitro quantification of carbohydrates and lipids in the parasite biomass to update the biomass function may also increase the accuracy of model predictions. Furthermore, quantification of ATP, NADH and NADPH production/consumption may be done and considered as an additional or alternative objective function to model energy consumption and maintenance of redox balance. Published data on *in vitro* intracellular metabolite concentrations from malaria parasite cell extracts (Olszewski *et al*, 2009) may be used to calculate metabolite flux and these values may be incorporated into the model as additional constraints.

5.4 Validation of novel targets

Eighteen novel targets were identified by the model, and preliminary experimental data shown in Chapter 4 suggests that one of these novel targets, UMP-CMP kinase, is essential for parasite development. Gemcitabine, an antineoplastic agent that was predicted to target UMP-CMP kinase, halted the development of parasites from trophozoites to schizonts. This suggests that the compound is targeting a reaction/reactions that contribute to DNA replication which includes UMP-CMP kinase. Additional confirmatory tests are required to ascertain that this target is indeed essential to the parasite. *In vitro* gene knockout may be the most definitive way of validating the essentiality of the said gene. The CRISPR-Cas9 system as well as the use of alternative protein-RNA complexes (e.g., Cpf1) have been successful in knocking out genes in *P. falciparum* and *P. berghei* (Wagner *et al*, 2014; Ghorbal *et al*, 2014; Carrasquilla & Owusu, 2016; Zetsche *et al*, 2015) and such technique may be utilised to confirm the importance of UMP-CMP kinase in parasite growth. Testing the compound against purified parasite protein may also be useful in providing supporting evidence that gemcitabine targets the said kinase; but it is important to remember that gemcitabine requires phosphorylation by ribonucleotide reductase into an active metabolite that then targets UMP-CMP kinase (Hsu *et al*, 2005; Mini *et al*, 2006).

Similarly, other predicted novel targets may be validated using the previously mentioned techniques. Taking into consideration the cost and time involved in validating these targets, prioritisation is essential. One way is to concentrate on novel targets that are consistent with the predictions of the other malaria models. Another way is to look into gene expression data

on these novel targets and selecting those that are highly and/or constitutively expressed throughout the parasite life cycle (Le Roch *et al*, 2003). In light of finding druggable targets, novel targets may also be ranked based on having the least homology to human enzymes (Yeh *et al*, 2004). Other databases similar to DrugBank, such as STITCH (Kuhn *et al*, 2014), may be exploited to identify compounds that may be used for *in vitro* growth inhibition assays.

5.5   Conclusion

This study demonstrates the value of extensive curation and utilisation of parasite-specific constraints in the improvement of existing metabolic network models. A highly curated metabolic model of *Plasmodium falciparum*, iFT342, was developed by taking advantage of already existing malaria metabolic models in conjunction with the extensive use of available databases containing information on *P. falciparum* and associated reactions and metabolites to ensure correctness of genes, reactions and species incorporated into the model. The iFT342 has updated gene and reaction annotations as well as additional species identifiers that will facilitate ease in comparison with other models. The model has zero dead-end metabolites and has the highest percentage of live reactions (87.5%) compared with highly curated models such as the human (71.5%), *E. coli* (64.3%) and yeast (56.1%) models. With the addition of experimentally measured biomass composition (Forth, 2012) and metabolite fluxes for glucose and 18 amino acids, iFT342 was able to model *in vitro* parasite growth in restricted glucose environment with remarkable fidelity (r = 0.985, p-value = 1.36×10^{-6}). Through single gene knockout analysis, the model was able to significantly enrich the total number of experimentally validated essential (gold standard) genes (PPV = 76.6%, enrichment score = 1.44, p-value = 1.69×10^{-6}). Finally, as proof of concept, inhibition of parasite growth was demonstrated using gemcitabine, which was predicted to target UMP-CMP kinase, a novel target predicted by the model. Gemcitabine had an IC_{50} in the low micromolar range (2.62 ± 0.53 μM) and was shown to inhibit development of the parasite from the trophozoite to the schizont stage.

References

Achard F, Vaysseix G & Barillot E (2001) XML, bioinformatics and data integration. *Bioinformatics* **17:** 115–25

Ahn S-Y, Shin M-Y, Kim Y-A, Yoo J-A, Kwak D-H, Jung Y-J, Jun G, Ryu S-H, Yeom J-S, Ahn J-Y, Chai J-Y & Park J-W (2008) Magnetic separation: a highly effective method for synchronization of cultured erythrocytic Plasmodium falciparum. *Parasitol. Res.* **102:** 1195–200

Akkemik E, Budak H & Ciftci M (2010) Effects of some drugs on human erythrocyte 6-phosphogluconate dehydrogenase: an in vitro study. *J. Enzyme Inhib. Med. Chem.* **25:** 476–9

Alker AP, Lim P, Sem R, Shah NK, Yi P, Bouth DM, Tsuyuoka R, Maguire JD, Fandeur T, Ariey F, Wongsrichanalai C & Meshnick SR (2007) Pfmdr1 and in vivo resistance to artesunate-mefloquine in falciparum malaria on the Cambodian-Thai border. *Am. J. Trop. Med. Hyg.* **76:** 641–7

Allen RJW & Kirk K (2004) Cell volume control in the Plasmodium-infected erythrocyte. *Trends Parasitol.* **20:** 7-10–1

Allen SM, Lim EE, Jortzik E, Preuss J, Chua HH, MacRae JI, Rahlfs S, Haeussler K, Downton MT, McConville MJ, Becker K & Ralph SA (2015) Plasmodium falciparum glucose-6-phosphate dehydrogenase 6-phosphogluconolactonase is a potential drug target. *FEBS J.* **282:** 3808–23

Allman EL, Painter HJ, Samra J, Carrasquilla M & Llinás M (2016) Metabolomic Profiling of the Malaria Box Reveals Antimalarial Target Pathways. *Antimicrob. Agents Chemother.* **60:** 6635–6649

Alvarez-Coque MCG, Hernández MJM, Camañas RMV & Fernández CM (1988) Some observations on the reaction of cysteine with o-phthalaldehyde. *Spectrochim. Acta Part A Mol. Spectrosc.* **44A:** 1461–1464

Arisue N & Hashimoto T (2015) Phylogeny and evolution of apicoplasts and apicomplexan parasites. *Parasitol. Int.* **64:** 254–9

Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, Grosdidier A, Hernandez C, Ioannidis V, Kuznetsov D, Liechti R, Moretti S, Mostaguir K, Redaschi N, Rossier G, Xenarios I, et al (2012) ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* **40:** W597-603

Asahi H & Kanazawa T (1994) Continuous cultivation of intraerythrocytic Plasmodium falciparum in a serum-free medium with the use of a growth-promoting factor. *Parasitology* **109:** 397–401

Asante KP, Abokyi L, Zandoh C, Owusu R, Awini E, Sulemana A, Amenga-Etego S, Adda R, Boahen O, Segbaya S, Mahama E, Bart-Plange C, Chandramohan D & Owusu-Agyei S (2010) Community perceptions of malaria and malaria treatment behaviour in a rural district of Ghana: implications for artemisinin combination therapy. *BMC Public Health* **10:** 409

Aurrecoechea C, Barreto A, Brestelli J, Brunk BP, Cade S, Doherty R, Fischer S, Gajria B, Gao X, Gingle A, Grant G, Harb OS, Heiges M, Hu S, Iodice J, Kissinger JC, Kraemer ET, Li W, Pinney DF, Pitts B, et al (2013) EuPathDB: the eukaryotic pathogen database. *Nucleic Acids Res.* **41:** D684-91

Aurrecoechea C, Brestelli J, Brunk BP, Dommer J, Fischer S, Gajria B, Gao X, Gingle A, Grant G, Harb OS, Heiges M, Innamorato F, Iodice J, Kissinger JC, Kraemer ET, Li W, Miller JA, Nayak V, Pennington C, Pinney DF, et al (2009) PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res.* **37:** D539-43

Aziz RK, Devoid S, Disz T, Edwards RA, Henry CS, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Stevens RL, Vonstein V & Xia F (2012) SEED servers: high-performance access to the SEED genomes, annotations, and metabolic models. *PLoS One* **7:** e48053

Babbitt SE, Altenhofen L, Cobbold SA, Istvan ES, Fennell C, Doerig C, Llinás M & Goldberg DE

(2012) Plasmodium falciparum responds to amino acid starvation by entering into a hibernatory state. *Proc. Natl. Acad. Sci. U. S. A.* **109:** E3278-87

Bain BJB (2006) Blood cells: a practical guide 4th ed. John Wiley & Sons

Bardou P, Mariette J, Escudié F, Djemiel C & Klopp C (2014) jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics* **15:** 293

Barrett MP (1997) The pentose phosphate pathway and parasitic protozoa. *Parasitol. Today* **13:** 11–6

Bártfai R, Hoeijmakers WAM, Salcedo-Amaya AM, Smits AH, Janssen-Megens E, Kaan A, Treeck M, Gilberger T-W, Françoijs K-J & Stunnenberg HG (2010) H2A.Z demarcates intergenic regions of the plasmodium falciparum epigenome that are dynamically marked by H3K9ac and H3K4me3. *PLoS Pathog.* **6:** e1001223

Bartolomeo MP & Maisano F (2006) Validation of a reversed-phase HPLC method for quantitative amino acid analysis. *J. Biomol. Tech.* **17:** 131–7

Bass CC & Johns FM (1912) The cultivation of malarial Plasmodia (Plasmodium vivax and Plasmodium falciparum) in vitro. *J. Exp. Med.* **16:** 567–79

Bastian M, Heymann S & Jacomy M (2009) Gephi: An Open Source Software for Exploring and Manipulating Networks. *Third Int. AAAI Conf. Weblogs Soc. Media***:** 361–362

Bazzani S, Hoppe A & Holzhütter H-G (2012) Network-based assessment of the selectivity of metabolic drug targets in Plasmodium falciparum with respect to human liver metabolism. *BMC Syst. Biol.* **6:** 118

Bernard T, Bridge A, Morgat A, Moretti S, Xenarios I & Pagni M (2014) Reconciliation of metabolites and biochemical reactions for metabolic networks. *Brief. Bioinform.* **15:** 123–35

Birwé H & Hesse A (1991) High-performance liquid chromatographic determination of urinary cysteine and cystine. *Clin. Chim. Acta.* **199:** 33–42

Blume M, Hliscs M, Rodriguez-Contreras D, Sanchez M, Landfear S, Lucius R, Matuschewski K & Gupta N (2011) A constitutive pan-hexose permease for the Plasmodium life cycle and transgenic models for screening of antimalarial sugar analogs. *FASEB J.* **25:** 1218–29

Bolton EE, Wang Y, Thiessen PA & Bryant SH (2008) PubChem: integrated platform of small molecules and biological activities. *Annu. Rep. Comput. Chem.* **4:** 217–241

Bonate PL, Arthaud L, Cantrell WR, Stephenson K, Secrist JA & Weitman S (2006) Discovery and development of clofarabine: a nucleoside analogue for treating cancer. *Nat. Rev. Drug Discov.* **5:** 855–63

Bornstein BJ, Keating SM, Jouraku A & Hucka M (2008) LibSBML: an API library for SBML. *Bioinformatics* **24:** 880–1

Botté CY, Yamaryo-Botté Y, Rupasinghe TWT, Mullin KA, MacRae JI, Spurck TP, Kalanon M, Shears MJ, Coppel RL, Crellin PK, Maréchal E, McConville MJ & McFadden GI (2013) Atypical lipid composition in the purified relict plastid (apicoplast) of malaria parasites. *Proc. Natl. Acad. Sci. U. S. A.* **110:** 7506–11

Bouyer G, Egée S & Thomas SL (2006) Three types of spontaneously active anionic channels in malaria-infected human red blood cells. *Blood Cells. Mol. Dis.* **36:** 248–54

Bozdech Z & Ginsburg H (2005) Data mining of the transcriptome of Plasmodium falciparum: the pentose phosphate pathway and ancillary processes. *Malar. J.* **4:** 17

Bozdech Z, Llinás M, Pulliam BL, Wong ED, Zhu J & DeRisi JL (2003) The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum. *PLoS Biol.* **1:** E5

Bruggeman FJ & Westerhoff H V. (2007) The nature of systems biology. *Trends Microbiol.* **15:** 45–50

Budavari S, O'Neil MJ, Smith A & Heckelman PE (1989) The Merck Index Rahway, New Jersey: Merck

Carrasquilla M & Owusu CK (2016) A CRISPR outlook for apicomplexans. *Nat. Rev. Microbiol.* **14:** 668

Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA, Holland TA, Keseler IM, Kothari

A, Kubo A, Krummenacker M, Latendresse M, Mueller LA, Ong Q, Paley S, Subhraveti P, Weaver DS, Weerasinghe D, Zhang P & Karp PD (2014) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.* **42:** D459-71

Chan XWA, Wrenger C, Stahl K, Bergmann B, Winterberg M, Müller IB & Saliba KJ (2013) Chemical and genetic validation of thiamine utilization as an antimalarial drug target. *Nat. Commun.* **4:** 2060

Chanda I, Pan A & Dutta C (2005) Proteome composition in Plasmodium falciparum: higher usage of GC-rich nonsynonymous codons in highly expressed genes. *J. Mol. Evol.* **61:** 513–23

Chaouiya C, Keating SM, Berenguier D, Naldi A, Thieffry D, van Iersel MP, Le Novère N & Helikar T (2015) The Systems Biology Markup Language (SBML) Level 3 Package: Qualitative Models, Version 1, Release 1. *J. Integr. Bioinform.* **12:** 270

Chaudhary K, Darling JA, Fohl LM, Sullivan WJ, Donald RGK, Pfefferkorn ER, Ullman B & Roos DS (2004) Purine salvage pathways in the apicomplexan parasite Toxoplasma gondii. *J. Biol. Chem.* **279:** 31221–7

Chavali AK, Blazier AS, Tlaxca JL, Jensen PA, Pearson RD & Papin JA (2012a) Metabolic network analysis predicts efficacy of FDA-approved drugs targeting the causative agent of a neglected tropical disease. *BMC Syst. Biol.* **6:** 27

Chavali AK, D'Auria KM, Hewlett EL, Pearson RD & Papin JA (2012b) A metabolic network approach for the identification and prioritization of antimicrobial drug targets. *Trends Microbiol.* **20:** 113–23

Chavali AK, Whittemore JD, Eddy JA, Williams KT & Papin JA (2008) Systems analysis of metabolism in the pathogenic trypanosomatid Leishmania major. *Mol. Syst. Biol.* **4:** 177

Chong CR, Chen X, Shi L, Liu JO & Sullivan DJ (2006) A clinical drug library screen identifies astemizole as an antimalarial agent. *Nat. Chem. Biol.* **2:** 415–6

Cielecka-Piontek J, Zalewski P, Jelińska A & Garbacki P (2013) UHPLC: The greening face of liquid chromatography. *Chromatographia* **76:** 1429–1437

Clarke JL, Scopes DA, Sodeinde O & Mason PJ (2001) Glucose-6-phosphate dehydrogenase-6-phosphogluconolactonase. A novel bifunctional enzyme in malaria parasites. *Eur. J. Biochem.* **268:** 2013–9

Cobbold SA, Chua HH, Nijagal B, Creek DJ, Ralph SA & McConville MJ (2016) Metabolic Dysregulation Induced in Plasmodium falciparum by Dihydroartemisinin and Other Front-Line Antimalarial Drugs. *J. Infect. Dis.* **213:** 276–86

Combrinck JM, Mabotha TE, Ncokazi KK, Ambele MA, Taylor D, Smith PJ, Hoppe HC & Egan TJ (2013) Insights into the role of heme in the mechanism of action of antimalarials. *ACS Chem. Biol.* **8:** 133–7

Cottret L, Wildridge D, Vinson F, Barrett MP, Charles H, Sagot M-F & Jourdan F (2010) MetExplore: a web server to link metabolomic experiments and genome-scale metabolic networks. *Nucleic Acids Res.* **38:** W132-7

