UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

iTRAQ-mediated proteome analysis of the response of *Plasmodium falciparum* to a novel antimalarial compound

Ntombikayise Precious Nkomo

Submitted in partial fulfilment of the requirements for the degree

Magister Scientiae

(Microbiology)

In the Faculty of Natural and Agricultural Sciences, Department of

Microbiology and Plant Pathology

University of Pretoria

Pretoria

March 2009

Supervisor: Prof. Jacques Theron

Co-supervisors: Prof. LynMarie Birkholtz

Dr. Dalu Mancama

© University of Pretoria

DECLARATION

I, the undersigned, hereby declare that the dissertation submitted herewith for the degree MSc. to the University of Pretoria, contains my own independent work and has not been submitted for any degree at any other university.

_____

Ntombikayise Precious Nkomo

2009

TABLE OF CONTENTS

DECLARATION ii
LIST OF TABLES AND FIGURES vi
Table 1: Summary of proteins identified following separation by SDS-PAGE and 1D-RP- vii
HPLC/MS/MS vii
Table 2: Summary of proteins identified following separation by 2-D SXC-RP HLPC vii
MS/MS vii
ACKNOWLEDGEMENTS ix
ABBREVIATIONS x
PREFACE xii
CHAPTER 1: LITERATURE REVIEW 1
1 INTRODUCTION 2
1.1 Transmission and exposure 5
1.2 Malaria pathologies 7
1.2.1 Anaemia 7
1.2.2 Hypoglycemia 8
1.2.3 Cerebral malaria 8
1.3 Drugs used in the treatment of malaria 9
1.3.1 Mechanism of action of antifolates 11
1.3.2 Mechanism of action of chloroquine 12
1.3.3 Artemisinin and its derivatives 14
1.3.4 Mechanism of action of artemisinin and its derivatives 15
1.4 Plants as sources of antimalarial drugs 16
1.4.1 Reference #31 (Ref#31) 17
1.5 Proteomics profiling and its use in elucidating the effects of drug pertubation 19
1.6 Aims 24
1.6.1 Objectives: 24
CHAPTER 2: SEPARATION AND MASS SPECTROMETRY ANALYSIS OF *P. FALCIPARUM* PROTEINS USING SDS-PAGE AND ION EXCHANGE CHROMATOGRAPHY 26

ABSTRACT.....27
2 INTRODUCTION.....28
2.1 MATERIALS AND METHODS.....34
2.1.1 Labelling of proteins with iTRAQ reagents for SDS-PAGE separation 34
2.1.2 Separation of protein mixtures by gel electrophoresis (SDS-PAGE) and in-gel digestion of proteins.....35
2.1.3 Labelling of proteins with iTRAQ reagents for strong cationic exchange 37
2.1.4 Sample optimisation.....37
2.1.5 LC/MS/MS.....38
2.2 RESULTS.....42
2.2.1 Protein separation by SDS-PAGE.....42
2.2.2 One-Dimensional Liquid Chromatography (RP-HPLC).....43
Table 1: Summary of proteins identified following separation by SDS-PAGE 46
2.2.3 Two-Dimensional Liquid Chromatography (SCX-RP-HPLC).....47
Table 2: Summary of proteins identified following separation by 2-D SXC-RP HLPC MS/MS.....49
2.2.4 Depletion of haemoglobin to enhance detection of *Plasmodium* proteins 54
2.3 DISCUSSION.....55
CHAPTER 3: INVESTIGATING THE MORPHOLOGICAL EFFECT OF A NOVEL ANTIMALARIAL COMPOUND REF#31 ON THE INTRAERYTHROCYTIC DEVELOPMENT CYCLE OF *PLASMODIUM FALCIPARUM*, AND IDENTIFICATION OF PROTEINS DIFFERENTIALLY EXPRESSED BY *PLASMODIUM FALCIPARUM* AFTER TREATMENT WITH REF#31.....67
3 INTRODUCTION.....66
3.1 MATERIALS AND METHODS.....69
3.1.1 Drugs.....69
3.1.2 Morphological study of the drug (Ref#31).....69
3.1.3 Culturing of parasites and drug treatment (Ref#31).....70
3.1.4 Protein extraction.....71

3.1.5 Determination of protein concentration 72
3.1.6 Depletion of haemoglobin to enhance protein identification 72
3.1.7 Separation of protein mixtures by gel electrophoresis (SDS-PAGE) and modified in-gel digestion 73
3.1.8 One-Dimensional Liquid Chromatography (RP-HPLC) 74
3.1.9 MS data analysis 75
3.2 RESULTS 76
3.2.1 Effect of Ref#31 on the life cycle of *P. falciparum* strain 3D7 76
3.2.2 Protein Separation by SDS-PAGE 79
3.2.3 One-Dimensional Liquid Chromatography (RP-HPLC) 81
3.2.4 Differential protein expression analysis following Ref#31 treatment 85
3.2.5 Post-translational analysis of protein modifications following Ref#31 treatment 90
3.3 DISCUSSION 98
4 SUMMARY 104
5 REFERENCES 108
APPENDIX 1 (Supplementary CD)
APPENDIX 2 (Supplementary CD)

LIST OF TABLES AND FIGURES

CHAPTER 1

Figure 1: Global distribution of malaria.

Figure 2: Life cycle of *Plasmodium falciparum* (Johnson, 2004).

Figure 3: Chemical structures of antimalarial antifolates.

Figure 4: The structures of quinine and its derivatives (Ridley, 2002).

Figure 5: The structure of artemisinin (A) and its derivatives (B) (Balint, 2001).

Figure 6: *Dicoma anomala*.

Figure 7: Nature of the chemical compound Ref#31.

Table 1: Bioassay results: *In vitro* data

CHAPTER 2

Figure 1: The concept of iTRAQ reagent chemistry (Gluckmann *et al.*, 2007)**.**

Figure 2: iTRAQ reagent structure showing the fragmentation sites during MS/MS (Applied Biosystems, 2004).

Figure 2.1: MudPIT (Multidimensional Protein Identification Technology).

Figure 3: SDS-PAGE of iTRAQ-labelled samples.

Figure 4: Nano-ESI-MS/MS spectra of the proteolytic peptide FFESFGDLSTPDAVMGNPK.

Figure 5: Nano-ESI-MS/MS spectra of the proteolytic peptide THINLVVIGHVDSGK$^{(iTRAQ}$

Figure 6: An illustration of the number of proteins identified by 2-D SCX-RP chromatography coupled to MS/MS.

Figure 7: Protein page-stained SDS-PAGE analysis of the depletion of Hb (haemoglobin).

Table 1: Summary of proteins identified following separation by SDS-PAGE and 1D-RP-HPLC/MS/MS

Table 2: Summary of proteins identified following separation by 2-D SXC-RP HLPC MS/MS

CHAPTER 3

Figure 1: Morphology study of *Plasmodium falciparum* 3D7.

Figure 2: Protein page-stained SDS-PAGE analysis of 6 hr control versus drug-treated samples.

Figure 3: Protein page-stained SDS-PAGE analysis of 12 hr control versus drug treated samples.

Figure 4: Protein page-stained SDS-PAGE analysis of 18 hr control versus drug-treated samples.

Figure 5: Comparison of two databases (PlasmoDB v5.5 and MSDB database).

Figure 6: Illustration of the summary statistics generated for each protein.

Figure 7: Venn diagram summarising proteins estimated to be differentially expressed due to the effects of Ref#31 treatment, time, and interaction between these two variables.

Figure 8: Proportion of *P. falciparum* peptide sequences observed to be acetylated.

Figure 9: Proportion of *P. falciparum* peptide sequences observed to be methylated.

Figure 10: Proportion of *P. falciparum* peptide sequences observed to be ubiquitinated.

Figure 11: Proportion of *P. falciparum* peptide sequences observed to be phosphorylated.

Figure 12: Illustration of the acetylation and methylation status of amino acid residues belonging to histone H2A.

Table 1: *P. falciparum* proteins found to be differentially expressed greater or lower than 1.4-fold following treatment with Ref#31

APPENDIX 1: (Supplementary CD)

Table1: Summary of proteins identified following separation by 2-D chromatography (MS/MS)

Table 2: Summary of six protein mix identified by trypsin in-gel digestion

APPENDIX 2: (Supplementary CD)

Table 1: Proteins identified using the PlasmoDB v 5.5 database

Table 2: Proteins identified using the MSDB database

Table 3: Summary of proteins identified by trypsin in-gel digestion (6 hr control) (1D RP-HPLC)

Table 4: Summary of proteins identified by trypsin in-gel digestion (6 hr drug-treated) 1D RP-HPLC

Table 5: Summary of proteins identified by trypsin in-gel digestion (12 hr control) (1D RP-HPLC)

Table 6: Summary of proteins identified by trypsin in-gel digestion (12 hr drug-treated) (1D RP- HPLC)

Table 7: Summary of proteins identified by trypsin in-gel digestion (18 hr control) (1D RP-HPLC)

Table 8: Summary of proteins identified by trypsin in-gel digestion (18 hr drug-treated) (1D RP-HPLC)

Table 9: Differentially expressed *P. falciparum* proteins clustered by treatment with Ref#31

Table 10: Differentially expressed *P. falciparum* proteins clustered by time

Table 11: Differentially expressed *P. falciparum* proteins clustered by treatment and time interaction

Table 12: Detected Acetylation sites (an addition of an acetyl group, either at the N-terminus of the protein or at lysine)

Table 13: Detected Methylation sites (an addition of a methyl group, usually at lysine or arginine residues)

Table 14: Detected Ubiquitination sites (post-translational modification of a protein by covalent attachment (via an isopeptide bond one or more ubiquitin monomers)

Table 15: Cleavage of alpha tubulin C-terminal tyrosine residue by tubulin carboxypeptidase

Table 16: Detected phosphorylation sites (Addition of a phosphate (PO_4) group)

ACKNOWLEDGEMENTS

I would like to express my appreciation and thanks to the following people:

Dr. Dalu Mancama, Professor Jacques Theron, Prof. Lyn-Marie Birkholtz. Thank you for your guidance during the preparation of this dissertation. It was a great privilege to work with you!

The Systems Biology group at the CSIR for support and guidance.

Council for Scientific and Industrial Research (CSIR) for financial support.

Dr. Stoychev, this would not have been possible without your help, thanks!

Sinky, you have always supported, encouraged and believed in me. Thank you for the love and support.

To my family, I would not have done this without your love and support, thank you.

ABBREVIATIONS

ACN	Acetonitrile
COOH	Carboxyl group
CQ	Chloroquine
CQR	Chloroquine resistance
DHFR	Dihydrofolate reductase
DHPS	Dihydropteroate synthase
DIGE	Differential In-Gel Electrophoresis
DMSO	Dimethylsulfoxide
ESI	Electrospray ionisation
FA	Formic acid
H_3PO_4	Phosphoric acid
Hb	Haemoglobin
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
Hpi	Hours post inhibition
HPLC	High performance liquid chromatography
IC_{50}	Half maximal inhibitory concentration
IC_{99}	Inhibitory concentration 99
ICAT	Isotope-Coded Affinity Tags
IDA	Information dependent acquisition
iTRAQ	Isobaric tags for relative and absolute quantification
KCl	Potassium chloride
kDa	Kilodaltons
KH_2PO_4	Monopotassium phosphate

MS	Mass Spectrometry
MudPIT	Multidimensional protein identification technology
NI-NTA	Nickel-Nitrilotriacetic acid
PRBC	Parasitized red blood cells
PfCRT	*Plasmodium falciparum* chloroquine resistance transporter
Pfmdr1	*Plasmodium falciparum* multiple drug resistance gene 1
PTM	Post-translational modifications
RBC	Red blood cells
Ref#31	Reference #31
RP	Reverse Phase
RPMI	Roswell Park Memorial Institute medium
SILAC	Stable isotope labelling with amino acids in cell culture
SCX	Strong Cation Exchange
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SP	SDX and PYR combination, Fansidar
SPR	SP resistance
TCP	Tubulin carboxypeptidase
TFA	Trifluoroacetic acid
2DGE	Two-dimensional differential gel electrophoresis

PREFACE

The malaria parasite *Plasmodium falciparum* is one of four species of *Plasmodium* in the phylum *Apicomplexa* that infect humans, and is responsible for virtually all deaths from this disease, as well as high levels of morbidity. Malaria has a profound impact on the public health and education systems, economies and social structures of many developing countries, particularly Africa (Sims *et al.*, 2006). Proteomic analysis of this deadly parasite began nearly two decades ago with investigations of 2D gel patterns of parasite proteins before and after the application of antimalarial drugs (Pinswasdi *et al.*, 1987). Current efforts focus on research into novel compounds and on measures to prevent or delay resistance once new drugs are introduced. To develop new antimalarial drugs, ethnobotanical investigation in traditional medicine can be an important source of new leads. African traditional medicine uses numerous plants that can be a source of new antimalarials (Karou *et al.*, 2003). The aim of this research was to develop and employ new protocols in proteomics to investigate the effects on *P. falciparum* of a novel naturally derived antimalarial compound, Reference #31 (Ref#31). This knowledge contributes towards better understanding the mechanisms underlying the efficacy of Ref#31, and may ultimately be useful to improving the compound's further development as an antimalarial.

Chapter 1: this chapter introduces the novel antimalarial compound Ref#31, iTRAQ technology, and different techniques that have been used over the years to study the complex proteome of *P. falciparum*.

Chapter 2: describes the different approaches used to optimise iTRAQ technology and maximise the amount of *Plasmodium* hits identified. Cation exchange and SDS-PAGE were used to fractionate protein samples. Data analysis was performed with Protein Pilot 2.0 (Applied Biosystems). It also describes an in-house approach to improve parasite proteome

characterisation, by depleting levels of haemoglobin present in cell culture protein lysates. The approach uses Ni-NTA (Nickel-Nitrilotriacetic acid) agarose (Qiagen), a tetradentate chelating adsorbent.

Chapter 3: investigates the morphological effects of the novel plant-derived antimalarial lead Ref#31, on the parasite's intraerythrocytic development cycle. Cultured *P. falciparum* parasites were treated for 48 hrs with different drug concentrations and the parasitemia was determined by Giemsa stain and light microscopy. Blood smears were taken every 6 hr for a period of 48 hrs. The second half of this work describes the analysis of parasite proteins following parasite treatment with Ref#31. Three time points were analysed, 6, 12 and 18 hrs.

CHAPTER 1: LITERATURE REVIEW

1 INTRODUCTION

One of the oldest infectious diseases on the planet, malaria still casts a formidable shadow on many parts of the world, causing hundreds of millions of new infections and millions of deaths each year (Breman *et al.,* 2001). There were an estimated 247 million malaria cases among 3.3 billion people at risk in 2006, causing nearly a million deaths, mostly of children under 5 years (WHO, 2008). Malaria not only causes considerable problems to human health and an economic burden in developing countries, but with increasing global travel it also poses a threat to human health worldwide. The African continent shows the greatest effects of this disease, accounting for more than 90% of the burden. Despite advances in understanding malaria aetiology and development of interventions, more than 50% of the world's population is exposed to malaria (Breman *et al.,* 2004). Combating malaria requires significant financial and organisational resources, yet malaria itself restrains economical development, creating a vicious circle in developing countries (Pink *et al.,* 2002). The devastating socio-economic and public health impact of malaria is mostly experienced in sub-Saharan Africa, and is galvanised by the emergence and rapid spread of drug-resistant parasites and the lack of a licensed vaccine (Birkholtz *et al.,* 2008*).* Malaria is caused by the infection of hemaprotozoan parasite belonging to the genus *Plasmodium*. Four species of *Plasmodium* cause malaria in humans, namely *P. falciparum, P. vivax, P. ovale* and *P. malariae,* but infection with *P. falciparum* is the most dangerous, often leading to death if left untreated (Stevenson *et al.,* 2001). Over the past 50 years, many new drugs have been introduced to treat malaria infections. Unfortunately, parasite resistance to the most affordable antimalarial drugs has severely limited the available therapeutic options for Africa, where falciparum malaria takes its greatest toll (Bloland *et al.,* 2003). New and inexpensive drugs against malaria are therefore urgently needed in all parts of the world where the disease is endemic (Rosenthal *et al.,* 2001). A major step forward in the battle

against malaria was the recent completion of the genome sequence of *P. falciparum*. This has opened up the possibility of conducting genome-wide studies of transcription and translation in the parasite (Gardner *et al*., 2002).

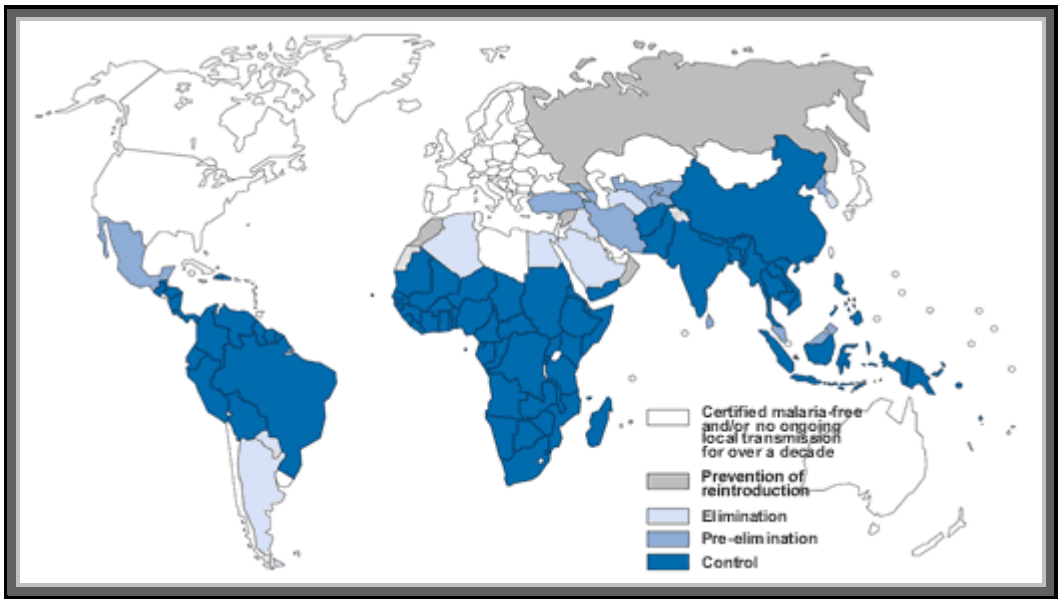


Figure 1: Global distribution of malaria. Malaria, 2008. (Source: World Health Organization, 2008).

The proteome of *P. falciparum*, a single-celled eukaryotic parasite, is highly dynamic due to its complex life cycle in two separate hosts and the numerous morphological stages, each with distinct patterns of protein expression that coordinate survival and reproduction within *Anopheles* and human hosts. This proteome is further complicated by the considerable degree of strain variation in *P. falciparum* populations (Sims *et al.*, 2006). Many *Plasmodium* proteins have low complexity regions consisting of long hydrophobic stretches, amino acid repeats or segments highly rich in amino acids encoded by AT-biased codons, notably lysine and asparagine (Vedadi *et al.*, 2007).

Fewer reports are available of the parasite's proteome than of the transcriptome due to technical constraints. The most comprehensive study to date of the *P. falciparum* life cycle

proteome provided information of 46% (roughly 2400 proteins) of all possible gene products using MudPIT (Florens *et al.,* 2002). Two-dimensional differential gel electrophoresis) 2DGE proteomics has been applied extensively to study proteins correlated with disease states, an earlier study using 2DGE investigated sex-specific and blood meal-induced proteins in the midgut of *Anopheles* (Johnson *et al.,* 2003). Sims and co-workers made progress in characterizing certain subpopulations of the *P. falciparum* proteome (Sims *et al.,* 2006). In 2008, Foth and co-workers used two-dimensional differential gel electrophoresis (2D-DIGE) in quantitative proteomics analyses to investigate the extent of post-transcriptional gene regulation and PTMs (Post-translational modifications) during the late phases of the *P. falciparum* intra-erythrocytic developmental cycle. A total of 623 proteins/isoforms analysed generated more than 9000 individual protein abundance measurements (Foth *et al.*, 2008).

1.1 Transmission and exposure

Malaria parasites are transmitted by the female of various species of *Anopheles* mosquitoes. Infection is initiated during blood-feeding by infected mosquitoes, which inoculate sporozoites into the mammalian host. Sporozoites migrate through the lymphatics and bloodstream to the liver and invade hepatocytes (Figure 2), where the parasite matures into merozoites, the so-called liver-stage of infection (Stevenson *et al.*, 2001). At the end of this pre-erythrocytic cycle, thousands of merozoites are released into the blood and infect human erythrocytes where they asexually reproduce in the intra-erythrocytic developmental cycle from ring, trophozoite and schizont stages, eventually producing new merozoites to infect uninfected erythrocytes. A subset of merozoites develops into gametocytes that are taken up by mosquitoes and develop into gametes that fertilize to form a zygote in the mosquito midgut (Johnson *et al.*, 2004).