Cowman AF & Crabb BS (2006) Invasion of red blood cells by malaria parasites. *Cell* **124:** 755–66

Cranmer SL, Magowan C, Liang J, Coppel RL & Cooke BM (1995) An alternative to serum for cultivation of Plasmodium falciparum in vitro. *Trans. R. Soc. Trop. Med. Hyg.* **91:** 363–5

Creek DJ, Chua HH, Cobbold SA, Nijagal B, MacRae JI, Dickerman BK, Gilson PR, Ralph SA & McConville MJ (2016) Metabolomics-Based Screening of the Malaria Box Reveals both Novel and Established Mechanisms of Action. *Antimicrob. Agents Chemother.* **60:** 6650–6663

Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, Caudy M, Garapati P, Gillespie M, Kamdar MR, Jassal B, Jupe S, Matthews L, May B, Palatnik S, Rothfels K, Shamovsky V, Song H, Williams M, Birney E, et al (2014) The Reactome pathway knowledgebase. *Nucleic Acids Res.* **42:** D472-7

Dar O, Khan MS & Adagu I (2008) The potential use of methotrexate in the treatment of falciparum malaria: in vitro assays against sensitive and multidrug-resistant falciparum

strains. *Jpn. J. Infect. Dis.* **61:** 210–1

Degtyarenko K, de Matos P, Ennis M, Hastings J, Zbinden M, McNaught AD, Alcántara R, Darsow M, Guedj M & Ashburner M (2008) ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Res.* **36:** D344-50

Dei-Cas E, Wattez A & Vernes A (1985) Influence of medium osmolality on the in vitro growth of Plasmodium falciparum: a morphologic and radioisotopic study. *In Vitro Cell. Dev. Biol.* **21:** 93–8

Denis MB, Tsuyuoka R, Poravuth Y, Narann TS, Seila S, Lim C, Incardona S, Lim P, Sem R, Socheat D, Christophel EM & Ringwald P (2006) Surveillance of the efficacy of artesunate and mefloquine combination for the treatment of uncomplicated falciparum malaria in Cambodia. *Trop. Med. Int. Health* **11:** 1360–6

DeWalski K, Watermann H, Kapp L & Zimring JC (2015) Donor genetics, biology, and metabolomics of blood storage. *Pathology* **47:** S35

Divo AA, Geary TG, Davis NL & Jensen JB (1985) Nutritional requirements of Plasmodium falciparum in culture. I. Exogenously supplied dialyzable components necessary for continuous growth. *J. Protozool.* **32:** 59–64

Dondorp AM, Nosten F, Yi P, Das D, Phyo AP, Tarning J, Lwin KM, Ariey F, Hanpithakpong W, Lee SJ, Ringwald P, Silamut K, Imwong M, Chotivanich K, Lim P, Herdman T, An SS, Yeung S, Singhasivanon P, Day NPJ, et al (2009) Artemisinin resistance in Plasmodium falciparum malaria. *N. Engl. J. Med.* **361:** 455–67

Dorsey G, Njama D, Kamya MR, Cattamanchi A, Kyabayinze D, Staedke SG, Gasasira A & Rosenthal PJ (2002) Sulfadoxine/pyrimethamine alone or with amodiaquine or artesunate for treatment of uncomplicated malaria: a longitudinal randomised trial. *Lancet (London, England)* **360:** 2031–8

Downie MJ, Kirk K & Mamoun C Ben (2008) Purine salvage pathways in the intraerythrocytic malaria parasite Plasmodium falciparum. *Eukaryot. Cell* **7:** 1231–7

Duarte NC, Herrgård MJ & Palsson BØ (2004) Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res.* **14:** 1298–309

Eastman RT, Pattaradilokrat S, Raj DK, Dixit S, Deng B, Miura K, Yuan J, Tanaka TQ, Johnson RL, Jiang H, Huang R, Williamson KC, Lambert LE, Long C, Austin CP, Wu Y & Su X-Z (2013) A class of tricyclic compounds blocking malaria parasite oocyst development and transmission. *Antimicrob. Agents Chemother.* **57:** 425–35

Ebrahim A, Lerman JA, Palsson BØ & Hyduke DR (2013) COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst. Biol.* **7:** 74

Edwards JS & Palsson BØ (1999) Systems properties of the Haemophilus influenzae Rd metabolic genotype. *J. Biol. Chem.* **274:** 17410–6

Elliott JL, Saliba KJ & Kirk K (2001) Transport of lactate and pyruvate in the intraerythrocytic malaria parasite, Plasmodium falciparum. *Biochem. J.* **355:** 733–9

Fang X, Reifman J & Wallqvist A (2014) Modeling metabolism and stage-specific growth of Plasmodium falciparum HB3 during the intraerythrocytic developmental cycle. *Mol. Biosyst.* **10:** 2526–37

Feachem RGA, Phillips AA, Hwang J, Cotter C, Wielgosz B, Greenwood BM, Sabot O, Rodriguez MH, Abeyasinghe RR, Ghebreyesus TA & Snow RW (2010) Shrinking the malaria map: progress and prospects. *Lancet (London, England)* **376:** 1566–78

Feist AM, Zielinski DC, Orth JD, Schellenberger J, Herrgård MJ & Palsson BØ (2010) Model-driven evaluation of the production potential for growth-coupled products of Escherichia coli. *Metab. Eng.* **12:** 173–86

Fitch CD (2004) Ferriprotoporphyrin IX, phospholipids, and the antimalarial actions of quinoline drugs. *Life Sci.* **74:** 1957–72

Flores M V, Berger-Eiszele SM & Stewart TS (1997) Long-term cultivation of Plasmodium falciparum in media with commercial non-serum supplements. *Parasitol. Res.* **83:** 734–6

Förster J, Famili I, Fu P, Palsson BØ & Nielsen J (2003) Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network. *Genome Res.* **13:** 244–53

Forth T (2012) Metabolic Systems Biology of the Malaria Parasite (Doctoral Thesis). *Univ. Leeds*

Foth BJ, Stimmler LM, Handman E, Crabb BS, Hodder AN & McFadden GI (2005) The malaria parasite Plasmodium falciparum has only one pyruvate dehydrogenase complex, which is located in the apicoplast. *Mol. Microbiol.* **55:** 39–53

Foth BJ, Zhang N, Chaal BK, Sze SK, Preiser PR & Bozdech Z (2011) Quantitative time-course profiling of parasite and host cell proteins in the human malaria parasite Plasmodium falciparum. *Mol. Cell. Proteomics* **10:** M110.006411

Francis SE, Sullivan DJ & Goldberg DE (1997) Hemoglobin metabolism in the malaria parasite Plasmodium falciparum. *Annu. Rev. Microbiol.* **51:** 97–123

Frevert U (2004) Sneaking in through the back entrance: the biology of malaria liver stages. *Trends Parasitol.* **20:** 417–24

Gardner JJ & Boyle NR (2017) The use of genome-scale metabolic network reconstruction to predict fluxes and equilibrium composition of N-fixing versus C-fixing cells in a diazotrophic cyanobacterium, Trichodesmium erythraeum. *BMC Syst. Biol.* **11:** 4

Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Eisen JA, Rutherford K, Salzberg SL, Craig A, Kyes S, Chan M-S, Nene V, Shallom SJ, et al (2002) Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature* **419:** 498–511

Geary TG, Divo AA, Bonanni LC & Jensen JB (1985) Nutritional requirements of Plasmodium falciparum in culture. III. Further observations on essential nutrients and antimetabolites. *J. Protozool.* **32:** 608–13

Gerald N, Mahajan B & Kumar S (2011) Mitosis in the human malaria parasite Plasmodium falciparum. *Eukaryot. Cell* **10:** 474–82

Gevi F, D'Alessandro A, Rinalducci S & Zolla L (2012) Alterations of red blood cell metabolome during cold liquid storage of erythrocyte concentrates in CPD-SAGM. *J. Proteomics* **76:** 168–80

Gevorgyan A, Bushell ME, Avignone-Rossa C & Kierzek AM (2011) SurreyFBA: a command line tool and graphics user interface for constraint-based modeling of genome-scale metabolic reaction networks. *Bioinformatics* **27:** 433–4

Ghorbal M, Gorman M, Macpherson CR, Martins RM, Scherf A & Lopez-Rubio J-J (2014) Genome editing in the human malaria parasite Plasmodium falciparum using the CRISPR-Cas9 system. *Nat. Biotechnol.* **32:** 819–21

Ginsburg H (2006) Progress in in silico functional genomics: the malaria Metabolic Pathways database. *Trends Parasitol.* **22:** 238–40

Ginsburg H & Abdel-Haleem AM (2016) Malaria Parasite Metabolic Pathways (MPMP) Upgraded with Targeted Chemical Compounds. *Trends Parasitol.* **32:** 7–9

Ginsburg H, Handeli S, Friedman S, Gorodetsky R & Krugliak M (1986a) Effects of red blood cell potassium and hypertonicity on the growth of Plasmodium falciparum in culture. *Z. Parasitenkd.* **72:** 185–99

Ginsburg H, Kutner S, Zangwil M & Cabantchik ZI (1986b) Selectivity properties of pores induced in host erythrocyte membrane by Plasmodium falciparum. Effect of parasite maturation. *Biochim. Biophys. Acta* **861:** 194–6

Ginsburg H & Stein WD (2004) The new permeability pathways induced by the malaria parasite in the membrane of the infected erythrocyte: comparison of results using different experimental techniques. *J. Membr. Biol.* **197:** 113–34

Goheen MM, Clark MA, Kasthuri RS & Cerami C (2016) Biopreservation of RBCs for in vitro Plasmodium falciparum culture. *Br. J. Haematol.* **175:** 741–744

Gomes AR, Bushell E, Schwach F, Girling G, Anar B, Quail MA, Herd C, Pfander C, Modrzynska K, Rayner JC & Billker O (2015) A genome-scale vector resource enables high-throughput reverse genetic screening in a malaria parasite. *Cell Host Microbe* **17:** 404–13

Goodman CD & McFadden GI (2007) Fatty acid biosynthesis as a drug target in apicomplexan parasites. *Curr. Drug Targets* **8:** 15–30

Gornicki P (2003) Apicoplast fatty acid biosynthesis as a target for medical intervention in apicomplexan parasites. *Int. J. Parasitol.* **33:** 885–96

Greenbaum DC, Mackey Z, Hansell E, Doyle P, Gut J, Caffrey CR, Lehrman J, Rosenthal PJ, McKerrow JH & Chibale K (2004) Synthesis and structure-activity relationships of parasiticidal thiosemicarbazone cysteine protease inhibitors against Plasmodium falciparum, Trypanosoma brucei, and Trypanosoma cruzi. *J. Med. Chem.* **47:** 3212–9

Guiguemde WA, Shelat AA, Garcia-Bustos JF, Diagana TT, Gamo F-J & Guy RK (2012) Global phenotypic screening for antimalarials. *Chem. Biol.* **19:** 116–29

Hartwig CL, Rosenthal AS, D'Angelo J, Griffin CE, Posner GH & Cooper RA (2009) Accumulation of artemisinin trioxane derivatives within neutral lipids of Plasmodium falciparum malaria parasites is endoperoxide-dependent. *Biochem. Pharmacol.* **77:** 322–36

Hastings J, de Matos P, Dekker A, Ennis M, Harsha B, Kale N, Muthukrishnan V, Owen G, Turner S, Williams M & Steinbeck C (2013) The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. *Nucleic Acids Res.* **41:** D456-63

Heikkilä T, Ramsey C, Davies M, Galtier C, Stead AMW, Johnson AP, Fishwick CWG, Boa AN & McConkey GA (2007) Design and synthesis of potent inhibitors of the malaria parasite dihydroorotate dehydrogenase. *J. Med. Chem.* **50:** 186–91

Heller SR & McNaught AD (2009) The IUPAC international chemical identifier (InChI). *Chem. Int.* **31:** 7

Heller SR, McNaught AD, Pletnev I, Stein S & Tchekhovskoi D (2015) InChI, the IUPAC International Chemical Identifier. *J. Cheminform.* **7:** 23

Herrgård MJ, Swainston N, Dobson PD, Dunn WB, Arga KY, Arvas M, Blüthgen N, Borger S, Costenoble R, Heinemann M, Hucka M, Le Novère N, Li P, Liebermeister W, Mo ML, Oliveira AP, Petranovic D, Pettifer S, Simeonidis E, Smallbone K, et al (2008) A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat. Biotechnol.* **26:** 1155–60

Hicks ND & Schall JJ (2014) Dynamics of clonal diversity in natural infections of the malaria parasite Plasmodium mexicanum in its free-ranging lizard host. *Parasitol. Res.* **113:** 2059–67

Hsu C-H, Liou J-Y, Dutschman GE & Cheng Y-C (2005) Phosphorylation of Cytidine, Deoxycytidine, and Their Analog Monophosphates by Human UMP/CMP Kinase Is Differentially Regulated by ATP and Magnesium. *Mol. Pharmacol.* **67:** 806–14

Huang DW, Sherman BT & Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37:** 1–13

Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, Cuellar AA, Dronov S, Gilles ED, Ginkel M, Gor V, Goryanin II, Hedley WJ, Hodgman TC, Hofmeyr J-H, Hunter PJ, et al (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19:** 524–31

Huthmacher C, Hoppe A, Bulik S & Holzhütter H-G (2010) Antimalarial drug targets in Plasmodium falciparum predicted by stage-specific metabolic network analysis. *BMC Syst. Biol.* **4:** 120

Hyde JE (2007) Targeting purine and pyrimidine metabolism in human apicomplexan parasites. *Curr. Drug Targets* **8:** 31–47

Jensen MD, Conley M & Helstowski LD (1983) Culture of Plasmodium falciparum: the role of pH, glucose, and lactate. *J. Parasitol.* **69:** 1060–7

Joet T, Eckstein-Ludwig U, Morin C & Krishna S (2003) Validation of the hexose transporter of Plasmodium falciparum as a novel drug target. *Proc. Natl. Acad. Sci. U. S. A.* **100:** 7476–9

Joët T & Krishna S (2004) The hexose transporter of Plasmodium falciparum is a worthy drug target. *Acta Trop.* **89:** 371–374

Jones E, Oliphant T & Peterson P (2015) SciPy: Open source scientific tools for Python, 2001. *URL http//www. scipy. org* **73:** 86

Jortzik E, Mailu BM, Preuss J, Fischer M, Bode L, Rahlfs S & Becker K (2011) Glucose-6-phosphate dehydrogenase-6-phosphogluconolactonase: a unique bifunctional enzyme from Plasmodium falciparum. *Biochem. J.* **436:** 641–50

Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T & Yamanishi Y (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **36:** D480-4

Kanehisa M & Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28:** 27–30

Kanehisa M, Sato Y, Kawashima M, Furumichi M & Tanabe M (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44:** D457-62

Kaspar H, Dettmer K, Gronwald W & Oefner PJ (2009) Advances in amino acid analysis. *Anal. Bioanal. Chem.* **393:** 445–52

Ke H, Lewis IA, Morrisey JM, McLean KJ, Ganesan SM, Painter HJ, Mather MW, Jacobs-Lorena M, Llinás M & Vaidya AB (2015) Genetic investigation of tricarboxylic acid metabolism during the Plasmodium falciparum life cycle. *Cell Rep.* **11:** 164–74

Kell DB & Goodacre R (2014) Metabolomics and systems pharmacology: why and how to model the human metabolic network for drug discovery. *Drug Discov. Today* **19:** 171–82

Khodayari A, Zomorrodi AR, Liao JC & Maranas CD (2014) A kinetic model of Escherichia coli core metabolism satisfying multiple sets of mutant flux data. *Metab. Eng.* **25:** 50–62

Kim TY, Sohn SB, Kim Y Bin, Kim WJ & Lee SY (2012) Recent advances in reconstruction and applications of genome-scale metabolic models. *Curr. Opin. Biotechnol.* **23:** 617–23