There are many factors regulating the transmission of malaria, such as poverty, social and economical factors. The combination of these factors is different for every geographical region (WHO, 2008). Poor rural populations cannot afford adequate housing and insecticide impregnated bed nets that would protect them from exposure to mosquitoes, and they often lack the knowledge to recognize malaria and access to health care (WHO, 2008). Human activities can create breeding sites for larvae such as standing water, irrigation ditches and burrow pits. Agricultural work such as harvesting, irrigation schemes, population movements and deforestation all contribute to deteriorating situations (Nosten *et al.*, 2002). Raising domestic animals near the household may provide alternative sources of blood meals for *Anopheles* mosquitoes and thus decrease human exposure (Sachs *et al.*, 2002). Moreover, these combinations are not constant and are likely to change with the general weather conditions or modernization. Therefore, effective regional methods of

control of the infection might be useless in another region and the efficiency of the given methods might change with time. Control of the infection at the individual level, with drugs and vaccines, might thus be of advantage in combination with intervention at community level (Johnson *et al.*, 2004).

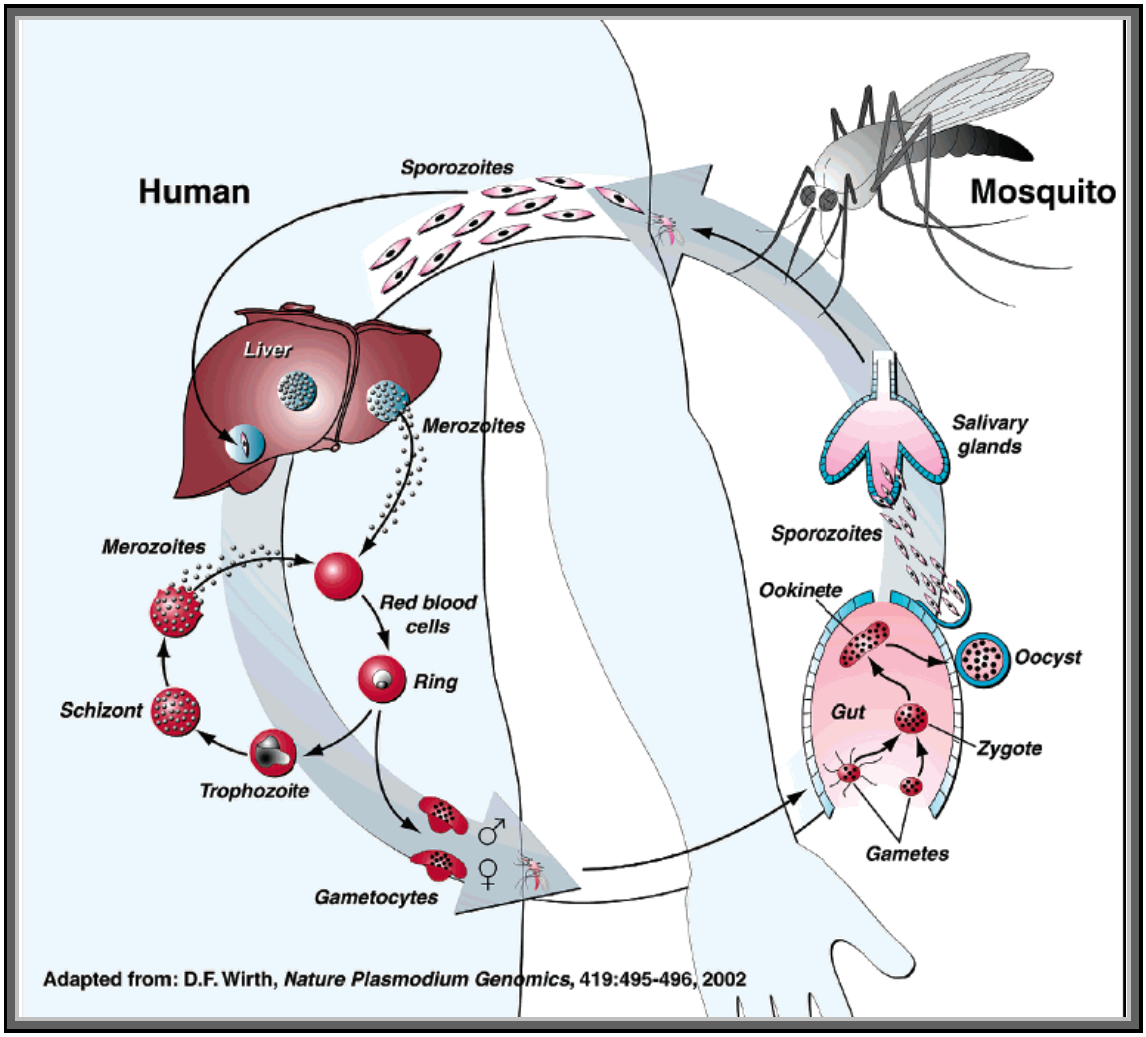


Figure 2: Life cycle of *P. falciparum* (Johnson *et al.*, 2004).

1.2 Malaria pathologies

1.2.1 Anaemia

Malaria is associated with several pathologies, including fever, haemoglobinuria, cerebral malaria, hypoglycaemia, pulmonary oedema and jaundice. Although some of these (like fever) are found in almost every *Plasmodia* infection, others (including cerebral malaria, severe anaemia and renal failure) are only found in a minority of patients (Schofield *et al.*, 2005).

Severe anaemia is one of the most serious complications of falciparum malaria. All children with clinically significant malaria have some degree of anaemia. The pathogenesis of malarial anaemia is complex. The anaemia of malaria is multifactorial, involving both the destruction of red blood cells (RBC) and the decreased production of RBC (Menendez *et al.*, 2000). With acute uncomplicated falciparum malaria, anaemia develops 48hrs after the onset of fever. Mechanisms responsible include rupture of parasitized RBC. The severity of anaemia in children and adults with acute falciparum malaria correlates with the density of parasitaemia and schizontaemia, and it may be deduced that rupture of parasitized RBC, both in the circulation and sequestered in deep vasculature, is possibly the most important mechanism of anaemia (Menendez *et al.*, 2000). The second mechanism involves destruction of parasitized and unparasitized red blood cells through phagocytosis. In response to acute malaria, there is proliferation and hyperactivity of macrophages in the reticuloendothelial system (RES), with phagocytosis of both parasitized and unparasitized RBC, which have abnormally rigid membranes (Menendez *et al.*, 2000; Collins *et al.*, 2003).

1.2.2 Hypoglycemia

Hypoglycemia is a common complication of severe falciparum malaria, occurring in 8–30% of individuals affected with this disease. It is usually ascribed to increased glucose use and impaired glucose production caused by the inhibition of gluconeogenesis (Thien *et al.*, 2006). Patients most vulnerable for this complication include pregnant women, patients with severe disease and patients on quinine therapy (Mishra *et al.*, 2006). The mechanism underlying hypoglycemia in malaria is not fully explained by pathophysiological data. Apart from the hyperinsulinemic hypoglycemia caused by quinine and quinidine treatment, the mechanisms suggested to underlie hypoglycemia include increased glucose consumption and insufficient or decreased production (White 2003; Thien *et al.*, 2006). Theoretically, increased glucose consumption in an individual with malaria could be due to increased consumption by the individual or increased consumption by the parasite, or a combination of both. Glucose is the obligatory fuel of *Plasmodia*, and parasitized erythrocytes consume up to 30–75 times more sugar than their uninfected counterparts (WHO, 2000). Quinine stimulates insulin release *in vivo* in both healthy humans and individuals affected with malaria. This insulin release occurs mainly in adults, especially in pregnant women (White 2003).

1.2.3 Cerebral malaria

Cerebral malaria is one of the most severe complications of *P. falciparum* infection. Cerebral malaria is caused by obstruction of small vessels of the brain by sequestered parasites. The symptoms of cerebral malaria range from confusion or stupor to obtundation and deep coma, and are frequently associated convulsions which may be subclinical (Crawley *et al.*, 2001). The *Plasmodium* parasite invades erythrocytes as part of its asexual life cycle within the human host. As the parasite matures within these parasitised red blood

cells (PRBC) to trophozoite and schizont stages, they disappear from the peripheral circulation and localise specifically in the vascular beds of key organs such as the brain, a process termed sequestration (Medana and Turner, 2006). Sequestration occurs principally during the second half of the intra-erythrocytic asexual growth phase of the parasite, following the adherence of mature parasites to endothelial cells through electron-dense knobs on the PRBC surface (Beeson *et al.,* 2002). This process (cytoadherence) is presumably advantageous to the parasite, since it prevents the passage of abnormal erythrocytes through the spleen (Philip and Rosenthal, 2008).

1.3 Drugs used in the treatment of malaria

Although malaria is a curable disease if promptly diagnosed and adequately treated, a limited number of drugs for its treatment are available. The majority of antimalarial drugs target the intra-erythrocytic stage of the parasite life cycle, resulting in prevention of the development of disease in humans. Quinine has been the mainstay of treatment of severe malaria since the introduction of Cinchona bark to European medicine in the 1630s. The drug is one of the four main alkaloids isolated from the bark of the *Cinchona* tree (Artesunate Malaria Trial (SEAQUAMAT) group, 2005). The value of chloroquine was recognised near the end of World War II (Wellems *et al.,* 2001). Today, there are chloroquine-resistant strains of *P. falciparum* throughout Africa (Nosten *et al.,* 2002). Intense demands for an effective quinine substitute launched the discovery and evaluation of a series of organic compounds. The success of chloroquine and its heavy use through the decades led to chloroquine resistance in *P. falciparum and P. vivax* (Wellems *et al.,* 2001; Payne 1987).

Sulfadoxine/pyrimethamine, which was widely recommended as alternative drug therapy, has been adopted as a first-line treatment in some countries of East Africa (Malawi, Kenya, Tanzania) but is becoming rapidly ineffective (Ogutu *et al.*, 2000). Sulfadoxine/ pyrimethamine has been considered a single drug rather than a combination in which each component acts independently, since, although pyrimethamine inhibits dihydrofolate reductase (DHFR) and sulfadoxine inhibits dihydropteroate synthase (DHPS), inhibition of both enzymes prevents synthesis of folic acid in parasites (Kremsner *et al.*, 2004). Pyrimethamine is also structurally related to proguanil, and to cycloguanil and chlorproguanil, other drugs that inhibit DHFR. Its use as a prophylactic agent for malaria began half a century ago (Kremsner *et al.*, 2004).

Mefloquine, an antimalarial drug that is structurally related to quinine, was introduced for therapy in 1971. It has been found to be effective against all malarial species, including *P. falciparum.* However, resistance has since been reported in various parts of the tropics in the naturally resistant parasite populations (Milijaona *et al.*, 2000; Calvosa *et al.*, 2001; Caniato *et al.*, 2003). The discovery of qinghaosu (artemisinin) in China in 1972 and the subsequent synthesis of artemether and artesunate have provided highly effective alternatives to quinine (WHO, 1993-2000). Artemisinin or qinghaosu, an endoperoxide sesquiterpene lactone produced by aerial parts of *Artemisia annua*, is effective even against multi-drug resistant strains of the malarial parasite (Dhingra *et al.*, 2000). There is currently limited, if any, resistance to artemisinins in malaria parasites (Philip *et al.*, 2008). Although all artemisinins have rapid antiparasitic activity, they have short half-lives, such that the standard 3-day treatment course is commonly followed by recrudescence of infecting parasites and recurrent illness within days to weeks (Woodrow *et al.*, 2005). The short half-lives of artemisinins limit the possibility of selection for resistance. Nonetheless, recent

heavy use of artemisinins, including monotherapy, has created selective pressure. Some parasites isolated from French Guiana and Senegal recently showed diminished *in vitro* sensitivity to artemether, (Jambou *et al.,* 2005) and the efficacies of artemisinin-based combination agents have apparently decreased along the Thailand–Cambodia border (Wongsrichanalai *et al.,* 2008; Philip *et al.,* 2008).