Kirk K, Horner HA & Kirk J (1996) Glucose uptake in Plasmodium falciparum-infected erythrocytes is an equilibrative not an active process. *Mol. Biochem. Parasitol.* **82:** 195–205

Kirk K & Lehane AM (2014) Membrane transport in the malaria parasite and its host erythrocyte. *Biochem. J.* **457:** 1–18

Kohl M, Wiese S & Warscheid B (2011) Cytoscape: software for visualization and analysis of biological networks. *Methods Mol. Biol.* **696:** 291–303

Krishna S, Woodrow CJ, Burchmore RJ, Saliba KJ & Kirk K (2000) Hexose transport in asexual stages of Plasmodium falciparum and kinetoplastidae. *Parasitol. Today* **16:** 516–21

Krugliak M, Zhang J & Ginsburg H (2002) Intraerythrocytic Plasmodium falciparum utilizes only a fraction of the amino acids derived from the digestion of host cell cytosol for the biosynthesis of its proteins. *Mol. Biochem. Parasitol.* **119:** 249–56

Krungkrai SR & Krungkrai J (2016) Insights into the pyrimidine biosynthetic pathway of human malaria parasite Plasmodium falciparum as chemotherapeutic target. *Asian Pac. J. Trop. Med.* **9:** 525–34

Kuhn M, Szklarczyk D, Pletscher-Frankild S, Blicher TH, von Mering C, Jensen LJ & Bork P (2014) STITCH 4: integration of protein-chemical interactions with user data. *Nucleic Acids Res.* **42:** D401-7

Kumar A, Suthers PF & Maranas CD (2012) MetRxn: a knowledgebase of metabolites and reactions spanning metabolic models and databases. *BMC Bioinformatics* **13:** 6

Kwiatkowski DP (1989) Febrile temperatures can synchronize the growth of Plasmodium falciparum in vitro. *J. Exp. Med.* **169:** 357–61

Lambros C & Vanderberg JP (1979) Synchronization of Plasmodium falciparum erythrocytic stages in culture. *J. Parasitol.* **65:** 418–20

Lang M, Stelzer M & Schomburg D (2011) BKM-react, an integrated biochemical reaction database. *BMC Biochem.* **12:** 42

Latendresse M, Krummenacker M, Trupp M & Karp PD (2012) Construction and completion of flux balance models from pathway databases. *Bioinformatics* **28:** 388–96

Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, Maciejewski A, Arndt D, Wilson M, Neveu

V, Tang A, Gabriel G, Ly C, Adamjee S, Dame ZT, Han B, Zhou Y & Wishart DS (2014) DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* **42:** D1091-7

Lee D, Smallbone K, Dunn WB, Murabito E, Winder CL, Kell DB, Mendes P & Swainston N (2012) Improving metabolic flux predictions using absolute gene expression data. *BMC Syst. Biol.* **6:** 73

LeRoux M, Lakshmanan V & Daily JP (2009) Plasmodium falciparum biology: analysis of in vitro versus in vivo growth conditions. *Trends Parasitol.* **25:** 474–81

Lew VL, Tiffert T & Ginsburg H (2003) Excess hemoglobin digestion and the osmotic stability of Plasmodium falciparum-infected red blood cells. *Blood* **101:** 4189–94

Li W, Mo W, Shen D, Sun L, Wang J, Lu S, Gitschier JM & Zhou B (2005) Yeast model uncovers dual roles of mitochondria in action of artemisinin. *PLoS Genet.* **1:** e36

Li Z, Wang R-S & Zhang X-S (2011) Two-stage flux balance analysis of metabolic networks for drug target identification. *BMC Syst. Biol.* **5 Suppl 1:** S11

Liu JO, Istvan ES, Gluzman IY, Gross J & Goldberg DE (2006) Plasmodium falciparum ensures its amino acid supply with multiple acquisition pathways and redundant proteolytic enzyme systems. *Proc. Natl. Acad. Sci. U. S. A.* **103:** 8840–5

Logan-Klumpler FJ, de Silva N, Boehme U, Rogers MB, Velarde G, McQuillan JA, Carver T, Aslett M, Olsen C, Subramanian S, Phan I, Farris C, Mitra S, Ramasamy G, Wang H, Tivey A, Jackson A, Houston R, Parkhill J, Holden M, et al (2012) GeneDB--an annotation database for pathogens. *Nucleic Acids Res.* **40:** D98-108

Lutz M (2013) Learning python O'Reilly Media, Inc

Magariños MP, Carmona SJ, Crowther GJ, Ralph SA, Roos DS, Shanmugam D, van Voorhis WC & Agüero F (2012) TDR Targets: a chemogenomics resource for neglected diseases. *Nucleic Acids Res.* **40:** D1118-27

Magrane M & UniProt Consortium (2011) UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford).* **2011:** bar009

Makhorin A (2008) GLPK (GNU linear programming kit).

Martin RE & Kirk K (2004) The malaria parasite's chloroquine resistance transporter is a member of the drug/metabolite transporter superfamily. *Mol. Biol. Evol.* **21:** 1938–49

Mazumdar J & Striepen B (2007) Make it or take it: fatty acid metabolism of apicomplexan parasites. *Eukaryot. Cell* **6:** 1727–35

McNamara JM & Stearne DJ (2010) Flexible nonlinear periodization in a beginner college weight training class. *J. strength Cond. Res.* **24:** 17–22

McRobert L & McConkey GA (2002) RNA interference (RNAi) inhibits growth of Plasmodium falciparum. *Mol. Biochem. Parasitol.* **119:** 273–278

Medeiros MCI, Hamer GL & Ricklefs RE (2013) Host compatibility rather than vector-host-encounter rate determines the host range of avian Plasmodium parasites. *Proceedings. Biol. Sci.* **280:** 20122947

Mi-Ichi F, Kano S & Mitamura T (2007) Oleic acid is indispensable for intraerythrocytic proliferation of Plasmodium falciparum. *Parasitology* **134:** 1671–1677

Milani KJ, Schneider TG & Taraschi TF (2015) Defining the morphology and mechanism of the hemoglobin transport pathway in Plasmodium falciparum-infected erythrocytes. *Eukaryot. Cell* **14:** 415–26

Milner E, Sousa J, Pybus B, Auschwitz J, Caridha D, Gardner S, Grauer K, Harris E, Hickman M, Kozar MP, Lee P, Leed S, Li Q, Melendez V, Moon J, Ngundam F, O'Neil M, Parriott S, Potter B, Sciotti R, et al (2012) Ketotifen is an antimalarial prodrug of norketotifen with blood schizonticidal and liver-stage efficacy. *Eur. J. Drug Metab. Pharmacokinet.* **37:** 17–22

Mini E, Nobili S, Caciagli B, Landini I & Mazzei T (2006) Cellular pharmacology of gemcitabine. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **17 Suppl 5:** v7-12

Miotto O, Almagro-Garcia J, Manske M, Macinnis B, Campino S, Rockett KA, Amaratunga C, Lim P, Suon S, Sreng S, Anderson JM, Duong S, Nguon C, Chuor CM, Saunders D, Se Y, Lon C, Fukuda MM, Amenga-Etego L, Hodgson AVO, et al (2013) Multiple populations of

artemisinin-resistant Plasmodium falciparum in Cambodia. *Nat. Genet.* **45:** 648–55

Mita T, Tanabe K & Kita K (2009) Spread and evolution of Plasmodium falciparum drug resistance. *Parasitol. Int.* **58:** 201–9

Molnár-Perl I (2001) Derivatization and chromatographic behavior of the o-phthaldialdehyde amino acid derivatives obtained with various SH-group-containing additives. *J. Chromatogr. A* **913:** 283–302

Monk J, Nogales J & Palsson BØ (2014) Optimizing genome-scale network reconstructions. *Nat. Biotechnol.* **32:** 447–52

Mons B, Janse CJ, Boorsma EG & van der Kaay HJ (1985) Synchronized erythrocytic schizogony and gametocytogenesis of Plasmodium berghei in vivo and in vitro. *Parasitology* **91 (Pt 3):** 423–30

Moore GE & Woods LK (1977) Culture media for human cells—RPMI 1603, RPMI 1634, RPMI 1640 and GEM 1717. *Methods Cell Sci.* **3:** 503–509

Morgat A, Coissac E, Coudert E, Axelsen KB, Keller G, Bairoch A, Bridge A, Bougueleret L, Xenarios I & Viari A (2012) UniPathway: a resource for the exploration and annotation of metabolic pathways. *Nucleic Acids Res.* **40:** D761-9

Mueckler M (1994) Facilitative glucose transporters. *Eur. J. Biochem.* **219:** 713–25

Navid A (2011) Applications of system-level models of metabolism for analysis of bacterial physiology and identification of new drug targets. *Brief. Funct. Genomics* **10:** 354–64

Neue UD (2002) HPLC Troubleshooting Guide Waters Corporation

Nixon GL, Pidathala C, Shone AE, Antoine T, Fisher N, O'Neill PM, Ward SA & Biagini GA (2013) Targeting the mitochondrial electron transport chain of Plasmodium falciparum: new strategies towards the development of improved antimalarials for the elimination era. *Future Med. Chem.* **5:** 1573–91

Nosten F & White NJ (2007) Artemisinin-based combination treatment of falciparum malaria. *Am. J. Trop. Med. Hyg.* **77:** 181–92

O'Neill PM, Barton VE & Ward SA (2010) The molecular mechanism of action of artemisinin--the debate continues. *Molecules* **15:** 1705–21

Oberhardt MA, Palsson BØ & Papin JA (2009) Applications of genome-scale metabolic reconstructions. *Mol. Syst. Biol.* **5:** 320

Ofulla A V, Okoye VC, Khan B, Githure JI, Roberts CR, Johnson AJ & Martin SK (1993) Cultivation of Plasmodium falciparum parasites in a serum-free medium. *Am. J. Trop. Med. Hyg.* **49:** 335–40

Olliaro P (2001) Mode of action and mechanisms of resistance for antimalarial drugs. *Pharmacol. Ther.* **89:** 207–19

Olszewski KL & Llinás M (2011) Central carbon metabolism of Plasmodium parasites. *Mol. Biochem. Parasitol.* **175:** 95–103

Olszewski KL, Mather MW, Morrisey JM, Garcia BA, Vaidya AB, Rabinowitz JD & Llinás M (2010) Branched tricarboxylic acid metabolism in Plasmodium falciparum. *Nature* **466:** 774–8

Olszewski KL, Morrisey JM, Wilinski D, Burns JM, Vaidya AB, Rabinowitz JD & Llinás M (2009) Host-parasite interactions revealed by Plasmodium falciparum metabolomics. *Cell Host Microbe* **5:** 191–9

Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM & Palsson BØ (2011) A comprehensive genome-scale reconstruction of Escherichia coli metabolism--2011. *Mol. Syst. Biol.* **7:** 535

Orth JD, Thiele I & Palsson BØ (2010) What is flux balance analysis? *Nat. Biotechnol.* **28:** 245–8

Overbeek R, Begley T, Butler RM, Choudhuri J V., Chuang H-Y, Cohoon M, de Crécy-Lagard V, Diaz N, Disz T, Edwards RA, Fonstein M, Frank ED, Gerdes S, Glass EM, Goesmann A, Hanson A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, et al (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* **33:** 5691–702

Painter HJ, Morrisey JM, Mather MW & Vaidya AB (2007) Specific role of mitochondrial electron transport in blood-stage Plasmodium falciparum. *Nature* **446:** 88–91

Palsson BØ (2015) Systems biology: constraint-based reconstruction and analysis Cambridge, United Kingdom: University Printing House

Parker MD, Hyde RJ, Yao SY, McRobert L, Cass CE, Young JD, McConkey GA & Baldwin SA (2000) Identification of a nucleoside/nucleobase transporter from Plasmodium falciparum, a novel target for anti-malarial chemotherapy. *Biochem. J.* **349:** 67–75

Patel AP, Staines HM & Krishna S (2008) New antimalarial targets: the example of glucose transport. *Travel Med. Infect. Dis.* **6:** 58–66

Payne SH & Loomis WF (2006) Retention and loss of amino acid biosynthetic pathways based on analysis of whole-genome sequences. *Eukaryot. Cell* **5:** 272–6

Pérez B, Teixeira C, Gomes JRB & Gomes P (2013) Development of Plasmodium falciparum protease inhibitors in the past decade (2002-2012). *Curr. Med. Chem.* **20:** 3049–68

Perkins SL & Schaer J (2016) A Modern Menagerie of Mammalian Malaria. *Trends Parasitol.* **32:** 772–782

Phillips MA, Lotharius J, Marsh K, White J, Dayan A, White KL, Njoroge JW, El Mazouni F, Lao Y, Kokkonda S, Tomchick DR, Deng X, Laird T, Bhatia SN, March S, Ng CL, Fidock D a, Wittlin S, Lafuente-Monasterio M, Benito FJG, et al (2015) A long-duration dihydroorotate dehydrogenase inhibitor (DSM265) for prevention and treatment of malaria. *Sci. Transl. Med.* **7:** 296ra111

Pinney JW, Shirley MW, McConkey GA & Westhead DR (2005) metaSHARK: software for automated metabolic network prediction from DNA sequence and its application to the genomes of Plasmodium falciparum and Eimeria tenella. *Nucleic Acids Res.* **33:** 1399–409

Placzek S, Schomburg I, Chang A, Jeske L, Ulbrich M, Tillack J & Schomburg D (2017) BRENDA in 2017: new perspectives and new tools in BRENDA. *Nucleic Acids Res.* **45:** D380–D388

PlasmoDB (2001) PlasmoDB: An integrative database of the Plasmodium falciparum genome. Tools for accessing and analyzing finished and unfinished sequence data. The Plasmodium Genome Database Collaborative. *Nucleic Acids Res.* **29:** 66–9

Plata G, Hsiao T, Olszewski KL, Llinás M & Vitkup D (2010) Reconstruction and flux-balance analysis of the Plasmodium falciparum metabolic network. *Mol. Syst. Biol.* **6:** 408

Ponce-de-León M, Montero F & Peretó J (2013) Solving gap metabolites and blocked reactions in genome-scale models: application to the metabolic network of Blattabacterium cuenoti. *BMC Syst. Biol.* **7:** 114

Prchal JT & Gregg XT (2005) Red cell enzymes. *Hematol. Am. Soc. Hematol. Educ. Progr.***:** 19–23

Price ND, Reed JL & Palsson BØ (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat. Rev. Microbiol.* **2:** 886–97

Ralph SA, van Dooren GG, Waller RF, Crawford MJ, Fraunholz MJ, Foth BJ, Tonkin CJ, Roos DS & McFadden GI (2004) Tropical infectious diseases: metabolic maps and functions of the Plasmodium falciparum apicoplast. *Nat. Rev. Microbiol.* **2:** 203–16

Ralton JE, Naderer T, Piraino HL, Bashtannyk TA, Callaghan JM & McConville MJ (2003) Evidence that intracellular beta1-2 mannan is a virulence factor in Leishmania parasites. *J. Biol. Chem.* **278:** 40757–63

Ravikrishnan A & Raman K (2015) Critical assessment of genome-scale metabolic networks: the need for a unified standard. *Brief. Bioinform.* **16:** 1057–68

Reed JL (2012) Shrinking the metabolic solution space using experimental datasets. *PLoS Comput. Biol.* **8:** e1002662

Reed JL, Vo TD, Schilling CH & Palsson BØ (2003) An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). *Genome Biol.* **4:** R54

Reitz M, Sacher O, Tarkhov A, Trumbach D & Gasteiger J (2004) Enabling the exploration of biochemical pathways. *Org. Biomol. Chem.* **2:** 3226–37

Robak T, Korycka A & Robak E (2006) Older and new formulations of cladribine. Pharmacology and clinical efficacy in hematological malignancies. *Recent Pat. Anticancer. Drug Discov.* **1:** 23–38