1.3.1 Mechanism of action of antifolates

The spread of chloroquine resistance (CQR) has resulted in the use of the sulfadoxine (SDX) and pyrimethamine (PYR) combination, Fansidar SP (Figure 3). Unfortunately, resistance to SP (SPR) was reported in Thailand soon after it was introduced (Hayton *et al.,* 2004; Reacher *et al.,* 1981), and now SPR is spread in South America and South East Asia. Antifolate antimalarial drugs interfere with folate metabolism, a pathway essential to malaria parasite survival. This class of drugs includes effective causal prophylactic and therapeutic agents, some of which act synergistically when used in combination. Resistance is caused by point mutations in dihydrofolate reductase (DHFR) and dihydropteroate synthase (DHPS), the two key enzymes in the folate biosynthetic pathway that are targeted by the antifolates. Sulfadoxine inhibits dihydropteroate synthase, a key enzyme in the biosynthesis of folate. Pyrimethamine is a competitive inhibitor of DHFR, a key enzyme in the redox cycle for production of tetrahydrofolate (Sibley *et al.,* 2001). The DHFR and DHPS inhibitors inhibit DNA synthesis, and their toxic effect on the parasite reaches a peak in the late erythrocytic schizont stage, precisely when DNA synthesis peaks (Gregson *et al*., 2005).

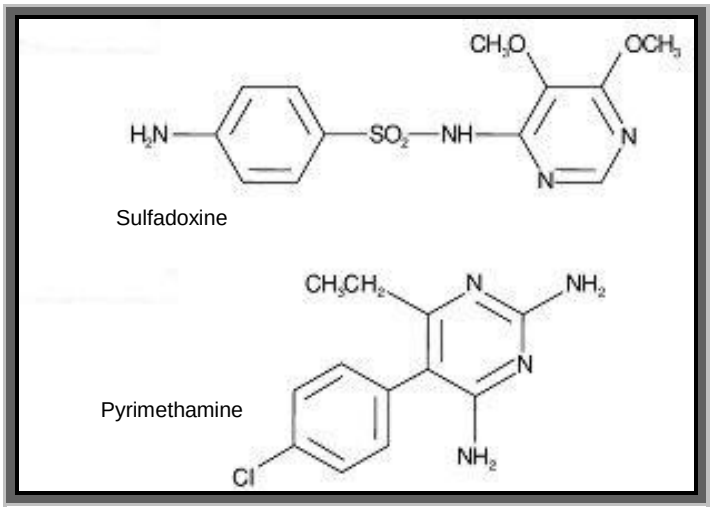


Figure 3: Chemical structures of antimalarial antifolates (Sibley *et al.*, 2001).

1.3.2 Mechanism of action of chloroquine

Chloroquine is a member of the quinoline–containing antimalarials, which include quinine and its derivatives mefloquine, amadiaquine and chloroquine (Figure 4). The need to understand the mechanisms of action of the quinoline antimalarials is becoming more urgent as levels of resistance to these drugs increase. A number of potential mechanisms of action have previously been proposed for chloroquine; however, chloroquine is now widely believed to act by interfering with heme metabolism in the digestive food vacuole of *P. falciparum* (Lamarque *et al.*, 2008). The food vacuole (FV) is a lysosome-like organelle where erythrocyte hemoglobin digestion occurs. It is a favourite target in the development of antimalarials (Lamarque *et al.*, 2008).

Resistance has been genetically linked to mutations in a parasite protein (PfCRT) that might confer resistance by inhibiting chloroquine accumulation in infected erythrocytes (Petri, 2003). Recent genetic transformation experiments of *P. falciparum* parasites provided evidence that mutations in PfCRT can affect parasite CQ response (Hayton *et al.*,

2004). Verapamil, which inhibits P-glycoprotein–mediated multidrug resistance (encoded by *mdr*) in mammalian tumour cells, partly reverses chloroquine resistance in malaria parasites grown *in vitro* (Djimde *et al.*, 2001). In *P. falciparum, mdr* homologues encoding P-glycoprotein–like molecules have been proposed as determinants of chloroquine resistance, and associations have been reported between chloroquine resistance and amplification or mutation of the *mdr*-like gene Pfmdr 1, which encodes Pgh1 (Foot *et al.*, 1990, Djimde *et al.*, 2001).

Chloroquine-resistant parasites accumulate chloroquine much less efficiently in their acidic food vacuoles than chloroquine-sensitive strains, suggesting that drug resistance results mainly from exclusion of the drug from its site of action rather than an alteration in the chloroquine target, although a decreased sensitivity of the target may also be involved (Hayton *et al.*, 2004).

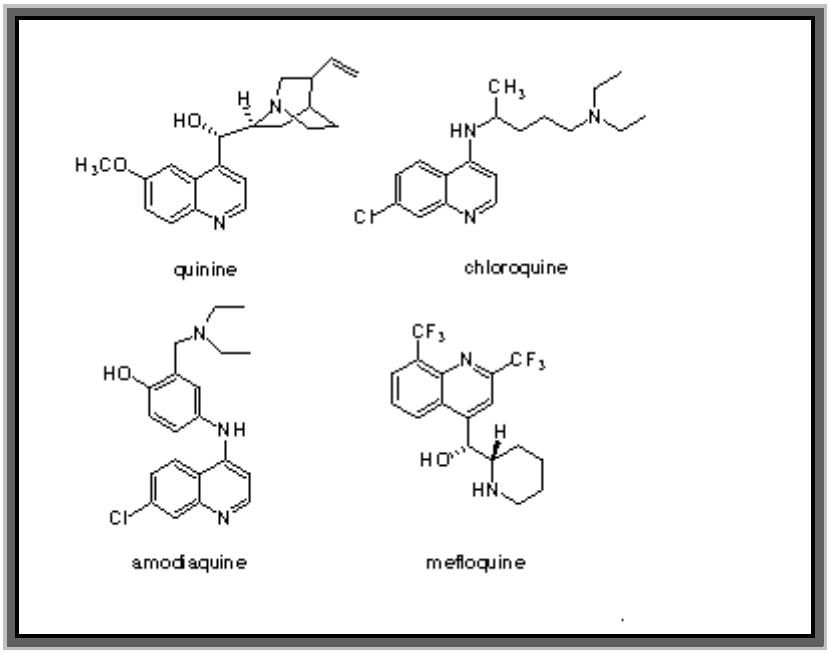


Figure 4: The structures of quinine and its derivatives (Ridley, 2002).

1.3.3 Artemisinin and its derivatives

Artemisia annua, a plant belonging to the Asteraceae family, is an annual herb native to China and occurs naturally as a part of steppe vegetation in northern parts of the Chahar and Suiya provinces (Dhingra *et al.*, 2000). Artemisinin, or qinghaosu, is effective even against multi-drug resistant strains of the malaria parasite (Figure 5) (Dhingra *et al.*, 2000). The isolation and characterization of the endoperoxide sesquiterpene lactone, artemisinin, from the aerial parts of *A. annua* leaves, is considered as one of the most novel discoveries in recent medicinal plant research (van Agtmael *et al.*, 1999). Artemisinin has a peroxide bridge to which its antimalarial properties are attributed. It has a unique structure and lacks a nitrogen-containing heterocyclic ring, which is found in most antimalarial compounds. The drug can be easily reduced with sodium borohydride, resulting in the formation of dihydroartemisinin, which has better antimalarial activity *in vitro* than artemisinin itself (van Agtmael *et al.*, 1999). Several semi-synthetic and synthetic derivatives of artemisinin are currently being developed or actively being used to treat malaria (Olliaro *et al.*, 2003).

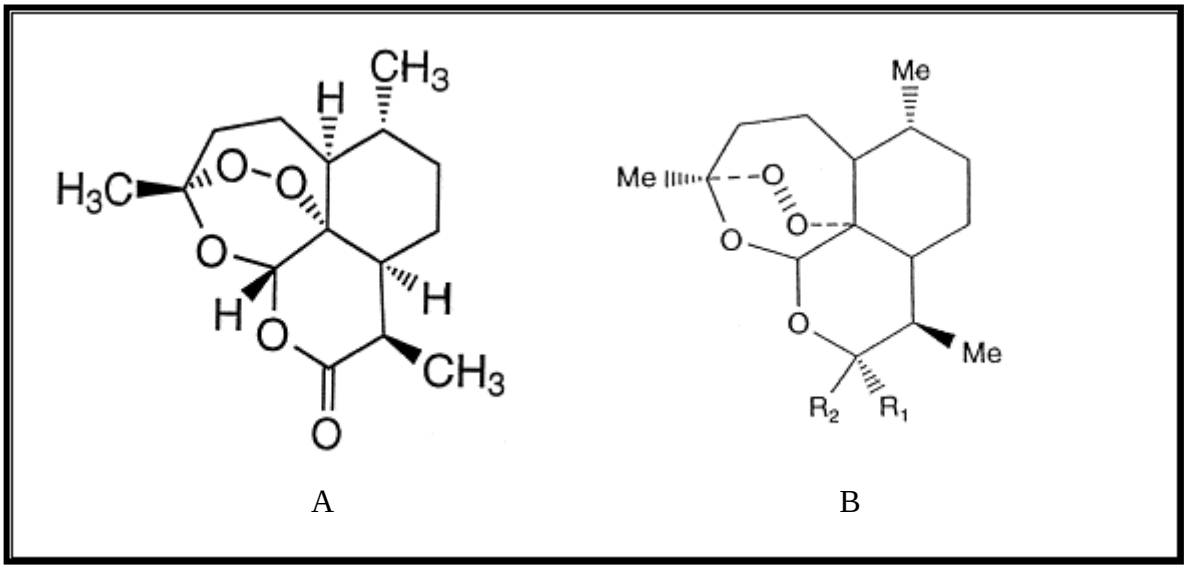


Figure 5: The structure of artemisinin (A) and its derivatives (B) (Balint, 2001).