Le Roch KG, Johnson JR, Florens L, Zhou Y, Santrosyan A, Grainger M, Yan SF, Williamson KC,

Holder A a, Carucci DJ, Yates JR & Winzeler E a (2004) Global analysis of transcript and protein levels across the Plasmodium falciparum life cycle. *Genome Res.* **14:** 2308–18

Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, Haynes JD, de la Vega P, Holder AA, Batalov S, Carucci DJ & Winzeler EA (2003) Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* **301:** 1503–8

Rocha I, Maia P, Evangelista P, Vilaça P, Soares S, Pinto JP, Nielsen J, Patil KR, Ferreira EC & Rocha M (2010) OptFlux: an open-source software platform for in silico metabolic engineering. *BMC Syst. Biol.* **4:** 45

Rossell S, van der Weijden CC, Lindenbergh A, van Tuijl A, Francke C, Bakker BM & Westerhoff H V (2006) Unraveling the complexity of flux regulation: a new method demonstrated for nutrient starvation in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.* **103:** 2166–71

Roth E (1990) Plasmodium falciparum carbohydrate metabolism: a connection between host cell and parasite. *Blood Cells* **16:** 453-60–6

Ruwende C & Hill A (1998) Glucose-6-phosphate dehydrogenase deficiency and malaria. *J. Mol. Med. (Berl).* **76:** 581–8

Saliba KJ, Krishna S & Kirk K (2004) Inhibition of hexose transport and abrogation of pH homeostasis in the intraerythrocytic malaria parasite by an O-3-hexose derivative. *FEBS Lett.* **570:** 93–6

Sanz S, Bandini G, Ospina D, Bernabeu M, Mariño K, Fernández-Becerra C & Izquierdo L (2013) Biosynthesis of GDP-fucose and other sugar nucleotides in the blood stages of Plasmodium falciparum. *J. Biol. Chem.* **288:** 16506–17

Sarma PS, Mandal AK & Khamis HJ (1998) Allopurinol as an additive to quinine in the treatment of acute complicated falciparum malaria. *Am. J. Trop. Med. Hyg.* **58:** 454–7

Scheibel LW & Miller JA (1969) Glycolytic and cytochrome oxidase activity in Plasmodia. *Mil. Med.* **134:** 1074–80

Schellenberger J, Park JO, Conrad TM & Palsson BØ (2010) BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics* **11:** 213

Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis NE, Rahmanian S, Kang J, Hyduke DR & Palsson BØ (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat. Protoc.* **6:** 1290–307

Schlitzer M (2007) Malaria chemotherapeutics part I: History of antimalarial drug development, currently used therapeutics, and drugs in clinical development. *ChemMedChem* **2:** 944–86

Schomburg I, Chang A, Ebeling C, Gremse M, Heldt C, Huhn G & Schomburg D (2004) BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res.* **32:** D431-3

Schuster FL (2002) Cultivation of plasmodium spp. *Clin. Microbiol. Rev.* **15:** 355–64

Schuster R (1988) Determination of amino acids in biological, pharmaceutical, plant and food samples by automated precolumn derivatization and high-performance liquid chromatography. *J. Chromatogr.* **431:** 271–84

Schwach F, Bushell E, Gomes AR, Anar B, Girling G, Herd C, Rayner JC & Billker O (2015) PlasmoGEM, a database supporting a community resource for large-scale experimental genetics in malaria parasites. *Nucleic Acids Res.* **43:** D1176-82

Segrè D, Vitkup D & Church GM (2002) Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. U. S. A.* **99:** 15112–7

Shameer S, Logan-Klumpler FJ, Vinson F, Cottret L, Merlet B, Achcar F, Boshart M, Berriman M, Breitling R, Bringaud F, Bütikofer P, Cattanach AM, Bannerman-Chukualim B, Creek DJ, Crouch K, de Koning HP, Denise H, Ebikeme C, Fairlamb AH, Ferguson MAJ, et al (2015) TrypanoCyc: a community-led biochemical pathways database for Trypanosoma brucei. *Nucleic Acids Res.* **43:** D637-44

Shock JL, Fischer KF & DeRisi JL (2007) Whole-genome analysis of mRNA decay in Plasmodium falciparum reveals a global lengthening of mRNA half-life during the intra-erythrocytic development cycle. *Genome Biol.* **8:** R134

Singh N & Siddiqi MI (2017) Computational evaluation of glutamine synthetase as drug target against infectious diseases: molecular modeling, substrate-binding analysis, and molecular dynamics simulation studies. *Med. Chem. Res.* **26:** 450–460

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. *Mol. Microbiol.* **75:** 1402–13

Smilkstein M, Sriwilaijaroen N, Kelly JX, Wilairat P & Riscoe M (2004) Simple and inexpensive fluorescence-based technique for high-throughput antimalarial drug screening. *Antimicrob. Agents Chemother.* **48:** 1803–6

Snounou G, Viriyakosol S, Jarra W, Thaithong S & Brown KN (1993) Identification of the four human malaria parasite species in field samples by the polymerase chain reaction and detection of a high prevalence of mixed infections. *Mol. Biochem. Parasitol.* **58:** 283–92

Spangenberg T, Burrows JN, Kowalczyk P, McDonald S, Wells TNC & Willis P (2013) The open access malaria box: a drug discovery catalyst for neglected diseases. *PLoS One* **8:** e62906

Srivastava IK & Vaidya AB (1999) A mechanism for the synergistic antimalarial action of atovaquone and proguanil. *Antimicrob. Agents Chemother.* **43:** 1334–9

Staines HM, Ashmore S, Felgate H, Moore J, Powell T & Ellory JC (2006) Solute transport via the new permeability pathways in Plasmodium falciparum-infected human red blood cells is not consistent with a simple single-channel model. *Blood* **108:** 3187–94

Steiner F, Fabel S & McLeod F (2009) An automated system for at-line process analysis of biopharmaceutical fermentation reactions. Available at: http://www.dionex.com/en-us/webdocs/70923-lpn_2236-01.pdf

Stincone A, Prigione A, Cramer T, Wamelink MMC, Campbell K, Cheung E, Olin-Sandoval V, Grüning N-M, Krüger A, Tauqeer Alam M, Keller MA, Breitenbach M, Brindle KM, Rabinowitz JD & Ralser M (2015) The return of metabolism: biochemistry and physiology of the pentose phosphate pathway. *Biol. Rev. Camb. Philos. Soc.* **90:** 927–63

Stryer L (1988) Biochemistry 3rd ed.

Swain M (2013) PubChemPy. Available at: https://pypi.python.org/pypi/PubChemPy/1.0

Swainston N, Smallbone K, Hefzi H, Dobson PD, Brewer J, Hancho M, Zielinski DC, Ang KS, Gardiner NJ, Gutierrez JM, Kyriakopoulos S, Lakshmanan M, Li S, Liu JK, Martínez VS, Orellana CA, Quek L-E, Thomas A, Zanghellini J, Borth N, et al (2016) Recon 2.2: from reconstruction to model of human metabolism. *Metabolomics* **12:** 109

Ta TH, Hisam S, Lanza M, Jiram AI, Ismail N & Rubio JM (2014) First case of a naturally acquired human infection with Plasmodium cynomolgi. *Malar. J.* **13:** 68

Tarun AS, Vaughan AM & Kappe SHI (2009) Redefining the role of de novo fatty acid synthesis in Plasmodium parasites. *Trends Parasitol.* **25:** 545–50

Taverne J (2000) Albumax and global warming on the Web. *Parasitol. Today* **16:** 190

Thiele I, Swainston N, Fleming RMT, Hoppe A, Sahoo S, Aurich MK, Haraldsdottir H, Mo ML, Rolfsson O, Stobbe MD, Thorleifsson SG, Agren R, Bölling C, Bordel S, Chavali AK, Dobson PD, Dunn WB, Endler L, Hala D, Hucka M, et al (2013) A community-driven global reconstruction of human metabolism. *Nat. Biotechnol.* **31:** 419–25

Tilley L, Straimer J, Gnädig NF, Ralph SA & Fidock DA (2016) Artemisinin Action and Resistance in Plasmodium falciparum. *Trends Parasitol.* **32:** 682–96

Ting L-M, Shi W, Lewandowicz A, Singh V, Mwakingwe A, Birck MR, Ringia EAT, Bench G, Madrid DC, Tyler PC, Evans GB, Furneaux RH, Schramm VL & Kim K (2005) Targeting a novel Plasmodium falciparum purine recycling pathway with specific immucillins. *J. Biol. Chem.* **280:** 9547–54

Trager W & Jensen JB (2005) Human malaria parasites in continuous culture. 1976. *J. Parasitol.* **91:** 484–6

Treutiger CJ, Scholander C, Carlson J, McAdam KP, Raynes JG, Falksveden L & Wahlgren M (1999) Rouleaux-forming serum proteins are involved in the rosetting of Plasmodium falciparum-infected erythrocytes. *Exp. Parasitol.* **93:** 215–24

Triglia T, Menting JG, Wilson C & Cowman AF (1997) Mutations in dihydropteroate synthase are responsible for sulfone and sulfonamide resistance in Plasmodium falciparum. *Proc. Natl. Acad. Sci. U. S. A.* **94:** 13944–9

Uga H, Kuramori C, Ohta A, Tsuboi Y, Tanaka H, Hatakeyama M, Yamaguchi Y, Takahashi T, Kizaki M & Handa H (2006) A new mechanism of methotrexate action revealed by target screening with affinity beads. *Mol. Pharmacol.* **70:** 1832–9

Vaidya AB & Mather MW (2009) Mitochondrial evolution and functions in malaria parasites. *Annu. Rev. Microbiol.* **63:** 249–67

Vaughan AM, O'Neill MT, Tarun AS, Camargo N, Phuong TM, Aly ASI, Cowman AF & Kappe SHI (2009) Type II fatty acid synthesis is essential only for malaria parasite late liver stage development. *Cell. Microbiol.* **11:** 506–20

Vaughan JA (2007) Population dynamics of Plasmodium sporogony. *Trends Parasitol.* **23:** 63–70

Vivas L, Easton A, Kendrick H, Cameron A, Lavandera J-L, Barros D, de las Heras FG, Brady RL & Croft SL (2005) Plasmodium falciparum: stage specific effects of a selective inhibitor of lactate dehydrogenase. *Exp. Parasitol.* **111:** 105–14

Wagner JC, Platt RJ, Goldfless SJ, Zhang F & Niles JC (2014) Efficient CRISPR-Cas9–mediated genome editing in Plasmodium falciparum. *Nat. Methods* **11:** 915–8

Waldecker M, Dasanna AK, Lansche C, Linke M, Srismith S, Cyrklaff M, Sanchez CP, Schwarz US & Lanzer M (2017) Differential time-dependent volumetric and surface area changes and delayed induction of new permeation pathways in P. falciparum-infected hemoglobinopathic erythrocytes. *Cell. Microbiol.* **19:** 1–13

Wallqvist A, Fang X, Tewari SG, Ye P & Reifman J (2016) Metabolic host responses to malarial infection during the intraerythrocytic developmental cycle. *BMC Syst. Biol.* **10:** 58

Weisman JL, Liou AP, Shelat AA, Cohen FE, Guy RK & DeRisi JL (2006) Searching for new antimalarial therapeutics amongst known drugs (supplemental file: table 1 and 2). *Chem. Biol. Drug Des.* **67:** 409–16

Weiss MM, Oppenheim JD & Vanderberg JP (1981) Plasmodium falciparum: assay in vitro for inhibitors of merozoite penetration of erythrocytes. *Exp. Parasitol.* **51:** 400–7

Wernsdorfer WH (2012) Global challenges of changing epidemiological patterns of malaria. *Acta Trop.* **121:** 158–65

White NJ (2008) Plasmodium knowlesi: the fifth human malaria parasite. *Clin. Infect. Dis.* **46:** 172–3

Wickramasinghe SR, Inglis KA, Urch JE, Müller S, van Aalten DMF & Fairlamb AH (2006) Kinetic, inhibition and structural studies on 3-oxoacyl-ACP reductase from Plasmodium falciparum, a key enzyme in fatty acid biosynthesis. *Biochem. J.* **393:** 447–57

Willet GP & Canfield CJ (1984) Plasmodium falciparum: continuous cultivation of erythrocyte stages in plasma-free culture medium. *Exp. Parasitol.* **57:** 76–80

Woodrow CJ, Burchmore RJ & Krishna S (2000) Hexose permeation pathways in Plasmodium falciparum-infected erythrocytes. *Proc. Natl. Acad. Sci. U. S. A.* **97:** 9931–6

World Health Organization (2014) World malaria report 2013 World Health Organization

World Health Organization (2015) Guidelines for the treatment of malaria 3rd Ed. Geneva: WHO Press

World Health Organization (2016) World Malaria Report 2016 Geneva: World Health Organization

www.plasmogem.sanger.ac.uk Plasmo GEM news and updates. Available at: www.plasmogem.sanger.ac.uk [Accessed March 14, 2017]

www.sbml.org SBML Documents/Specifications. Available at: http://sbml.org/Documents/ Specifications [Accessed December 29, 2016]

Yang YZ, Little B & Meshnick SR (1994) Alkylation of proteins by artemisinin. Effects of heme, pH,

and drug structure. *Biochem. Pharmacol.* **48:** 569–73

Yayon A, Vande Waa JA, Yayon M, Geary TG & Jensen JB (1983) Stage-dependent effects of chloroquine on Plasmodium falciparum in vitro. *J. Protozool.* **30:** 642–7

Yeh I, Hanekamp T, Tsoka S, Karp PD & Altman RB (2004) Computational analysis of Plasmodium falciparum metabolism: organizing genomic information to facilitate drug discovery. *Genome Res.* **14:** 917–24

Yilmaz LS & Walhout AJ (2017) Metabolic network modeling with model organisms. *Curr. Opin. Chem. Biol.* **36:** 32–39

Yizhak K, Benyamini T, Liebermeister W, Ruppin E & Shlomi T (2010) Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. *Bioinformatics* **26:** i255-60

Yu M, Kumar TRS, Nkrumah LJ, Coppi A, Retzlaff S, Li CD, Kelly BJ, Moura PA, Lakshmanan V, Freundlich JS, Valderramos J-C, Vilcheze C, Siedner M, Tsai JH-C, Falkard B, Sidhu ABS, Purcell LA, Gratraud P, Kremer L, Waters AP, et al (2008) The fatty acid biosynthesis enzyme FabI plays a key role in the development of liver-stage malarial parasites. *Cell Host Microbe* **4:** 567–78

Zarchin S, Krugliak M & Ginsburg H (1986) Digestion of the host erythrocyte by malaria parasites is the primary target for quinoline-containing antimalarials. *Biochem. Pharmacol.* **35:** 2435–42

Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, van Der Oost J, Regev A, Koonin E V. & Zhang F (2015) Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **163:** 759–771

Zhang C & Hua Q (2015) Applications of Genome-Scale Metabolic Models in Biotechnology and Systems Medicine. *Front. Physiol.* **6:** 413

Appendix   Original set of experimentally validated essential *Plasmodium* spp. genes and associated reactions

Gene ID	Reaction/Gene name	EC number	References
PF3D7_0104400	4-hydroxy-3-methylbut-2-enyl diphosphate reductase (LytB)	1.17.1.2	Vinayak, S., & Sharma, Y. D. (2007). Inhibition of Plasmodium falciparum ispH (lytB) gene expression by hammerhead ribozyme. Oligonucleotides, 17(2), 189-200.
PF3D7_0106300	Calcium-transporting ATPase, putative	3.6.3.8	Eckstein-Ludwig, U., Webb, R. J., Van Goethem, I. D. A., East, J. M., Lee, A. G., Kimura, M., ... & Krishna, S. (2003). Artemisinins target the SERCA of Plasmodium falciparum. *Nature*, *424*(6951), 957-961.
PF3D7_0106900	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase, putative (IspD)	2.7.7.60	Wu, W., Herrera, Z., Ebert, D., Baska, K., Cho, S. H., DeRisi, J. L., & Yeh, E. (2015). A chemical rescue screen identifies a Plasmodium falciparum apicoplast inhibitor targeting MEP isoprenoid precursor biosynthesis. Antimicrobial agents and chemotherapy, 59(1), 356-364.
PF3D7_0204700	Hexose transporter		Woodrow, C. J., Penny, J. I., & Krishna, S. (1999). Intraerythrocytic Plasmodium falciparum expresses a high affinity facilitative hexose transporter. *Journal of Biological Chemistry*, *274*(11), 7272-7277.
PF3D7_0206300	3-Phosphoshikimate 1-carboxyvinyltransferase	2.5.1.19	Roberts, F., Roberts, C. W., Johnson, J. J., Kyle, D. E., Krell, T., Coggins, J. R., ... & Chakrabarti, D. (1998). Evidence for the shikimate pathway in apicomplexan parasites. Nature, 393(6687), 801-805.
PF3D7_0209300	2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	4.6.1.12	Crane, C. M., Kaiser, J., Ramsden, N. L., Lauw, S., Rohdich, F., Eisenreich, W., ... & Diederich, F. (2006). Fluorescent Inhibitors for IspF, an Enzyme in the Non-Mevalonate Pathway for Isoprenoid Biosynthesis and a Potential Target for Antimalarial Therapy. Angewandte Chemie International Edition, 45(7), 1069-1074.
PF3D7_0211400	Beta-ketoacyl-ACP synthase III (KASIII)	2.3.1.180, 2.3.1.41	Prigge, S. T., He, X., Gerena, L., Waters, N. C., & Reynolds, K. A. (2003). The initiating steps of a type II fatty acid synthase in Plasmodium falciparum are catalyzed by pfACP, pfMCAT, and pfKASIII. *Biochemistry*, *42*(4), 1160-1169.

Gene ID	Reaction/Gene name	EC number	References
PF3D7_0312400	Glycogen synthase kinase 3	2.7.11.26	Syin, C., Parzy, D., Traincard, F., Boccaccio, I., Joshi, M. B., Lin, D. T., ... & Langsley, G. (2001). The H89 cAMP-dependent protein kinase inhibitor blocks Plasmodium falciparum development in infected erythrocytes. *European Journal of Biochemistry*, *268*(18), 4842-4849.; Xiao, Z., Waters, N. C., Woodard, C. L., Li, Z., & Li, P. K. (2001). Design and synthesis of Pfmrk inhibitors as potential antimalarial agents. Bioorganic & medicinal chemistry letters, 11(21), 2875-2878.
PF3D7_0320500	Nicotinamidase, putative	3.5.1.19	O'Hara, J. K., Kerwin, L. J., Cobbold, S. A., Tai, J., Bedell, T. A., Reider, P. J., & Llinás, M. (2014). Targeting NAD+ metabolism in the human malaria parasite Plasmodium falciparum. PloS one, 9(4), e94061.
PF3D7_0322000	Peptidyl-prolyl cis-trans isomerase	5.2.1.8	Monaghan, P., Fardis, M., Revill, W. P., & Bell, A. (2005). Antimalarial effects of macrolactones related to FK520 (ascomycin) are independent of the immunosuppressive properties of the compounds. *Journal of Infectious Diseases*, *191*(8), 1342-1349.
PF3D7_0406400	Cytosolic glyoxalase II (cGloII)	3.1.2.6	Urscher, M., Przyborski, J. M., Imoto, M., & Deponte, M. (2010). Distinct subcellular localization in the cytosol and apicoplast, unexpected dimerization and inhibition of Plasmodium falciparum glyoxalases. Molecular microbiology, 76(1), 92-103.
PF3D7_0415300	Cdc2-related protein kinase 3 (CRK3)	2.7.11.22	Rangarajan, R., Bei, A., Henry, N., Madamet, M., Parzy, D., Nivez, M. P., ... & Sultan, A. (2006). Pbcrk-1, the Plasmodium berghei orthologue of P. falciparum cdc-2 related kinase-1 (Pfcrk-1), is essential for completion of the intraerythrocytic asexual cycle. *Experimental parasitology*, *112*(3), 202-207.

Gene ID	Reaction/Gene name	EC number	References
PF3D7_0417200	Bifunctional dihydrofolate reductase-thymidylate synthase	1.5.1.3, 2.1.1.45	Le Bras, J., & Durand, R. (2003). The mechanisms of resistance to antimalarial drugs in Plasmodium falciparum. Fundamental & clinical pharmacology, 17(2), 147-153.; Fidock, D. A., Nomura, T., & Wellems, T. E. (1998). Cycloguanil and Its Parent Compound Proguanil Demonstrate Distinct Activities against Plasmodium falciparum Malaria Parasites Transformed with Human Dihydrofolate Reductase. Molecular pharmacology, 54(6), 1140-1147; Dar, O., Khan, M. S., & Adagu, I. (2008). The potential use of methotrexate in the treatment of falciparum malaria: in vitro assays against sensitive and multidrug-resistant falciparum strains. Japanese journal of infectious diseases, 61(3), 210-1.; Jiang, L., Lee, P. C., White, J., & Rathod, P. K. (2000). Potent and selective activity of a combination of thymidine and 1843U89, a folate-based thymidylate synthase inhibitor, against Plasmodium falciparum. Antimicrobial agents and chemotherapy, 44(4), 1047-1050.; Nduati, E., Hunt, S., Kamau, E. M., & Nzila, A. (2005). 2, 4-diaminopteridine-based compounds as precursors for de novo synthesis of antifolates: a novel class of antimalarials. Antimicrobial agents and chemotherapy, 49(9), 3652-3657.
PF3D7_0508200	Sphingosine-N-acyltransferase	2.3.1.24	Gerold, P., & Schwarz, R. T. (2001). Biosynthesis of glycosphingolipids de-novo by the human malaria parasite Plasmodium falciparum. Molecular and biochemical parasitology, 112(1), 29-37.
PF3D7_0509800	Phosphatidylinositol 4-kinase (PI4K)	2.7.1.67	McNamara, C. W., Lee, M. C., Lim, C. S., Lim, S. H., Roland, J., Nagle, A., ... & Manary, M. J. (2013). Targeting Plasmodium PI (4) K to eliminate malaria. Nature, 504(7479), 248-253.

Gene ID	Reaction/Gene name	EC number	References
PF3D7_0510500	Topoisomerase I	5.99.1.2	Bodley, A. L., Cumming, J. N., & Shapiro, T. A. (1998). Effects of camptothecin, a topoisomerase I inhibitor, on Plasmodium falciparum. Biochemical pharmacology, 55(5), 709-711.
PF3D7_0511200	Stearoyl-CoA desaturase (SCD)	1.14.19.1	Gratraud, P., Huws, E., Falkard, B., Adjalley, S., Fidock, D. A., Berry, L., ... & Kremer, L. (2009). Oleic acid biosynthesis in Plasmodium falciparum: characterization of the stearoyl-CoA desaturase and investigation as a potential therapeutic target. PloS one, 4(9), e6889.
PF3D7_0512700	Orotate phosphoribosyltransferase	2.4.2.10	Scott, H. V., Gero, A. M., & O'Sullivan, W. J. (1986). In vitro inhibition of Plasmodium falciparum by pyrazofurin, an inhibitor of pyrimidine biosynthesis de novo. *Molecular and biochemical parasitology*, *18*(1), 3-15.
PF3D7_0513300	Purine nucleoside phosphorylase (PNP)	2.4.2.1	Kicska, G. A., Tyler, P. C., Evans, G. B., Furneaux, R. H., Schramm, V. L., & Kim, K. (2002). Purine-less Death in Plasmodium falciparum Induced by Immucillin-H, a Transition State Analogue of Purine Nucleoside Phosphorylase. Journal of Biological Chemistry, 277(5), 3226-3231.

Gene ID	Reaction/Gene name	EC number	References
PF3D7_0520900	S-adenosyl-L-homocysteine hydrolase (SAHH)	3.3.1.1	Messika, E., Golenser, J., Abu-Elheiga, L., Robert-Gero, M., Lederer, E., & Bachrach, U. (1990). Effect of sinefungin on macromolecular biosynthesis and cell cycle of Plasmodium falciparum. Tropical medicine and parasitology: official organ of Deutsche Tropenmedizinische Gesellschaft and of Deutsche Gesellschaft fur Technische Zusammenarbeit (GTZ), 41(3), 273-278.; Kitade, Y., Kozaki, A., Gotoh, T., Miwa, T., Nakanishi, M., & Yatome, C. (1999, November). Synthesis of S-adenosyl-L-homocysteine hydrolase inhibitors and their biological activities. In Nucleic acids symposium series (Vol. 42, No. 1, pp. 25-26). Oxford University Press.; Shuto, S., Minakawa, N., Niizuma, S., Kim, H. S., Wataya, Y., & Matsuda, A. (2002). New Neplanocin Analogues. 12. Alternative Synthesis and Antimalarial Effect of (6 'R)-6 '-C-Methylneplanocin A, a Potent AdoHcy Hydrolase Inhibitor 1. Journal of medicinal chemistry, 45(3), 748-751.; Bujnicki, J. M., Prigge, S. T., Caridha, D., & Chiang, P. K. (2003). Structure, evolution, and inhibitor interaction of S-adenosyl-L-homocysteine hydrolase from Plasmodium falciparum. Proteins: Structure, Function, and Bioinformatics, 52(4), 624-632.
PF3D7_0527300	Methionine aminopeptidase 1a, putative (METAP1a)	3.4.11.18	Chen, X., Chong, C. R., Shi, L., Yoshimoto, T., Sullivan, D. J., & Liu, J. O. (2006). Inhibitors of Plasmodium falciparum methionine aminopeptidase 1b possess antimalarial activity. Proceedings of the National Academy of Sciences, 103(39), 14548-14553.

Gene ID	Reaction/Gene name	EC number	References
PF3D7_0603300	Dihydroorotate dehydrogenase, mitochondrial precursor (DHODH)	1.3.5.2	Krungkrai, J., Krungkrai, S. R., & Phakanont, K. (1992). Antimalarial activity of orotate analogs that inhibit dihydroorotase and dihydroorotate dehydrogenase. Biochemical pharmacology, 43(6), 1295-1301.; McRobert, L., & McConkey, G. A. (2002). RNA interference (RNAi) inhibits growth of Plasmodium falciparum. Molecular and biochemical parasitology, 119(2), 273-278.; Baldwin, J., Michnoff, C. H., Malmquist, N. A., White, J., Roth, M. G., Rathod, P. K., & Phillips, M. A. (2005). High-throughput screening for potent and selective inhibitors of Plasmodium falciparum dihydroorotate dehydrogenase. Journal of Biological Chemistry, 280(23), 21847-21853.; Boa, A. N., Canavan, S. P., Hirst, P. R., Ramsey, C., Stead, A. M., & McConkey, G. A. (2005). Synthesis of brequinar analogue inhibitors of malaria parasite dihydroorotate dehydrogenase. Bioorganic & medicinal chemistry, 13(6), 1945-1967.; Heikkilä, T., Thirumalairajan, S., Davies, M., Parsons, M. R., McConkey, A. G., Fishwick, C. W., & Johnson, A. P. (2006). The first de novo designed inhibitors of Plasmodium falciparum dihydroorotate dehydrogenase. Bioorganic & medicinal chemistry letters, 16(1), 88-92.
PF3D7_0604700	Lactoylglutathione lyase	4.4.1.5	Thornalley, P. J., Strath, M., & Wilson, R. J. M. (1994). Antimalarial activity in vitro of the glyoxalase I inhibitor diester, Sp-bromobenzylglutathione diethyl ester. Biochemical pharmacology, 47(2), 418-420.
PF3D7_0615100	Enoyl-acyl carrier reductase	1.3.1.9	Surolia, N., & Surolia, A. (2001). Triclosan offers protection against blood stages of malaria by inhibiting enoyl-ACP reductase of Plasmodium falciparum. Nature medicine, 7(2), 167-173.; Spalding, M. D., & Prigge, S. T. (2008). Malaria pulls a FASt one. Cell host & microbe, 4(6), 509-511.

Gene ID	Reaction/Gene name	EC number	References
PF3D7_0621200	Pyridoxine biosynthesis protein PDX1 (PDX1)	4.-.-.-	Reeksting, S. B., Müller, I. B., Burger, P. B., Burgos, E. S., Salmon, L., Louw, A. I., ... & Wrenger, C. (2013). Exploring inhibition of Pdx1, a component of the PLP synthase complex of the human malaria parasite Plasmodium falciparum. Biochemical Journal, 449(1), 175-187.
PF3D7_0623000	Chorismate synthase	4.2.3.5	McRobert, L., & McConkey, G. A. (2002). RNA interference (RNAi) inhibits growth of Plasmodium falciparum. Molecular and biochemical parasitology, 119(2), 273-278.
PF3D7_0624000	Hexokinase	2.7.1.1	Harris, M. T., Walker, D. M., Drew, M. E., Mitchell, W. G., Dao, K., Schroeder, C. E., ... & Morris, J. C. (2013). Interrogating a hexokinase-selected small-molecule library for inhibitors of Plasmodium falciparum hexokinase. *Antimicrobial agents and chemotherapy*, *57*(8), 3731-3737.
PF3D7_0624700	N-acetylglucosaminylphosphatidylinositol deacetylase, putative	3.5.1.89	Smith, T. K., Gerold, P., Crossman, A., Paterson, M. J., Borissow, C. N., Brimacombe, J. S., ... & Schwarz, R. T. (2002). Substrate specificity of the Plasmodium falciparum glycosylphosphatidylinositol biosynthetic pathway and inhibition by species-specific suicide substrates. Biochemistry, 41(41), 12395-12406.
PF3D7_0625000	Ceramide-cholinephosphotransferase; sphingomyelin synthase 1, putative	2.7.8.27	Gerold, P., & Schwarz, R. T. (2001). Biosynthesis of glycosphingolipids de-novo by the human malaria parasite Plasmodium falciparum. Molecular and biochemical parasitology, 112(1), 29-37.
PF3D7_0625100	Ceramide-cholinephosphotransferase; sphingomyelin synthase 2, putative	2.7.8.27	Gerold, P., & Schwarz, R. T. (2001). Biosynthesis of glycosphingolipids de-novo by the human malaria parasite Plasmodium falciparum. Molecular and biochemical parasitology, 112(1), 29-37.

Gene ID	Reaction/Gene name	EC number	References
PF3D7_0626300	3-oxoacyl-acyl-carrier protein synthase I/II	2.3.1.41, 2.3.1.85	Waller, R. F., Ralph, S. A., Reed, M. B., Su, V., Douglas, J. D., Minnikin, D. E., ... & McFadden, G. I. (2003). A type II pathway for fatty acid biosynthesis presents drug targets in Plasmodium falciparum. Antimicrobial agents and chemotherapy, 47(1), 297-301.; Prigge, S. T., He, X., Gerena, L., Waters, N. C., & Reynolds, K. A. (2003). The initiating steps of a type II fatty acid synthase in Plasmodium falciparum are catalyzed by pfACP, pfMCAT, and pfKASIII. Biochemistry, 42(4), 1160-1169.
PF3D7_0720400	Ferrodoxin reductase-like protein	1.7.1.4	Seeber, F., Aliverti, A., & Zanetti, G. (2005). The plant-type ferredoxin-NADP+ reductase/ferredoxin redox system as a possible drug target against apicomplexan human parasites. Current pharmaceutical design, 11(24), 3159-3172.
PF3D7_0724300	3-demethylubiquinone-9 3-methyltransferase, putative	2.1.1.64	Massimine, K. M., McIntosh, M. T., Doan, L. T., Atreya, C. E., Gromer, S., Sirawaraporn, W., ... & Anderson, K. S. (2006). Eosin B as a novel antimalarial agent for drug-resistant Plasmodium falciparum. Antimicrobial agents and chemotherapy, 50(9), 3132-3141.
PF3D7_0802000	Glutamate dehydrogenase, putative (GDH3)	1.4.1.2	Aparicio, I. M., Marín-Menéndez, A., Bell, A., & Engel, P. C. (2010). Susceptibility of Plasmodium falciparum to glutamate dehydrogenase inhibitors—A possible new antimalarial target. Molecular and biochemical parasitology, 172(2), 152-155.
PF3D7_0804400	Methionine aminopeptidase 1c, putative (METAP1c)	3.4.11.18	Chen, X., Chong, C. R., Shi, L., Yoshimoto, T., Sullivan, D. J., & Liu, J. O. (2006). Inhibitors of Plasmodium falciparum methionine aminopeptidase 1b possess antimalarial activity. Proceedings of the National Academy of Sciences, 103(39), 14548-14553.