R_1= H_1 R2 = OH Dihydroartemisinin, R_1= H_1 R2 = OMe Artemether, R_1= H_1 R_2 = OEt Arteether, R_1= H_1 R_2 = $OCH_2C_6H_5CO_2H$ Artelinic acid,

R_2= H_1 R1 = $O_2CCH_2CH_2CO_2H$

1.3.4 Mechanism of action of artemisinin and its derivatives

Artemisinin and its derivatives kill malaria parasites more rapidly than any other antimalarial agent due to its unique pharmacodynamic action. They have cured millions of cases of malaria; they are also active against all drug-resistant parasite strains and are evolving into a worldwide mainstay (Cooper *et al.*, 2004). Artemisinin derivative-based combination therapy (ACT) has been advocated as the therapy of choice to handle widespread drug resistance in *P. falciparum* malaria, at the same time preventing recrudescence due to artemisinin monotherapy. A general issue with many of these combinations is side-effects and pharmacokinetic mismatch between artemisinin and the other drugs in the combination. Another major concern of all these combinations is their high cost (Kremsner *et al.*, 2004). The artemisinin derivative-based combinations under development are artesunate plus amodiaquine, artesunate plus sulfadoxine-pyrimethamine, artesunate plus mefloquine, dihydroartemisinin plus piperaquine, and artemether plus lumefantrine (Toovey *et al.*, 2004). The artemether plus lumefantrine combination is the only formulation that has been registered for use in Africa, although a study has shown irreversible hearing impairment (Toovey *et al.*, 2004). In Cambodia in 2000, artesunate-mefloquine became the first-line drug for the treatment of falciparum malaria. Unfortunately there have been reports of emerging resistance to this combination of antimalarials. Clinical and molecular evidence indicate that artesunate-mefloquine failures are occurring on the Cambodia–Thailand border (Wongsrichanalai *et al.*, 2008).

Earlier studies suggested that artemisinin acts by haem-dependent activation of the endoperoxide bridge occurring within the parasite's food vacuole (van Agtmael *et al.*, 1999). Most recently, Eckstein-Ludwig and colleagues suggested the ATPase6 (SERCA)-type protein as the major target of artemisinin in *P. falciparum*. Their investigations have shown that artemisinin completely inhibits the SERCA orthologue (PfATP6) of *P.*

falciparum expressed in *Xenopus* oocytes with similar potency to thapsigargin (another sesquiterpene lactone and highly specific SERCA inhibitor). Desoxyartemisinin lacks an endoperoxide bridge and is ineffective both as an inhibitor of PfATP6 and as an antimalarial. Chelation of iron by desferrioxamine abrogates the antiparasitic activity of artemisinins and correspondingly attenuates inhibition of PfATP6 (Eckstein-Ludwig *et al* 2003). Further investigations on resistance to artemisinin have also indicated that some polymorphisms in genes as PfCRT and MDR1 may modulate sensitivity to artemisinin. Several data support the hypothesis that PfCRT and MDR1 mutant alleles are associated with increased susceptibility of *P. falciparum* isolates to artemisinin (Price *et al.*, 2004). Most recently, mutations in PfATP6 have also been reported. Thus, a L263E mutation has been shown to alter artemisinin sensitivity of PfATP6 expressed in *Xenopus* oocytes; however, no changes have been reported in this protein in artemisinin-resistant parasites so far discovered in South-East Asia (Egan, 2009).

1.4 Plants as sources of antimalarial drugs

The success of the antimalarial drug quinine and the discovery of artemisinin, currently the most potent and useful antimalarial drug, both from plant sources, has led to the study of plant derivatives as antimalarial agents. Herbal remedies applied since ancient times for treating malaria has been widespread (Tran *et al.*, 2003). Many plants used in traditional medicines have in fact been found to be sources of safe drugs, from which several present-day pharmaceuticals have been developed. Plants are rich in a wide variety of secondary metabolites, such as tannins, terpenoids, alkaloids, and flavonoids, which have been found *in vitro* to have antimicrobial properties. The recognition and validation of traditional medicinal practices could lead to new plant-derived drugs (Tran *et al.*, 2003). Therefore, the search for novel properties in a variety of indigenous plant species in malaria endemic

countries promises to be a valid strategy in antimalarial drug discovery. Several studies have successfully identified compounds from plants that bear potential for antimalarial drug development. In 2003, for example, Tran and colleagues evaluated 42 plant extracts *in vitro* for antiplasmodial activity, of which 24 extracts were found to be active against *P. falciparum.* Similarly, anti-sporozoite activity has been reported for a plant, the so-called "Indian beer" (Ampelozizyphus amazonicus, *Rhamnaceae*), used in the Brazilian endemic area to prevent malaria.

1.4.1 Reference #31 (Ref#31)

Dicoma anomala ssp. *gerardii* is a highly variable plant characterized by its stiff, sharp-pointed bracts, large flower heads and mauve-white florets surrounded by many white bristles (Figure 6). The stem of this plant is thinly covered with hairs. Leaves are simple, alternate, stalk-less, linear or narrowly lanceolate, the upper surface glabrous, the lower surface white-hairy. Flower heads are terminal, solitary or in small groups and have stiff, narrow, sharply-pointed bracts and slender, white, mauve, purple or pink tubular florets. The pappus consists of numerous barbellate bristles. Flowering occurs from November to July, with peak flowering in February and March. It is widely distributed in sub-Saharan Africa, resulting in pronounced morphological variety. It grows in summer rainfall areas and occurs in stony grasslands, hillsides or flat grassland and in savanna, on doleritic or sandy soils south of the Sahara, at altitudes ranging from 165 to 2 075 m (Mnengwane *et al.*, 2007).

In South Africa it is widely distributed in Limpopo, the North-West, Gauteng, Mpumalanga, the Free State, the Northern Cape and KwaZulu-Natal. It does not occur in the winter rainfall areas of the Western Cape. In KwaZulu-Natal it is mainly recorded in

the Tugela Basin, on sandy soils. Plant parts of *D. anomala* are said to be used for coughs and respiratory complaints. Roots are sometimes chewed by the Xhosa to induce vomiting when there is suspicion that poisoned food has been ingested (Mnengwane *et al.*, 2007).

Figure 6: *Dicoma anomala* (South African National Biodiversity Institute) National Herbarium, Pretoria (PRE).

Ref#31 is a compound extracted from the plant *D. anomala*. This compound is extracted from the roots of the *Dicoma* plant (Figure 7). It was isolated at the CSIR and tested against *P. falciparum in vitro*, and was shown to have antimalarial properties, as well as antibacterial and anticancer activities (Vinesh Maharaj, CSIR Bioprospecting, personal communication). Preliminary tests were also done on this compound using mice models; the data provided the first evidence of the efficacy of the compound as an antimalarial *in vivo* (Vinesh Maharaj, CSIR Bioprospecting, personal communication). Biological assaying was conducted on this compound to confirm activity, for which the results are summarised in Table 1. Currently, no knowledge exists on how Ref#31 works and uncovering this information would be critical to accelerating the development of this

compound by enabling researchers to improve it's selectivity, and to potentially enable the development of novel drug classes using a rational drug discovery approach.

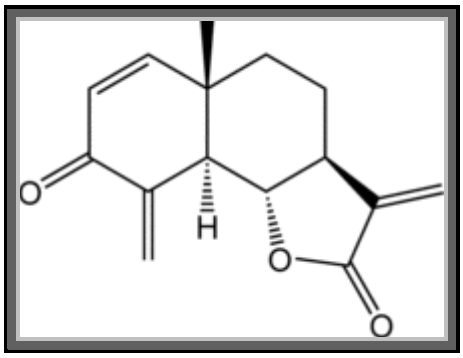


Figure 7: Structure of the chemical compound Ref#31.

Table 1: Bioassay results: *In vitro* data

Test sample	**D10 IC_{50} (µg/ml)**	**K1 IC_{50} (µg/ml)**	**CHO IC_{50} (µg/ml)**	**SI Index for D10**	**SI Index for KI**
Ref#31	0.27	0.22	4.2	16	19

SI (selective index) = cytotoxicity CHO IC_{50}/ antiplasmodial IC_{50}

D10= *P. falciparum* chloroquine sensitive strain

K1= *P. falciparum* chloroquine resistant strain

CHO= Cytotoxity against Chinese hamster ovarian cells

(Table 1: work done by the Bioprospecting group, CSIR)

1.5 Proteomics profiling and its use in elucidating the effects of drug pertubation

Proteins are responsible for ultimately carrying out the functions of genes, making these molecules key to deciphering biological function. The systematic analysis of the proteome (proteomics) is an innovative approach aimed at elucidating the complex interactions that occur between proteins. In particular, proteomics has become a powerful tool for predicting protein expression profiles, investigating protein localisation within cells and tissue, and

predicting the form and quantity in which proteins exist (Hood *et al.*, 2004). A global approach investigating the biological consequences of drug action is a potentially valuable strategy for understanding these complex interactions (Nirmalan *et al.*, 2004). Proteomics was thus chosen for this study to investigate the effects of Ref#31 on the malaria parasite *P. falciparum*, in order to gain a better understanding of the proteins and protein pathways affected by this compound. Recently, such global approaches to the study of drug activity, toxicity and resistance have increasingly turned to proteomics (Gardner *et al.*, 2002).

The first large-scale comparative study on the effects of treatment with the antimalarial drugs artemisinin and chloroquine on the proteome of *P. falciparum* was performed using metabolic labelling, followed by identification and quantification with MudPiT LC-MS (Prieto *et al.*, 2008). In 2005, Makanga and colleagues used proteomics to establish the relationship between the pharmacologically relevant concentration and time of exposure for the two components of CoArtem, artemether and lumefantrine, in terms of their inhibition concentrations, against *P. falciparum* cultures *in vitro* (Makanga *et al.*, 2005). They also provided qualitative and quantitative evidence for distinct proteomic changes induced by each drug. Recently, Foth and co-workers presented a quantitative time-course analysis of relative protein abundance for schizont-stage parasites (34 to 46 hours after invasion) based on two-dimensional differential gel electrophoresis of protein samples labeled with fluorescent dyes. Their final data indicated that both post-transcriptional and post-translational events are widespread and of presumably great biological significance during the intra-erythrocytic development of *P. falciparum* (Foth *et al.*, 2008).

Several technologies have been developed to enable large-scale proteomics profiling. These primarily include gel-based approaches such as 2-dimensional gel-based electrophoresis,

and high-throughput mass spectrometry-based approaches that employ isobaric tags. Modern genomics-based approaches such as proteomics offer increased hope for the discovery of promising new drug targets by virtue of their ability to characterise complex parasite biology and biochemistry. Proteomics is an increasingly popular approach to address both large and small-scale hypotheses related to microbial pathogens (Hunter *et al*., 2002). Its goal is the complete identification and quantification of a specific proteome, with the ultimate goal of revealing protein function as part of a complex, interrelated system.