Gene ID	Reaction/Gene name	EC number	References
PF3D7_0810800	Hydroxymethyldihydropterin pyrophosphokinase-dihydropteroate synthase (PPPK-DHPS)	2.5.1.15, 2.7.6.3	McCullough J. L., Maren T. H. (1974). Dihydropteroate synthetase from Plasmodium berghei: isolation, properties, and inhibition by dapsone and sulfadiazine. Molecular pharmacology, 10(1), 140-145.; Triglia, T., Menting, J. G., Wilson, C., & Cowman, A. F. (1997). Mutations in dihydropteroate synthase are responsible for sulfone and sulfonamide resistance in Plasmodium falciparum. Proceedings of the National Academy of Sciences, 94(25), 13944-13949.; Vinnicombe, H. G., & Derrick, J. P. (1999). Dihydropteroate synthase: an old drug target revisited. Biochem. Soc. Trans., 27, 53–58
PF3D7_0814900	Fe-superoxide dismutase	1.15.1.1	Soulere, L., Delplace, P., Davioud-Charvet, E., Py, S., Sergheraert, C., Perie, J., ... & Dive, D. (2003). Screening of Plasmodium falciparum iron superoxide dismutase inhibitors and accuracy of the SOD-assays. Bioorganic & medicinal chemistry, 11(23), 4941-4944.
PF3D7_0907900	Peptide deformylase (PDF)	3.5.1.88	Serero, A., Giglione, C., & Meinnel, T. (2001). Seeking new targets for antiparasitic agents: Response from A. Serero et al. *TRENDS in Parasitology*, *17*(1), 7-8.
PF3D7_0915000	Type II NADH:ubiquinone oxidoreductase (NDH2)	1.6.5.3, 1.6.5.9	Biagini, G. A., Viriyavejakul, P., O'neill, P. M., Bray, P. G., & Ward, S. A. (2006). Functional characterization and target validation of alternative complex I of Plasmodium falciparum mitochondria. Antimicrobial agents and chemotherapy, 50(5), 1841-1851.; Krungkrai, J., Kanchanarithisak, R., Krungkrai, S. R., & Rochanakij, S. (2002). Mitochondrial NADH dehydrogenase from Plasmodium falciparum and Plasmodium berghei. Experimental parasitology, 100(1), 54-61
PF3D7_0916600	Methyltransferase, putative	2.1.1.64	Massimine, K. M., McIntosh, M. T., Doan, L. T., Atreya, C. E., Gromer, S., Sirawaraporn, W., ... & Anderson, K. S. (2006). Eosin B as a novel antimalarial agent for drug-resistant Plasmodium falciparum. Antimicrobial agents and chemotherapy, 50(9), 3132-3141.

Gene ID	Reaction/Gene name	EC number	References
PF3D7_0918900	Gamma-glutamylcysteine synthetase	6.3.2.2	Luersen, K., Walter, R. D., Muller, S. (2000). Plasmodium falciparum-infected red blood cells depend on a functional glutathione de novo synthesis attributable to an enhanced loss of glutathione. Biochemical Journal, 346(2), 545-552.; Platel, D. F. N., Mangou, F., & Tribouley-Duret, J. (1999). Role of glutathione in the detoxification of ferriprotoporphyrin IX in chloroquine resistant Plasmodium berghei. Molecular and biochemical parasitology, 98(2), 215-223.; Meierjohann, S., Walter, R. D., & Müller, S. (2002). Regulation of intracellular glutathione levels in erythrocytes infected with chloroquine-sensitive and chloroquine-resistant Plasmodium falciparum. Biochemical journal, 368(3), 761-768.
PF3D7_0920800	Inosine-5'-monophosphate dehydrogenase	1.1.1.205	Webster, H. K., & Whaun, J. M. (1982). Antimalarial properties of bredinin. Prediction based on identification of differences in human host-parasite purine metabolism. Journal of Clinical Investigation, 70(2), 461.
PF3D7_0922400	Amino-deoxychorismate synthase	2.6.1.85	Camara, D., Bisanz, C., Barette, C., Van Daele, J., Human, E., Barnard, B., … & Maréchal, E. (2012). Inhibition of p-aminobenzoate and folate syntheses in plants and apicomplexan parasites by natural product rubreserine. *Journal of Biological Chemistry*, *287*(26), 22367-22376.
PF3D7_0922900	3-oxoacyl-(acyl-carrier protein) reductase	1.1.1.100, 2.3.1.85	Wickramasinghe, S. R., Inglis, K. A., Urch, J. E., Müller, S., Van Aalten, D. M., & Fairlamb, A. H. (2006). Kinetic, inhibition and structural studies on 3-oxoacyl-ACP reductase from Plasmodium falciparum, a key enzyme in fatty; acid biosynthesis. *Biochemical Journal*, *393*(2), 447-457.; Prigge, S. T., He, X., Gerena, L., Waters, N. C., & Reynolds, K. A. (2003). The initiating steps of a type II fatty acid synthase in Plasmodium falciparum are catalyzed by pfACP, pfMCAT, and pfKASIII. Biochemistry, 42(4), 1160-1169.

Gene ID	Reaction/Gene name	EC number	References
PF3D7_0923800	Thioredoxin reductase	1.8.1.9	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, 277(29), 25970-25975.; Luersen, K., Walter, R. D., Muller, S. (2000). Plasmodium falciparum-infected red blood cells depend on a functional glutathione de novo synthesis attributable to an enhanced loss of glutathione. Biochemical Journal, 346(2), 545-552.
PF3D7_0925700	Histone deacetylase	3.5.1.98	Singh, S. B., Zink, D. L., Liesch, J. M., Mosley, R. T., Dombrowski, A. W., Bills, G. F., ... & Goetz, M. A. (2002). Structure and chemistry of apicidins, a class of novel cyclic tetrapeptides without a terminal α-keto epoxide as inhibitors of histone deacetylase with potent antiprotozoal activities. *The Journal of organic chemistry*, *67*(3), 815-825.; Mai, A., Cerbara, I., Valente, S., Massa, S., Walker, L. A., & Tekwani, B. L. (2004). Antimalarial and antileishmanial activities of aroyl-pyrrolyl-hydroxyamides, a new class of histone deacetylase inhibitors. Antimicrobial agents and chemotherapy, 48(4), 1435-1436.
PF3D7_0926700	Glutamine-dependent NAD(+) synthetase, putative (NADSYN)	6.3.5.1	O'Hara, J. K., Kerwin, L. J., Cobbold, S. A., Tai, J., Bedell, T. A., Reider, P. J., & Llinás, M. (2014). Targeting NAD+ metabolism in the human malaria parasite Plasmodium falciparum. *PloS one*, *9*(4), e94061.
PF3D7_0934800	cAMP-dependent protein kinase catalytic subunit	2.7.11.11	Syin, C., Parzy, D., Traincard, F., Boccaccio, I., Joshi, M. B., Lin, D. T., ... & Langsley, G. (2001). The H89 cAMP-dependent protein kinase inhibitor blocks Plasmodium falciparum development in infected erythrocytes. *European Journal of Biochemistry*, *268*(18), 4842-4849.

Gene ID	Reaction/Gene name	EC number	References
PF3D7_1008900	Adenylate kinase	2.7.4.3	Ulschmid, J. K., Rahlfs, S., Schirmer, R. H., & Becker, K. (2004). Adenylate kinase and GTP: AMP phosphotransferase of the malarial parasite Plasmodium falciparum: central players in cellular energy metabolism. *Molecular and biochemical parasitology*, *136*(2), 211-220.; Kanaani, J., & Ginsburg, H. (1989). Metabolic interconnection between the human malarial parasite Plasmodium falciparum and its host erythrocyte. Regulation of ATP levels by means of an adenylate translocator and adenylate kinase. Journal of Biological Chemistry, 264(6), 3194-3199.
PF3D7_1012400	Hypoxanthine-guanine phosphoribosyltransferase (HGPRT)	2.4.2.8	Dawson, P. A., Cochran, D. A., Emmerson, B. T., & Gordon, R. B. (1993). Inhibition of Plasmodium falciparum hypoxanthine-guanine phosphoribosyltransferase mRNA by antisense oligodeoxynucleotide sequence. Molecular and biochemical parasitology, 60(1), 153-156.; Li, C. M., Tyler, P. C., Furneaux, R. H., Kicska, G., Xu, Y., Grubmeyer, C., ... & Schramm, V. L. (1999). Transition-state analogs as inhibitors of human and malarial hypoxanthine-guanine phosphoribosyltransferases. Nature Structural & Molecular Biology, 6(6), 582-587.
PF3D7_1012600	GMP synthetase	6.3.4.1, 6.3.5.2	McConkey, G. A. (2000). Plasmodium falciparum: isolation and characterisation of a gene encoding protozoan GMP synthase. Experimental parasitology, 94(1), 23-32.
PF3D7_1015300	Methionine aminopeptidase 1b, putative (METAP1b)	3.4.11.18	Chen, X., Chong, C. R., Shi, L., Yoshimoto, T., Sullivan, D. J., & Liu, J. O. (2006). Inhibitors of Plasmodium falciparum methionine aminopeptidase 1b possess antimalarial activity. Proceedings of the National Academy of Sciences, 103(39), 14548-14553.

Gene ID	Reaction/Gene name	EC number	References
PF3D7_1023200	Orotidine monophosphate decarboxylase	4.1.1.23	Scott, H. V., Gero, A. M., & O'Sullivan, W. J. (1986). In vitro inhibition of Plasmodium falciparum by pyrazofurin, an inhibitor of pyrimidine biosynthesis de novo. Molecular and biochemical parasitology, 18(1), 3-15.; Seymour, K. K., Lyons, S. D., Phillips, L., Rieckmann, K. H., & Christopherson, R. I. (1994). Cytotoxic effects of inhibitors of de novo pyrimidine biosynthesis upon Plasmodium falciparum. Biochemistry, 33(17), 5268-5274.; Krungkrai, S. R., DelFraino, B. J., Smiley, J. A., Prapunwattana, P., Mitamura, T., Horii, T., & Krungkrai, J. (2005). A Novel Enzyme Complex of Orotate Phosphoribosyltransferase and Orotidine 5 '-Monophosphate Decarboxylase in Human Malaria Parasite Plasmodium falciparum: Physical Association, Kinetics, and Inhibition Characterization†. Biochemistry, 44(5), 1643-1652.; Bello, A. M., Poduch, E., Liu, Y., Wei, L., Crandall, I., Wang, X., ... & Kotra, L. P. (2008). Structure–activity relationships of C6-uridine derivatives targeting plasmodia orotidine monophosphate decarboxylase. Journal of medicinal chemistry, 51(3), 439-448.; Kotra, L. P., Meza-Avina, M. E., Wei, L., Buhendwa, M. G., Poduch, E., Bello, A. M., & Pai, E. F. (2008). Inhibition of orotidine 5'-monophosphate decarboxylase and its therapeutic potential. Mini reviews in medicinal chemistry, 8(3), 239-247.
PF3D7_1026900	Acetyl-CoA carboxylase	6.4.1.2	Waller, R. F., Ralph, S. A., Reed, M. B., Su, V., Douglas, J. D., Minnikin, D. E., ... & McFadden, G. I. (2003). A type II pathway for fatty acid biosynthesis presents drug targets in Plasmodium falciparum. Antimicrobial agents and chemotherapy, 47(1), 297-301.

Gene ID	Reaction/Gene name	EC number	References
PF3D7_1029600	Adenosine deaminase (ADA)	3.5.4.4	Gero, A. M., Dunn, C. G., Brown, D. M., Pulenthiran, K., Gorovits, E. L., Bakos, T., & Weis, A. L. (2003). New malaria chemotherapy developed by utilization of a unique parasite transport system. Current pharmaceutical design, 9(11), 867-877.; Tyler, P. C., Taylor, E. A., Fröhlich, R. F., & Schramm, V. L. (2007). Synthesis of 5 '-Methylthio Coformycins: Specific Inhibitors for Malarial Adenosine Deaminase. Journal of the American Chemical Society, 129(21), 6872-6879.; Ting, L. M., Shi, W., Lewandowicz, A., Singh, V., Mwakingwe, A., Birck, M. R., ... & Evans, G. B. (2005). Targeting a novel Plasmodium falciparum purine recycling pathway with specific immucillins. Journal of Biological Chemistry, 280(10), 9547-9554.
PF3D7_1033100	S-adenosylmethionine decarboxylase/ornithine decarboxylase (AdoMetDC/ODC)	4.1.1.17, 4.1.1.50	Assaraf, Y. G., Golenser, J., Spira, D. T., & Bachrach, U. (1984). Polyamine levels and the activity of their biosynthetic enzymes in human erythrocytes infected with the malarial parasite, Plasmodium falciparum. Biochemical Journal, 222(3), 815-819.; Bitonti, A. J., Dumont, J. A., Bush, T. L., Edwards, M. L., Stemerick, D. M., McCann, P. P., & Sjoerdsma, A. (1989). Bis (benzyl) polyamine analogs inhibit the growth of chloroquine-resistant human malaria parasites (Plasmodium falciparum) in vitro and in combination with alpha-difluoromethylornithine cure murine malaria. Proceedings of the National Academy of Sciences, 86(2), 651-655.; Berger, B. J. (2000). Antimalarial activities of aminooxy compounds. Antimicrobial agents and chemotherapy, 44(9), 2540-2542.; Wright, P. S., Byers, T. L., Cross-Doersen, D. E., McCann, P. P., & Bitonti, A. J. (1991). Irreversible inhibition of S-adenosylmethionine decarboxylase in Plasmodium falciparum-infected erythrocytes: growth inhibition in vitro. Biochemical pharmacology, 41(11), 1713-1718.

Gene ID	Reaction/Gene name	EC number	References
PF3D7_1034400	Flavoprotein subunit of succinate dehydrogenase	1.3.5.1, 1.3.99.1, 6.2.1.4	Suraveratum, N., Krungkrai, S. R., Leangaramgul, P., Prapunwattana, P., & Krungkrai, J. (2000). Purification and characterization of Plasmodium falciparum succinate dehydrogenase. *Molecular and biochemical parasitology*, *105*(2), 215-222.
PF3D7_1113700	Lactoylglutathione lyase	4.4.1.5	Thornalley, P. J., Strath, M., & Wilson, R. J. M. (1994). Antimalarial activity in vitro of the glyoxalase I inhibitor diester, Sp-bromobenzylglutathione diethyl ester. Biochemical pharmacology, 47(2), 418-420.
PF3D7_1114800	Glycerol-3-phosphate dehydrogenase, putative (G3PDH)	1.1.1.8	Lindner, S. E., Sartain, M. J., Hayes, K., Harupa, A., Moritz, R. L., Kappe, S. H., & Vaughan, A. M. (2014). Enzymes involved in plastid-targeted phosphatidic acid synthesis are essential for Plasmodium yoelii liver-stage development. Molecular microbiology, 91(4), 679-693.
PF3D7_1115700	Falcipain-2A	3.4.22.-	Biot, C., Pradines, B., Sergeant, M. H., Gut, J., Rosenthal, P. J., & Chibale, K. (2007). Design, synthesis, and antimalarial activity of structural chimeras of thiosemicarbazone and ferroquine analogues. *Bioorganic & medicinal chemistry letters*, *17*(23), 6434-6438.
PF3D7_1116700	Dipeptidyl aminopeptidase 1	3.4.14.1	Klemba, M., Gluzman, I., & Goldberg, D. E. (2004). A Plasmodium falciparum dipeptidyl aminopeptidase I participates in vacuolar hemoglobin degradation. Journal of Biological Chemistry, 279(41), 43000-43007.
PF3D7_1118200	Heat shock protein 90, putative		Banumathy, G., Singh, V., Pavithra, S. R., & Tatu, U. (2003). Heat shock protein 90 function is essential for Plasmodium falciparum growth in human erythrocytes. *Journal of Biological Chemistry*, *278*(20), 18336-18345.

Gene ID	Reaction/Gene name	EC number	References
PF3D7_1126000	Threonine--tRNA ligase (ThrRS)	6.1.1.3	Ruan, B., Bovee, M. L., Sacher, M., Stathopoulos, C., Poralla, K., Francklyn, C. S., & Söll, D. (2005). A unique hydrophobic cluster near the active site contributes to differences in borrelidin inhibition among threonyl-tRNA synthetases. Journal of Biological Chemistry, 280(1), 571-577.
PF3D7_1127100	Deoxyuridine 5'-triphosphate nucleotidohydrolase, putative	3.6.1.23	Whittingham, J. L., Leal, I., Nguyen, C., Kasinathan, G., Bell, E., Jones, A. F., ... & Perez, L. M. R. (2005). dUTPase as a platform for antimalarial drug design: structural basis for the selectivity of a class of nucleoside inhibitors. Structure, 13(2), 329-338.
PF3D7_1129000	Spermidine synthase	2.5.1.16	Haider, N., Eschbach, M. L., de Souza Dias, S., Gilberger, T. W., Walter, R. D., & Lüersen, K. (2005). The spermidine synthase of the malaria parasite Plasmodium falciparum: molecular and biochemical characterisation of the polyamine synthesis enzyme. Molecular and biochemical parasitology, 142(2), 224-236.
PF3D7_1136500	Casein kinase 1, PfCK1	2.7.11.1	Waters, N. C., Woodard, C. L., & Prigge, S. T. (2000). Cyclin H activation and drug susceptibility of the Pfmrk cyclin dependent protein kinase from Plasmodium falciparum. *Molecular and biochemical parasitology*, *107*(1), 45-55.
PF3D7_1140000	Carbonic anhydrase, putative	4.2.1.1	Reungprapavut, S., Krungkrai, S. R., & Krungkrai, J. (2004). Plasmodium falciparum carbonic anhydrase is a possible target for malaria chemotherapy. Journal of enzyme inhibition and medicinal chemistry, 19(3), 249-256.
PF3D7_1147500	Farnesyltransferase beta subunit, putative	2.5.1.29, 2.5.1.58	Ohkanda, J., Lockman, J. W., Yokoyama, K., Gelb, M. H., Croft, S. L., Kendrick, H., ... & Hamilton, A. D. (2001). Peptidomimetic inhibitors of protein farnesyltransferase show potent antimalarial activity. *Bioorganic & medicinal chemistry letters*, *11*(6), 761-764.

Gene ID	Reaction/Gene name	EC number	References
PF3D7_1205700	Targeted glyoxalase II (tGloII)	3.1.2.6	Urscher, M., Przyborski, J. M., Imoto, M., & Deponte, M. (2010). Distinct subcellular localization in the cytosol and apicoplast, unexpected dimerization and inhibition of Plasmodium falciparum glyoxalases. Molecular microbiology, 76(1), 92-103.
PF3D7_1211900	Non-SERCA-type Ca2+ -transporting P-ATPase	3.6.3.2, 3.6.3.8	Eckstein-Ludwig, U., Webb, R. J., Van Goethem, I. D. A., East, J. M., Lee, A. G., Kimura, M., ... & Krishna, S. (2003). Artemisinins target the SERCA of Plasmodium falciparum. *Nature*, *424*(6951), 957-961.
PF3D7_1212500	Glycerol-3-phosphate acyltransferase	2.3.1.15	Nicolas, O., Margout, D., Taudon, N., Wein, S., Calas, M., Vial, H. J., & Bressolle, F. M. (2005). Pharmacological properties of a new antimalarial bisthiazolium salt, T3, and a corresponding prodrug, TE3. *Antimicrobial agents and chemotherapy*, *49*(9), 3631-3639.
PF3D7_1212800	Iron-sulfur subunit of succinate dehydrogenase	1.3.5.1, 1.3.99.1	Suraveratum, N., Krungkrai, S. R., Leangaramgul, P., Prapunwattana, P., & Krungkrai, J. (2000). Purification and characterization of Plasmodium falciparum succinate dehydrogenase. *Molecular and biochemical parasitology*, *105*(2), 215-222.
PF3D7_1235600	Serine hydroxymethyltransferase (SHMT)	2.1.2.1	Witschel, M. C., Rottmann, M., Schwab, A., Leartsakulpanich, U., Chitnumsub, P., Seet, M., ... & McNamara, C. (2015). Inhibitors of plasmodial serine hydroxymethyltransferase (SHMT): cocrystal structures of pyrazolopyrans with potent blood-and liver-stage activities. Journal of medicinal chemistry, 58(7), 3117-3130.
PF3D7_1236800	Protein-S-isoprenylcysteine O-methyltransferase, putative	2.1.1.100	Baron, R. A., Peterson, Y. K., Otto, J. C., Rudolph, J., & Casey, P. J. (2007). Time-dependent inhibition of isoprenylcysteine carboxyl methyltransferase by indole-based small molecules. Biochemistry, 46(2), 554-560.; Winter-Vann, A. M., Baron, R. A., Wong, W., dela Cruz, J., York, J. D., Gooden, D. M., ... & Casey, P. J. (2005). A small-molecule inhibitor of isoprenylcysteine carboxyl methyltransferase with antitumor activity in cancer cells. Proceedings of the National Academy of Sciences of the United States of America, 102(12), 4336-4341.

Gene ID	Reaction/Gene name	EC number	References
PF3D7_1238600	Sphingomyelin phosphodiesterase, putative	3.1.4.12	Hanada, K., Palacpac, N. M. Q., Magistrado, P. A., Kurokawa, K., Rai, G., Sakata, D., ... & Mitamura, T. (2002). Plasmodium falciparum phospholipase C hydrolyzing sphingomyelin and lysocholinephospholipids is a possible target for malaria chemotherapy. Journal of Experimental Medicine, 195(1), 23-34.
PF3D7_1239500	DNA gyrase subunit B	5.99.1.3	Gamage, S. A., Tepsiri, N., Wilairat, P., Wojcik, S. J., Figgitt, D. P., Ralph, R. K., & Denny, W. A. (1994). Synthesis and in vitro evaluation of 9-anilino-3, 6-diaminoacridines active against a multidrug-resistant strain of the malaria parasite Plasmodium falciparum. Journal of medicinal chemistry, 37(10), 1486-1494.; Chavalitshewinkoon-Petmitr, P., Pongvilairat, G., Auparakkitanon, S., & Wilairat, P. (2000). Gametocytocidal activity of pyronaridine and DNA topoisomerase II inhibitors against multidrug-resistant Plasmodium falciparum in vitro. Parasitology international, 48(4), 275-280.; Noonpakdee, W., Pothikasikorn, J., Nimitsantiwong, W., & Wilairat, P. (2003). Inhibition of Plasmodium falciparum proliferation in vitro by antisense oligodeoxynucleotides against malarial topoisomerase II. Biochemical and biophysical research communications, 302(4), 659-664.
PF3D7_1246100	Delta-aminolevulinic acid synthetase	2.3.1.37	Surolia, N., & Padmanaban, G. (1992). De novo biosynthesis of heme offers a new chemotherapeutic target in the human malarial parasite. Biochemical and Biophysical Research Communications, 187(2), 744-750.; Varadharajan, S., Dhanasekaran, S., Bonday, Z. Q., Rangarajan, P. N., Padmanaban, G. (2002). Involvement of deltaaminolaevulinate synthase encoded by the parasite gene in de novo haem synthesis by Plasmodium falciparum. Biochemical Journal. 367(Pt 2):321–327.

182

Gene ID	Reaction/Gene name	EC number	References
PF3D7_1246900	Rac-beta serine/threonine protein kinase, PfPKB	2.7.11.1	Waters, N. C., Woodard, C. L., & Prigge, S. T. (2000). Cyclin H activation and drug susceptibility of the Pfmrk cyclin dependent protein kinase from Plasmodium falciparum. *Molecular and biochemical parasitology*, *107*(1), 45-55.
PF3D7_1247800	Dipeptidyl aminopeptidase 2 (DPAP2)	3.4.14.1	Tanaka, T. Q., Deu, E., Molina-Cruz, A., Ashburne, M. J., Ali, O., Suri, A., ... & Williamson, K. C. (2013). Plasmodium dipeptidyl aminopeptidases as malaria transmission-blocking drug targets. Antimicrobial agents and chemotherapy, 57(10), 4645-4652.
PF3D7_1302600	Deoxyhypusine hydroxylase (DOHH)	1.14.99.29	Saeftel, M., Sarite, R. S., Njuguna, T., Holzgrabe, U., Ulmer, D., Hoerauf, A., & Kaiser, A. (2006). Piperidones with activity against Plasmodium falciparum. *Parasitology research*, *99*(3), 281-286.
PF3D7_1308200	Carbamoyl phosphate synthetase	6.3.4.16, 6.3.5.5	Flores, M. V., Atkins, D., Wade, D., O'Sullivan, W. J., Stewart, T. S. (1997). Inhibition of Plasmodium falciparum proliferation in vitro by ribozymes. Journal of Biological Chemistry, 272(27):16940–16945.
PF3D7_1311800	M1-family alanyl aminopeptidase (M1AAP)	3.4.11.2	Skinner-Adams, T. S., Stack, C. M., Trenholme, K. R., Brown, C. L., Grembecka, J., Lowther, J., ... & Whisstock, J. C. (2010). Plasmodium falciparum neutral aminopeptidases: new targets for anti-malarials. Trends in biochemical sciences, 35(1), 53-61.
PF3D7_1316600	Choline-phosphate cytidylyltransferase (CCT)	2.7.7.15	González-Bulnes, P., Bobenchik, A. M., Augagneur, Y., Cerdan, R., Vial, H. J., Llebaria, A., & Mamoun, C. B. (2011). PG12, a phospholipid analog with potent antimalarial activity, inhibits Plasmodium falciparum CTP: phosphocholine cytidylyltransferase activity. Journal of Biological Chemistry, 286(33), 28940-28947.

Gene ID	Reaction/Gene name	EC number	References
PF3D7_1318200	Glycerol-3-phosphate acyltransferase	2.3.1.15	Lindner, S. E., Sartain, M. J., Hayes, K., Harupa, A., Moritz, R. L., Kappe, S. H., & Vaughan, A. M. (2014). Enzymes involved in plastid-targeted phosphatidic acid synthesis are essential for Plasmodium yoelii liver-stage development. Molecular microbiology, 91(4), 679-693.; Nicolas, O., Margout, D., Taudon, N., Wein, S., Calas, M., Vial, H. J., & Bressolle, F. M. (2005). Pharmacological properties of a new antimalarial bisthiazolium salt, T3, and a corresponding prodrug, TE3. Antimicrobial agents and chemotherapy, 49(9), 3631-3639.
PF3D7_1323000	3-hydroxyacyl-ACP dehydratase	4.2.1.58, 4.2.1.59, 4.2.1.60, 4.2.1.61	Sharma, S. K., Kapoor, M., Ramya, T. N. C., Kumar, S., Kumar, G., Modak, R., ... & Surolia, A. (2003). Identification, characterization, and inhibition of Plasmodium falciparum β-hydroxyacyl-acyl carrier protein dehydratase (FabZ). Journal of Biological Chemistry, 278(46), 45661-45671.
PF3D7_1324900	Lactate dehydrogenase	1.1.1.27	Granchi, C., Bertini, S., Macchia, M., & Minutolo, F. (2010). Inhibitors of lactate dehydrogenase isoforms and their therapeutic potentials. Current medicinal chemistry, 17(7), 672-697.
PF3D7_1327600	Nicotinate-nucleotide adenylyltransferase (NMNAT)	2.7.7.18	O'Hara, J. K., Kerwin, L. J., Cobbold, S. A., Tai, J., Bedell, T. A., Reider, P. J., & Llinás, M. (2014). Targeting NAD+ metabolism in the human malaria parasite Plasmodium falciparum. PloS one, 9(4), e94061.
PF3D7_1342100	Aconitate hydratase	4.2.1.3	Hodges, M., Yikilmaz, E., Patterson, G., Kasvosve, I., Rouault, T. A., Gordeuk, V. R., & Loyevsky, M. (2005). An iron regulatory-like protein expressed in Plasmodium falciparum displays aconitase activity. Molecular and biochemical parasitology, 143(1), 29-38.

Gene ID	Reaction/Gene name	EC number	References
PF3D7_1343000	Phosphoethanolamine N-methyltransferase	2.1.1.103	Witola, W. H., El Bissati, K., Pessi, G., Xie, C., Roepe, P. D., & Mamoun, C. B. (2008). Disruption of the Plasmodium falciparum PfPMT gene results in a complete loss of phosphatidylcholine biosynthesis via the serine-decarboxylase-phosphoethanolamine-methyltransferase pathway and severe growth and survival defects. Journal of Biological Chemistry, 283(41), 27636-27643.
PF3D7_1345700	Isocitrate dehydrogenase (NADP), mitochondrial precursor (IDH)	1.1.1.42	Wrenger, C., & Müller, S. (2003). Isocitrate dehydrogenase of Plasmodium falciparum. European Journal of Biochemistry, 270(8), 1775-1783.
PF3D7_1347200	Nucleoside transporter 1 (NT1)		Frame, I. J., Deniskin, R., Rinderspacher, A., Katz, F., Deng, S. X., Moir, R. D., ... & Landry, D. W. (2015). Yeast-based high-throughput screen identifies Plasmodium falciparum equilibrative nucleoside transporter 1 inhibitors that kill malaria parasites. ACS chemical biology, 10(3), 775-783.
PF3D7_1351600	Glycerol kinase (GK)	2.7.1.30	Naidoo, K., & Coetzer, T. L. (2013). Reduced glycerol incorporation into phospholipids contributes to impaired intra-erythrocytic growth of glycerol kinase knockout Plasmodium falciparum parasites. *Biochimica et Biophysica Acta (BBA)-General Subjects*, *1830*(11), 5326-5334.
PF3D7_1354500	Adenylosuccinate synthetase (adsS)	6.3.4.4	Eaazhisai, K., Jayalakshmi, R., Gayathri, P., Anand, R. P., Sumathy, K., Balaram, H., & Murthy, M. R. N. (2004). Crystal structure of fully ligated adenylosuccinate synthetase from Plasmodium falciparum. Journal of molecular biology, 335(5), 1251-1264.
PF3D7_1356900	Protein kinase 5	2.7.11.22	Harmse, L., van Zyl, R., Gray, N., Schultz, P., Leclerc, S., Meijer, L., ... & Havlik, I. (2001). Structure-activity relationships and inhibitory effects of various purine derivatives on the in vitro growth of Plasmodium falciparum. *Biochemical pharmacology*, *62*(3), 341-348.