A common technique for proteomic studies has 2DGE, where proteins are separated by isoelectric point (pI) and molecular weight, and spot patterns are compared by sophisticated computer algorithms (Righetti *et al.,* 2004). The spot density on the gel is used to assess relative quantification through comparison with "matched" protein spots on parallel run 2D gels. For protein staining, many protocols are in use (Righetti *et al.,* 2004), whereby in practice Coomassie Brilliant Blue and silver staining have found most widespread applications. However, these stains are not ideal because of their relatively poor detection sensitivity (Coomassie Brilliant Blue) or diminished peptide recovery from in-gel-digested proteins for MS (silver staining). Both Coomassie Brilliant Blue and silver staining also have a rather limited dynamic range as far as quantification is concerned. The accuracy of quantification depends on the intrinsic characteristics of the visualization methods. More recently, a variety of improvements and alternatives that are more reproducible and have an increased linear dynamic range have been introduced. Significant quality improvements have been achieved with the introduction of fluorescent stains like that of the SYPRO family (Yan *et al.*, 2000). Proteins of interest are then identified by mass spectrometry (MS). Despite its robust nature for separating complex protein mixtures, significant shortfalls exist in the ability of 2DGE to comprehensively profile the entire set of proteins

expressed in a given cellular system. These include shortcomings in resolving proteins with extremes in isoelectric point, extremes in molecular weight, low-abundance proteins and membrane-associated or -bound proteins (Washburn *et al.*, 2001).

Another recently introduced novel approach in 2D gel-based quantitative proteomics is the application of fluorescent cyanine dyes (Cy2, Cy3, Cy5) to label proteins before they are separated on a 2D gel (2D difference gel electrophoresis (DIGE) (Dirksen *et al.*, 2005). Protein samples are labelled with up to three spectrally distinct, charge and mass-matched fluorescent dyes known as CyDye DIGE fluors (Marouga *et al.*, 2005; Unlu *et al.*, 1997). These fluorescent labels carry a *N*-hydroxysuccinimidyl ester functionality, designed to modify the ε-amino group of lysine residues in proteins. The three spectrally resolvable fluorophores carry a positive charge to compensate for the charge of the lysine that is lost upon labelling, thereby balancing the pI of the protein. The molecular masses of the CyDyes are approximately 450 Da, and will not significantly affect protein migration in the second dimension. Taken together, the characteristics of these labels allow the analysis of up to three pools of protein samples simultaneously on a single 2D gel (Dirksen *et al.*, 2005). This approach eliminates technical, i.e. gel-to-gel, variation which is a main limitation of 2D gel electrophoresis. In a standard protocol, two of the dyes (typically Cy3 and Cy5) are used to label two different pools of protein samples, while the third label (Cy2) is used to label an internal standard that consists of equal amounts of the two pools. This internal standard allows a correction for further experimental errors, thereby distinguishing biological from experimental variation (Alban *et al.*, 2003). While DIGE offers this advantage over silver/Coomassie staining, it is comparatively very expensive and therefore not practical for routine experiments comprising large sample numbers.

An increasingly employed alternative approach to gel-based proteomic analysis is two dimensional liquid chromatography–mass spectrometry (2D LC-MS) (Washburn *et al.*, 2001). However, relative quantitation of proteins using LC-MS is challenging. Isotopic labelling of peptides enables simultaneous identification and quantification of peptides using tandem MS and permits parallel proteome analysis of two or more samples (Aggarwal *et al.*, 2006). A few methods exist for quantifying proteins; these include iCAT, SILAC, and iTRAQ. In iCAT the two samples to be compared are labelled either covalently or metabolically with different isotopes, proteolysed, mixed in equal quantities and analysed via mass spectrometry (Elrick *et al.*, 2006). In SILAC usually a single amino acid is used. If the supplied amino acid is Lys or Arg, analysis of peptides from a trypsin digest that cleaves after these two amino acids will result in peptides containing only a single difference from the labelled amino acid. Therefore, the mass difference between peptides in the MS scan will be known and consistent (Thelen, 2007). iTRAQ (Isobaric tags for Relative and Absolute Quantification) uses four amine-specific isobaric reagents to label the primary amines of peptides from four different biological samples. The labelled peptides from each sample are mixed, separated using two-dimensional LC and analysed using MS and tandem mass spectrometry (MS/MS) (Aggarwal *et al.*, 2006). Most recently, advances in the sensitivity and specificity of MS technology have facilitated the development of new software tools that enable MS-based label-free absolute protein quantitation using spectral counting (Braisted *et al.*, 2008).

In order to undertake the task of relative quantitation more comprehensively, additional tools are required. While new methods continue to be explored to address the inherent limitations of 2DGE, significant efforts have focused on developing alternative and complementary methods for resolving proteins from complex mixtures. In recent years, this

has led to the establishment of technologies based on multidimensional high performance liquid chromatography (HPLC) coupled to mass spectrometry (MudPIT), which now represent highly-reproducible alternatives to gel-based profiling. The application, however, of these technologies in malaria research remains limited and significant advances would be gained by developing and incorporating these alternative approaches into existing research efforts. This project therefore set out to optimise and employ new technologies in MudPIT towards expanding the scope of ongoing proteomics research on *P. falciparum.* Establishing this capability would thereafter enable high-throughput mass spectrometry-based proteomics to be performed, in order to determine the effects of the naturally derived compound Ref#31 on the *P. falciparum* proteome.

1.6 Aims

The aim of this project was to develop and employ new protocols in MudPIT using iTRAQ reagents to facilitate comparative profiling of the *P. falciparum* proteome. This work will contribute significantly to potentially providing a deeper understanding of the role proteins play in the life cycle and development of the parasite. The work will also contribute to elucidating the potential role these molecules play in mediating the efficacy of a novel plant-derived anti-malarial lead, Ref#31.

1.6.1 Objectives:

- The development of LC-MS/MS-based MudPIT methodologies to enable high-throughput profiling of the *P. falciparum* proteome.
- The morphological determination of the effect of Ref#31 on the *in vitro P. falciparum* intraerythrocytic developmental cycle.

- Proteomic profiling of *P. falciparum* on the parasite's intraerythrocytic developmental cycle following treatment with Ref#31.
- Determination of qualitative and quantitative changes in protein expression between Ref#31-treated and untreated *P. falciparum*.

CHAPTER 2: SEPARATION AND MASS SPECTROMETRY ANALYSIS OF *P. FALCIPARUM* PROTEINS USING SDS-PAGE AND ION EXCHANGE CHROMATOGRAPHY

ABSTRACT

iTRAQ technology for protein quantification using mass spectrometry is a recent, powerful means of determining relative protein levels in up to four samples simultaneously. The primary aim of this work was to develop and optimise an iTRAQ-based methodology for identifying and accurately quantifying proteins in *P. falciparum* that are differentially expressed following antimalarial drug exposure. Evaluation of changes at the level of protein expression may allow identification of novel proteins and the mechanisms involved. Separation and analysis of proteins was performed using the multidimensional separation techniques MudPIT and SDS-PAGE. Subsequent protein peptide detection was performed on a mass spectrometer (QSTAR Elite), and proteins identified and quantified using Protein Pilot 2.0 (Applied Biosystems, California, USA). Initially, a total of only 11 proteins from whole parasite lysate were identified at a high confidence level of 95% using SDS-PAGE as the first protein separation dimension and subsequent reverse-phase tandem mass spectrometry (RP-MS/MS), while 20 proteins were recovered when strong cation exchange (SCX) was used as the first dimension. Subsequently, a total of 158 *P. falciparum* proteins were identified at a high confidence level of 95% after a series of optimised steps were incorporated into the protocol. Significantly more proteins of empirical value were identified at lower confidence levels. This work has provided the foundation for establishing a functional MS laboratory for the high-throughput analysis and quantification of the *P. falciparum* proteome; however, more work remains to be done in terms of increasing the number of proteins identified with high confidence. It has also provided grounds to exploit a depletion strategy that is aimed at removing haemoglobin and leaving behind *Plasmodium* proteins. In-depth proteome studies of *P. falciparum* are severely hampered by the presence of haemoglobin, which accounts for more than 98% of the total erythrocyte soluble protein content (Ringrose *et al.*, 2008). A depletion approach

was developed using Ni-NTA (Nickel-Nitrilotriacetic acid) agarose, which resulted in an increase in the number of proteins subsequently identified through mass spectrometry.

2 INTRODUCTION

A newly developed technique termed iTRAQ (isobaric tags for relative and absolute quantification), allows two or more protein samples to be simultaneously analysed and quantified through multiplexing in a single experiment using MS technology (Washburn *et al.*, 2001). Each protein sample is proteolytically digested with trypsin, labelled with one of the isobaric tags, mixed with the other different iTRAQ-labelled samples, and analyzed by tandem mass spectrometry (MS/MS). Proteins are identified using the peptide product ion spectra and are quantified by the relative intensities of the 4-plex iTRAQ reporter ions (Song *et al.*, 2008). These reagents are non-polymeric, consisting of a reporter group, a balance group and a peptide reactive group (see Figure 1). The peptide reactive group covalently links an iTRAQ reagent isobaric tag with each lysine side chain and N-terminus group of a peptide, labelling all peptides in a given sample digest. Since all peptides are tagged, proteome coverage is expanded. The balance group ensures that an iTRAQ reagent-labelled peptide displays the same mass, whether labelled with iTRAQ reagent 114, 115, 116 or 117. During MS/MS, the isobaric tag cleaves at the sites indicated in Figure 2. As a result of fragmentation, there is neutral loss of the balance group, and the iTRAQ reagent reporter groups are generated, displaying diagnostic ions in the low-mass region between *m/z* of 114-117. These diagnostic ions are the key components that make these reagents superior for use in quantitative experiments (Applied Biosystems, 2004; Aggarwal *et al.*, 2006).

iTRAQ offers a number of significant advantages. Firstly, because all peptides are labelled it is possible to obtain data on more peptides as compared with other methods, thereby increasing confidence in the quantification of each protein (Unwin *et al.*, 2005). Secondly, because the tags are isobaric, the signal from up to four samples is summed in MS mode, providing an increase in sensitivity. Since four different iTRAQ reagents are available, comparative analysis of a set of two to four samples is feasible within a single MS run. For iTRAQ, tandem mass spectrometry (MS/MS), must be carried out on all peptides to obtain quantification, whereas for ICAT or SILAC MS/MS is carried out only on proteins that are differentially expressed (Unwin *et al.*, 2006).

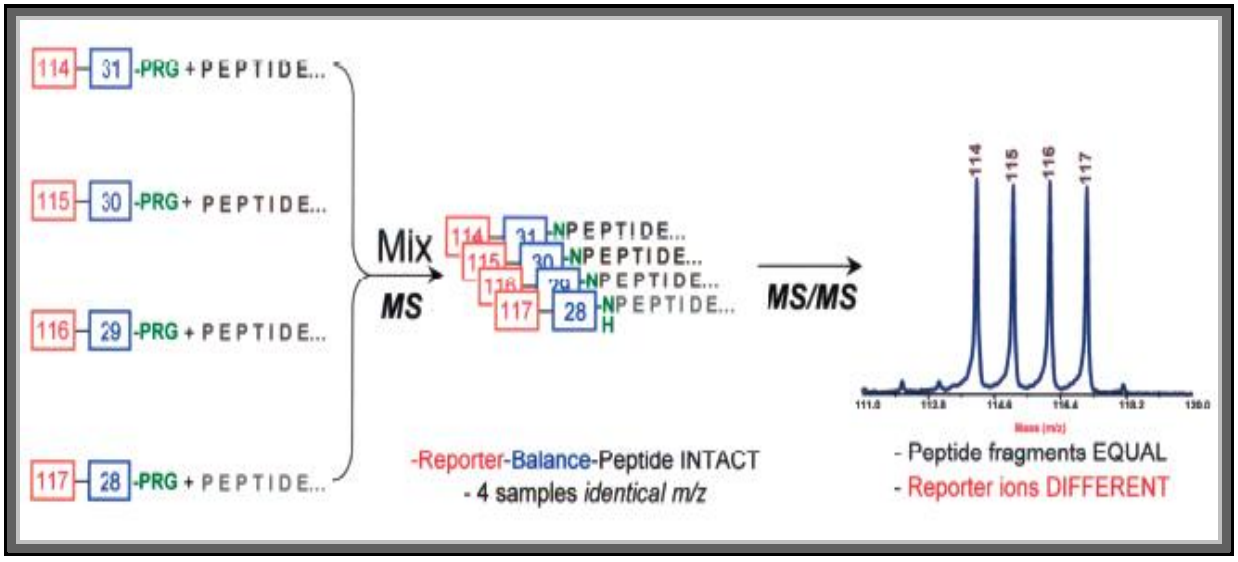


Figure 1: The concept of iTRAQ reagent chemistry (Gluckmann *et al.*, 2007)**.**

The figure shows the example of a 4-plex experiment. The tryptic peptides of each protein sample are labelled with one of the four iTRAQ reagents and then pooled prior to nano-LC separation and MS. While the labelled peptides are indistinguishable in MS mode, MS/MS analysis allows protein identification via fragmented peptides and quantitation via peak areas of the four signature ion signals.

The amine specificity of these reagents makes most peptides in a sample amenable to this labelling strategy with no loss of information from samples, such as post-translational modification. In addition, the multiplexing capacity of these reagents allows for information replication with certain LC-MS/MS experimental approaches, providing statistical validation within any given experiment (Zieske, 2006). Separation and analysis of the iTRAQ-tagged peptides is often performed using multidimensional protein identification technology (MudPiT) (Figure 2.1), combining strong cationic exchange (SCX) and reverse phase (RP) separation coupled to a tandem mass spectrometer. In this technique the peptides first bind to the SCX separation column. The peptides are then eluted from the first dimension using salt injections or salt gradients, which are compatible with a RP separation column. In the second dimension, hydrophobic and hydrophilic peptides carrying the same charge differ in their retention. Hydrophilic peptides are eluted first, followed by hydrophobic peptides. All peptides eluted from the nano-separation column are ionized using positive Electro Spray Ionization (ESI) and fragmented in MS/MS. The generated MS/MS spectra can be searched against the appropriate databases using proprietary software, e.g. Protein Pilot (Applied Biosystems) (Mitulovic *et al.*, 2004).

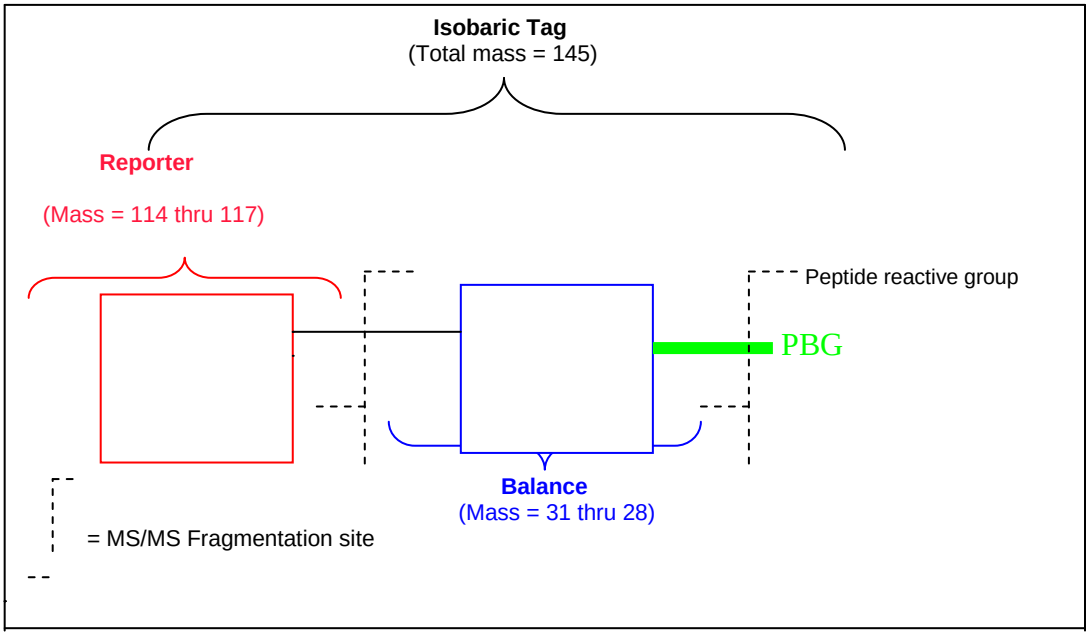


Figure 2: iTRAQ reagent structure showing the fragmentation sites generated during MS/MS (Applied Biosystems, 2004).

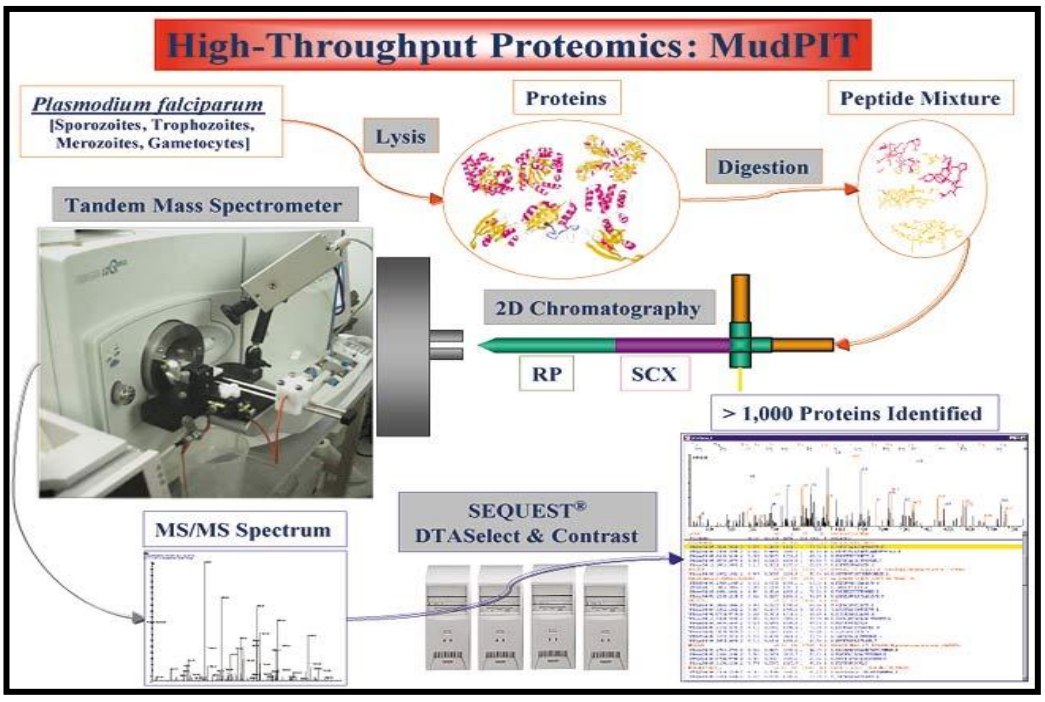


Figure 2.1: MudPIT (Multidimensional Protein Identification Technology). This technology is used for the separation and identification of complex protein and peptide mixtures. Rather than using traditional 2D gel electrophoresis, MudPIT separates peptides in 2D liquid chromatography. In this way, the separation can be interfaced directly with the ion source of a mass spectrometer.

The development of precise and transparent methods for analysis of proteomic data is one of the crucial challenges in protein sciences (Aebersold, *et al.,* 2003). Software for data evaluation support is needed for quantification, because proteomics yield huge amounts of data (Boehm *et al.,* 2004). These computer programs must be capable of providing results at the protein level. A number of software packages are currently available for analyzing iTRAQ data. These include the commercial packages MassTRAQ, Pro-Quant™ (Applied Biosystems (ABI), Darmstadt, Germany), ProteinPilot™ (ABI), Mascot™ 2.2 (Matrix Science, London, UK) (Boehm *et al.*, 2007) and the freely available package i-Tracker (Shadforth, *et al.,* 2005). For this study, iTRAQ analysis was performed using ProteinPilot™ (ABI), which is provided with the mass spectrometer used for peptide characterisation.

One of the major challenges in identifying proteins in a complex proteome such as *Plasmodium* is the presence of other proteins in wide dynamic concentration. In particular, high confidence detection of relatively low abundant proteins by MS is often impeded by the presence of highly abundant proteins. A highly promising step for most analysis is to deplete as many of the interfering proteins as possible. The ability to remove these proteins specifically, reproducibly, and with high selectivity is increasingly important in proteomic studies, and success will lead to an increasing list of lower abundant proteins being identified (Zolotarjova *et al.,* 2005). A range of methods to deplete high abundance proteins have been evaluated in the past, e.g. Cibacron blue (Echan *et al.,* 2005; Govorukhina *et al.,* 2003). While dye-based kits bind the majority of albumin, they also bind a large number of non-specific proteins, resulting in potential losses. Ideally, for biomarker discovery it is desirable to deplete as many high abundance proteins as possible, while minimizing incidental losses. The most critical consideration for major protein

depletion is the extent to which undesired non-specific losses of proteins occur during major protein depletion, and the potential positive impact that major protein depletion has on detection of lower abundance proteins (Echan *et al.*, 2005).

In order to increase the amount of *Plasmodium* proteins identified during MudPIT at a high confidence level (≥95%), a new depletion strategy that uses the tetradentate-chelating adsorbent Nickel-Nitrilotriacetic acid agarose (Qiagen) (Ni-NTA) was explored to remove human haemoglobin. Ni-NTA agarose is composed of Ni-NTA coupled to sepharose CL-6B and offers high binding capacity and minimal non-specific binding. This unique Ni-NTA was patented for its ability to bind 6xHis-tagged proteins. A Ni (II) binding site on the haemoglobin beta chain has previously been identified (Levine, 1998), and recently Ringrose and colleagues (2008) took advantage of this knowledge and explored whether immobilized Ni (II) is suited to efficiently deplete haemoglobin. The team successfully depleted haemoglobin and eluted the Ni (II) bound proteins with imidazole. The need to identify proteins expressed by the malaria parasite is often hindered by the presence of haemoglobin even after successive protein lysate washes, particularly when pre-fractionation of protein lysates containing haemoglobin is not possible due to unavailability of suitable equipment. Depletion of such proteins, followed by separation and mass spectral analysis can provide access to previously unreachable protein identities.