Gene ID	Reaction/Gene name	EC number	References
PF3D7_1360800	Falcilysin	3.4.23.1, 3.4.24.-, 4.4.1.21	Zhang, P., Nicholson, D. E., Bujnicki, J. M., Su, X., Brendle, J. J., Ferdig, M., ... & Chiang, P. K. (2002). Angiogenesis inhibitors specific for methionine aminopeptidase 2 as drugs for malaria and leishmaniasis. *Journal of biomedical science*, *9*(1), 34-40.
PF3D7_1364900	Ferrochelatase (FC)	4.99.1.1	Nagaraj, V. A., Sundaram, B., Varadarajan, N. M., Subramani, P. A., Kalappa, D. M., Ghosh, S. K., & Padmanaban, G. (2013). Malaria parasite-synthesized heme is essential in the mosquito and liver stages and complements host heme in the blood stages of infection. PLoS Pathog, 9(8), e1003522.
PF3D7_1367500	NADH-cytochrome b5 reductase, putative	1.7.1.3	Saxena, N., Saxena, A., Dutta, G. P., Ghatak, S., & Pandey, V. C. (1987). Effect of Plasmodium yoelli nigeriensis infection and chloroquine on the hepatic mixed function oxidase system of mice. Molecular and biochemical parasitology, 24(3), 283-287.
PF3D7_1367700	Alanine--tRNA ligase (AlaRS)	6.1.1.7	Corvaisier, S., Bordeau, V., & Felden, B. (2003). Inhibition of transfer messenger RNA aminoacylation and trans-translation by aminoglycoside antibiotics. Journal of Biological Chemistry, 278(17), 14788-14797.
PF3D7_1401800	Choline kinase (CK)	2.7.1.32	Choubey, V., Maity, P., Guha, M., Kumar, S., Srivastava, K., Puri, S. K., & Bandyopadhyay, U. (2007). Inhibition of Plasmodium falciparum choline kinase by hexadecyltrimethylammonium bromide: a possible antimalarial mechanism. Antimicrobial agents and chemotherapy, 51(2), 696-706.
PF3D7_1403700	Sphingosine-N-acyltransferase	2.3.1.24	Gerold, P., & Schwarz, R. T. (2001). Biosynthesis of glycosphingolipids de-novo by the human malaria parasite Plasmodium falciparum. Molecular and biochemical parasitology, 112(1), 29-37.

Gene ID	Reaction/Gene name	EC number	References
PF3D7_1405600	Ribonucleotide reductase small subunit	1.17.4.1	Chakrabarti, D., Schuster, S. M., & Chakrabarti, R. (1993). Cloning and characterization of subunit genes of ribonucleotide reductase, a cell-cycle-regulated enzyme, from Plasmodium falciparum. Proceedings of the National Academy of Sciences, 90(24), 12020-12024.; Barker, R. H., Metelev, V., Rapaport, E., & Zamecnik, P. (1996). Inhibition of Plasmodium falciparum malaria using antisense oligodeoxynucleotides. Proceedings of the National Academy of Sciences, 93(1), 514-518.; Lytton, S. D., Mester, B., Libman, J., Shanzer, A., & Cabantchik, Z. I. (1994). Mode of action of iron (III) chelators as antimalarials: II. Evidence for differential effects on parasite iron-dependent nucleic acid synthesis. Blood, 84(3), 910-915.
PF3D7_1407800	Plasmepsin IV (PM4)	3.4.23.-	Coombs, G. H., Goldberg, D. E., Klemba, M., Berry, C., Kay, J., & Mottram, J. C. (2001). Aspartic proteases of Plasmodium falciparum and other parasitic protozoa as drug targets. Trends in parasitology, 17(11), 532-537.
PF3D7_1407900	Plasmepsin I (PMI)	3.4.23.38	Nöteberg, D., Hamelink, E., Hultén, J., Wahlgren, M., Vrang, L., Samuelsson, B., & Hallberg, A. (2003). Design and Synthesis of Plasmepsin I and Plasmepsin II Inhibitors with Activity in Plasmodium f alciparum-Infected Cultured Human Erythrocytes. Journal of medicinal chemistry, 46(5), 734-746.; Romeo, S., Dell'Agli, M., Parapini, S., Rizzi, L., Galli, G., Mondani, M., ... & Bosisio, E. (2004). Plasmepsin II inhibition and antiplasmodial activity of Primaquine–Statinedouble-drugs'. Bioorganic & medicinal chemistry letters, 14(11), 2931-2934.

Gene ID	Reaction/Gene name	EC number	References
PF3D7_1408000	Plasmepsin II	3.4.23.39	Nöteberg, D., Hamelink, E., Hultén, J., Wahlgren, M., Vrang, L., Samuelsson, B., & Hallberg, A. (2003). Design and Synthesis of Plasmepsin I and Plasmepsin II Inhibitors with Activity in Plasmodium f alciparum-Infected Cultured Human Erythrocytes. Journal of medicinal chemistry, 46(5), 734-746.; Romeo, S., Dell'Agli, M., Parapini, S., Rizzi, L., Galli, G., Mondani, M., ... & Bosisio, E. (2004). Plasmepsin II inhibition and antiplasmodial activity of Primaquine–Statinedouble-drugs'. Bioorganic & medicinal chemistry letters, 14(11), 2931-2934.
PF3D7_1414400	Serine/threonine protein phosphatase	3.1.3.16	Blisnick, T., Vincensini, L., Fall, G., & Braun-Breton, C. (2006). Protein phosphatase 1, a Plasmodium falciparum essential enzyme, is exported to the host cell and implicated in the release of infectious merozoites. *Cellular microbiology*, *8*(4), 591-601.
PF3D7_1415700	Serine C-palmitoyltransferase, putative	2.3.1.50	Gerold, P., & Schwarz, R. T. (2001). Biosynthesis of glycosphingolipids de-novo by the human malaria parasite Plasmodium falciparum. Molecular and biochemical parasitology, 112(1), 29-37.

Gene ID	Reaction/Gene name	EC number	References
PF3D7_1419300	Glutathione S-transferase (GST)	2.5.1.18	Fritz-Wolf, K., Becker, A., Rahlfs, S., Harwaldt, P., Schirmer, R. H., Kabsch, W., & Becker, K. (2003). X-ray structure of glutathione S-transferase from the malarial parasite Plasmodium falciparum. Proceedings of the National Academy of Sciences, 100(24), 13821-13826.; Harwaldt, P., Rahlfs, S., & Becker, K. (2002). Glutathione S-transferase of the malarial parasite Plasmodium falciparum: characterization of a potential drug target. Biological chemistry, 383(5), 821-830.; Liebau, E., Bergmann, B., Campbell, A. M., Teesdale-Spittle, P., Brophy, P. M., Luersen, K., Walter, R. D. (2002) The glutathione S-transferase from Plasmodium falciparum. Molecular and Biochemical Parasitology, 124(1-2):85–90.; Perbandt, M., Burmeister, C., Walter, R. D., Betzel, C., & Liebau, E. (2004). Native and inhibited structure of a Mu class-related glutathione S-transferase from Plasmodium falciparum. Journal of Biological Chemistry, 279(2), 1336-1342.
PF3D7_1419800	Glutathione reductase	1.8.1.7	Biot, C., Bauer, H., Schirmer, R. H., & Davioud-Charvet, E. (2004). 5-Substituted tetrazoles as bioisosteres of carboxylic acids. Bioisosterism and mechanistic studies on glutathione reductase inhibitors as antimalarials. Journal of medicinal chemistry, 47(24), 5972-5983.; Ya, Z., Hempelmann, E., & Schirmer, R. H. (1988). Glutathione reductase inhibitors as potential antimalarial drugs: Effects of nitrosoureas on Plasmodium falciparum in vitro. Biochemical pharmacology, 37(5), 855-860.
PF3D7_1420600	Pantothenate kinase, putative	2.7.1.33	Saliba, K. J., Ferru, I., & Kirk, K. (2005). Provitamin B5 (pantothenol) inhibits growth of the intraerythrocytic malaria parasite. *Antimicrobial agents and chemotherapy*, *49*(2), 632-637.

Gene ID	Reaction/Gene name	EC number	References
PF3D7_1425400	DNA-directed DNA polymerase, putative	2.7.7.7	Barker, R. H., Metelev, V., Rapaport, E., & Zamecnik, P. (1996). Inhibition of Plasmodium falciparum malaria using antisense oligodeoxynucleotides. Proceedings of the National Academy of Sciences, 93(1), 514-518.
PF3D7_1426900	Ubiquinol-cytochrome c reductase hinge protein, putative	1.10.2.2	Fieser, L. F., Berliner, E., Bondhus, F. J., Chang, F. C., Dauben, W. G., Ettlinger, M. G., ... & Heymann, H. (1948). Naphthoquinone Antimalarials. I. General Survey1, 2. Journal of the American Chemical Society, 70(10), 3151-3155.
PF3D7_1434600	Methionine aminopeptidase 2 (METAP2)	3.4.11.18	Chen, X., Xie, S., Bhat, S., Kumar, N., Shapiro, T. A., & Liu, J. O. (2009). Fumagillin and fumarranol interact with P. falciparum methionine aminopeptidase 2 and inhibit malaria parasite growth in vitro and in vivo. Chemistry & biology, 16(2), 193-202.
PF3D7_1439400	Ubiquinol-cytochrome c reductase iron-sulfur subunit, putative	1.10.2.2	Fieser, L. F., Berliner, E., Bondhus, F. J., Chang, F. C., Dauben, W. G., Ettlinger, M. G., ... & Heymann, H. (1948). Naphthoquinone Antimalarials. I. General Survey1, 2. Journal of the American Chemical Society, 70(10), 3151-3155.
PF3D7_1439900	Triose phosphate isomerase	5.3.1.1	Barker, R. H., Metelev, V., Rapaport, E., & Zamecnik, P. (1996). Inhibition of Plasmodium falciparum malaria using antisense oligodeoxynucleotides. *Proceedings of the National Academy of Sciences*, *93*(1), 514-518.;
PF3D7_1440300	Delta-aminolevulinic acid dehydratase	4.2.1.24	Bonday, Z. Q., Dhanasekaran, S., Rangarajan, P. N., & Padmanaban, G. (2000). Import of host δ-aminolevulinate dehydratase into the malarial parasite: identification of a new drug target. Nature medicine, 6(8), 898-903.
PF3D7_1444300	1-acyl-sn-glycerol-3-phosphate acyltransferase, putative (LPAAT)	2.3.1.51	Lindner, S. E., Sartain, M. J., Hayes, K., Harupa, A., Moritz, R. L., Kappe, S. H., & Vaughan, A. M. (2014). Enzymes involved in plastid-targeted phosphatidic acid synthesis are essential for Plasmodium yoelii liver-stage development. Molecular microbiology, 91(4), 679-693.

Gene ID	Reaction/Gene name	EC number	References
PF3D7_1444800	Fructose-bisphosphate aldolase	4.1.2.13	Wanidworanun, C., Nagel, R. L., & Shear, H. L. (1999). Antisense oligonucleotides targeting malarial aldolase inhibit the asexual erythrocytic stages of Plasmodium falciparum. Molecular and biochemical parasitology, 102(1), 91-101.; Jewett, T. J., & Sibley, L. D. (2003). Aldolase forms a bridge between cell surface adhesins and the actin cytoskeleton in apicomplexan parasites. Molecular cell, 11(4), 885-894.
PF3D7_1446200	M17 leucyl aminopeptidase	3.4.11.1	Nankya-Kitaka, M. F., Curley, G. P., Gavigan, C. S., Bell, A., & Dalton, J. P. (1998). Plasmodium chabaudi and P. falciparum: inhibition of aminopeptidase and parasite growth by bestatin and nitrobestatin. Parasitology research, 84(7), 552-558.
PF3D7_1446800	Heme detoxification protein (HDP)	4.99.1.8	Martiney, J. A., Cerami, A., & Slater, A. F. (1996). Inhibition of hemozoin formation in Plasmodium falciparum trophozoite extracts by heme analogs: possible implication in the resistance to malaria conferred by the beta-thalassemia trait. Molecular Medicine, 2(2), 236.
PF3D7_1453800	Glucose 6-phosphate dehydrogenase-6-phosphogluconolactone (GluPho)	3.1.1.31	Preuss, J., Hedrick, M., Sergienko, E., Pinkerton, A., Mangravita-Novo, A., Smith, L., ... & Becker, K. (2012). High-throughput screening for small-molecule inhibitors of Plasmodium falciparum glucose-6-phosphate dehydrogenase 6-phosphogluconolactonase. Journal of biomolecular screening, 17(6), 738-751.
PF3D7_1462800	Glyceraldehyde-3-phosphate dehydrogenase	1.2.1.12	Robien, M. A., Bosch, J., Buckner, F. S., Van Voorhis, W. C., Worthey, E. A., Myler, P., ... & Lauricella, A. (2006). Crystal structure of glyceraldehyde-3-phosphate dehydrogenase from Plasmodium falciparum at 2.25 Å resolution reveals intriguing extra electron density in the active site. *Proteins: Structure, Function, and Bioinformatics*, *62*(3), 570-577.

Gene ID	Reaction/Gene name	EC number	References
PF3D7_1467300	1-deoxy-D-xylulose 5-phosphate reductoisomerase	1.1.1.267	Jomaa, H., Wiesner, J., Sanderbrand, S., Altincicek, B., Weidemeyer, C., Hintz, M., ... & Soldati, D. (1999). Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. Science, 285(5433), 1573-1576.
PF3D7_1469600	Biotin carboxylase subunit of acetyl CoA carboxylase, putative	6.3.4.14	Waller, R. F., Ralph, S. A., Reed, M. B., Su, V., Douglas, J. D., Minnikin, D. E., ... & McFadden, G. I. (2003). A type II pathway for fatty acid biosynthesis presents drug targets in Plasmodium falciparum. *Antimicrobial agents and chemotherapy*, *47*(1), 297-301.
PF3D7_1472200	Histone deacetylase, putative	3.5.1.98	Singh, S. B., Zink, D. L., Liesch, J. M., Mosley, R. T., Dombrowski, A. W., Bills, G. F., ... & Goetz, M. A. (2002). Structure and chemistry of apicidins, a class of novel cyclic tetrapeptides without a terminal α-keto epoxide as inhibitors of histone deacetylase with potent antiprotozoal activities. *The Journal of organic chemistry*, *67*(3), 815-825.; Mai, A., Cerbara, I., Valente, S., Massa, S., Walker, L. A., & Tekwani, B. L. (2004). Antimalarial and antileishmanial activities of aroyl-pyrrolyl-hydroxyamides, a new class of histone deacetylase inhibitors. Antimicrobial agents and chemotherapy, 48(4), 1435-1436.
PF3D7_1472900	Dihydroorotase, putative	3.5.2.3	Seymour, K. K., Yeo, A. E. T., Rieckmann, K. H., & Christopherson, R. I. (1997). dCTP levels are maintained in Plasmodium falciparum subjected to pyrimidine deficiency or excess. Annals of Tropical Medicine & Parasitology, 91(6), 603-609.; Krungkrai, J., Krungkrai, S. R., & Phakanont, K. (1992). Antimalarial activity of orotate analogs that inhibit dihydroorotase and dihydroorotate dehydrogenase. Biochemical pharmacology, 43(6), 1295-1301.

Gene ID	Reaction/Gene name	EC number	References
PF3D7_1476700	Lysophospholipase, putative	3.1.1.5	Zidovetzki, R., Sherman, I. W., Prudhomme, J., & Crawford, J. (1994). Inhibition of Plasmodium falciparum lysophospholipase by anti-malarial drugs and sulphydryl reagents. Parasitology, 108(03), 249-255.
PF3D7_1476800	Lysophospholipase, putative	3.1.1.5	Zidovetzki, R., Sherman, I. W., Prudhomme, J., & Crawford, J. (1994). Inhibition of Plasmodium falciparum lysophospholipase by anti-malarial drugs and sulphydryl reagents. Parasitology, 108(03), 249-255.
PFC10_API0015	RNA polymerase B (rpoB)	2.7.7.6	Lin, Q., Katakura, K., & Suzuki, M. (2002). Inhibition of mitochondrial and plastid activity of Plasmodium falciparum by minocycline. FEBS letters, 515(1-3), 71-74.