The primary aim of this work was to develop and optimise an iTRAQ-based methodology for identifying and accurately quantifying proteins in *P. falciparum* that are differentially expressed following antimalarial drug exposure. This work also led to the development of a method that employs Ni-NTA agarose to deplete haemoglobin from *Plasmodium* samples, leaving behind mostly *Plasmodium* proteins, in order to facilitate improved comparison,

through online LC-MS/MS, of differentially expressed proteins following Ref#31 treatment.

2.1 MATERIALS AND METHODS

2.1.1 Labelling of proteins with iTRAQ reagents for SDS-PAGE separation

Protein samples used for the optimisation of the iTRAQ-labelling protocol were obtained from a synchronized culture of *P. falciparum* 3D7 parasites grown for a period of five days according to the conditions detailed in Chapter 3, Section 3.1.3. Two 75-ml flasks contained cultures treated with DMSO (control) and two flasks contained cultures treated with Ref#31. Protein samples were extracted and pre-processed according to the methods outlined in Chapter 3, Section 3.1.4. The iTRAQ-labelling of protein samples was performed according to the recommendations set out in the manufacturer's kit (Applied Biosystems). Briefly, 6 volumes of acetone (at -20^{o}C) were added to 150 µl of protein sample (50 µg protein) (at 4^{o}C) in a 1-ml Eppendorf tube, and inverted three times. The reaction was incubated at -20^{o}C for 5 hrs until a precipitate formed, then centrifuged at 6000×g for 10 min. The acetone was decanted, and the precipitated samples dissolved in 20 µl of dissolution buffer (0.5 M triethylammonium bicarbonate, pH 8.5). A BSA protein assay was carried out for the determination of the protein concentration (detailed in Chapter 3, Section 3.1.5), and then 25 µg of the protein was transferred into a 1.5-ml microfuge tube. The sample was reduced by adding (0.5 M) 2 µl of 50 mM Tris-(2-carboxylethyl) phosphine (TCEP) for 1 hr at 37°C. The protein cysteine residues were blocked by adding 0.037 g/ml of iodoacetamide and the samples incubated at room temperature for 30 min. The sample was then inspected for the presence of precipitatation, and urea (250 µl) added to dissolve the precipitate. The sample must be completely dissolved before adding a tag to ensure complete labelling. Ethanol (70 µl of a 100% ethanol) was used to solubilise one

vial of iTRAQ reagent and this was then added to the sample. Each vial of iTRAQ reagent is sufficient to isotopically label up to 100 µg of protein. Cultures treated with Ref#31 were labelled with T114 tag and cultures treated with DMSO (control) were labelled with T115. The protein mixtures were labelled for 2 hrs at room temperature. Samples were then mixed and dried *in vacuo*. Following labelling, two different approaches were investigated to reduce the complexity of the *P. falciparum* proteome prior to peptide characterisation on the mass spectrometer. The first approach, detailed below in Section 2.1.2, involved protein lysate separation by conventional SDS-PAGE. The second approach, detailed in Section 2.1.3, involved peptide separation using online strong cationic exchange HPLC.

2.1.2 Separation of protein mixtures by gel electrophoresis (SDS-PAGE) and in-gel digestion of proteins

The proteins were analysed on an SDS-PAGE with a 4% stacking gel and a 15% separating gel. The stacking gel (0.5 M Tris-HCl, pH 6.8, 20% SDS) and separating gel (1.5 M Tris-HCl, pH 8.8, 20% SDS) were prepared from acrylamide (30% acrylamide, 0.8% bisacrylamide) and electrophoresis buffer (0.5 M Tris-HCl pH 8.8, 10% w/v SDS, 0.1 M glycine) stock solutions. The gel solutions were polymerised by adding 100 µl of 10% ammonium persulfate and 5 µl of TEMED *(N,N,N'N'*-tetramethylethylenediamine, Sigma Aldrich, USA). Protein samples were diluted 1:1 in reducing sample buffer (0.5 M Tris-HCl, pH 6.8, 10% w/v SDS, 0.1% v/v glycerol, 0.5% w/v bromophenol blue, 0.05% β-mercaptoethanol) and boiled at 70°C for 4 min. The electrophoresis was carried out in electrophoresis buffer using a BioRad electrophoresis system (BioRad, USA) with a voltage of 100 V for 35 min until the bromophenol blue marker reached the bottom of the gel. Protein molecular weight marker (PageBlue™, prestained protein ladder, Fermentas) was used.

Cultures treated with Ref#31 were labelled with T114 and cultures treated with DMSO (as the control) were labelled with T115. The combined mixtures of proteins labelled with T114 and T115 were reconstituted in SDS-PAGE sample buffer, and loaded and run in a single lane on a 15% gradient SDS-PAGE gel. After SDS-PAGE separation the proteins were visualized by protein page blue stain. The single lane was then divided into a total of 15 protein bands, and these were then excised using a scalpel and washed with 50% w/v NH_4HCO_3, by vortexing the pieces for 20 min. The gel pieces were subsequently dehydrated in 100% v/v ACN at room temperature until the gel pieces turned white. The pieces were then dried *in vacuo* for 10 min. A 0.25 µg/ml solution of trypsin (Whitehead Scientific) was added to the dehydrated gel pieces and incubated at 37°C overnight. The digested gel pieces were then vortexed and subsequently centrifuged (6000×g) at room temperature for 10 min to precipitate the solution to the bottom. The solution in the tubes was sonicated for 20 min on ice in a sonic water bath (4°C). The supernatant was then transferred from the tubes with gel pieces into a separate clean tube. The new tube was used to transfer the supernatant from the gel pieces to this corresponding tube. To the digested gel pieces remaining in each digestion tube, 100 µl of the extraction solvent was added (50% v/v ACN containing 0.1% v/v TFA in Milli-Q water). The solution inside the tubes was sonicated for 20 min (37°C) and the supernatant was transferred to a corresponding empty tube. This step was repeated twice. The supernatant was filtered using 0.22-µm filters (Agilent Technology). The supernatant tubes were then dried *in vacuo*. The extracted peptides were subsequently analysed by a QSTAR Elite LC/MS/MS system (Section 2.1.4).

2.1.3 Labelling of proteins with iTRAQ reagents for strong cationic exchange

The method for iTRAQ-labelling of proteins prior to separation using online strong cationic exchange HPLC was similar to that outlined in Section 2.1.2, with the following modifications. Acetone-precipitated samples were dissolved in 26 µl (instead of 20 µl) of dissolution buffer (0.5 M triethylammonium bicarbonate). The pH of the buffer was determined with a pH indicator strip to be ~8.5. A BSA protein assay was carried out for the determination of the protein concentration (Chapter 3, Section 3.1.5), and then an increased protein concentration of 50 µg (instead of 25 µg) protein was aliquoted into 1.5-ml microfuge tubes. The samples were reduced by adding 2 µl of 50 mM TCEP for 1 hr at 37°C. The protein cysteine residues were blocked by adding 1 µl (0.037 g/ml) of iodoacetamide (200 mM iodoacetamide in isopropanol) and the samples were incubated at room temperature for 10 min (instead of 30 min). Digestion of proteins was performed overnight at 37°C by addition of 10 µg/ml of trypsin (Whitehead Scientific) dissolved in 10 µl of Milli-Q water. Ethanol (70 µl) was used to solubilise the iTRAQ reagent which was then added to the sample. The protein mixtures were isotopically labelled for 1 hr at room temperature by adding the content of one vial of iTRAQ reagent (suited for 100 µg protein). Each sample was filtered using a 0.22-µm filter (Agilent Technologies). The samples were then mixed at equal ratios and dried by centrifugal evaporation.

2.1.4 Sample optimisation

2.1.4.1 Solid Phase Extraction column (SPE) sample clean-up

Two sets of spin columns were investigated to reduce the complexity of the *P. falciparum* proteome prior to peptide characterisation on the mass spectrometer. Both columns used the same principle, they are used primarily to clean up samples for analysis and/or

concentrate samples to improve detection. Both columns (C8 Octyl and C18 spin column, Agilent Technologies) use the following methodology. The columns were tapped lightly to allow resin to settle at the bottom. The top and the bottom cap were removed from the column, and the column placed into a 1.5-ml Eppendorf receiver tube. The resin was activated by adding 200 µl of activation solution (50% methanol) to rinse the walls of the spin column and wet the resin. The column was then centrifuged at 1500×g for 1 min. The flow-through was discarded and 200 µl of equilibration solution (0.5%TFA/5%ACN) added to the resin. The column was then centrifuged at 1500×g for 1 min and the flow-through discarded. The sample (170 µl) was loaded on top of the resin bed and the column placed into a 1.5-ml Eppendorf receiver tube. The column was then centrifuged at 1500×g for 1 min. The flow-through was retained to confirm sample binding, and the column was washed by adding 200 µl of wash solution (0.5%TFA/5%ACN) and centrifuged at 1500×g for 1 min. A new 1.5-ml Eppendorf receiver tube was placed under the column, and 20 µl of elution buffer (70% ACN) added to the resin bed and centrifuged at 1500×g for 1 min. The flow-through was gently dried in a centrifugal evaporator and resuspended in 2% ACN/0.1% FA for MS analysis. The iTRAQ-labelled and unlabelled samples were resuspended in 0.5% TFA/2%ACN and then mixed in equal ratios to run in the MS as one sample.

2.1.5 LC/MS/MS

2.1.5.1 One-Dimensional Liquid Chromatography (RP-HPLC)

Protein identification was performed using a QSTAR Elite LC/MS/MS (Applied Biosystems), interfaced with an Agilent 1100 micro-pump and an autosampler. Prior to RP nano-LC MS/MS analysis, the samples were resuspended in 100 µl of 2% Acetonitrile/0.1% Formic acid (2% ACN/0.1% FA) (pH 2.9). Tryptic peptides were

separated by RP-HPLC on a C18 Zorbax 300SB nanocapillary column (150 mm × 0.075, 3.5 µm, Agilent 1100 series). A solvent system consisting of solvent A (0.5% ACN and 0.5% FA) and solvent B (95% ACN and 0.1% FA) was used. Peptides were eluted into the mass spectrometer using a gradient of 12 to 35% solvent B (95% ACN and 0.1% FA) in 10 min, 35 to 100% solution B in 50 min. The mass spectrometric parameters were as follows: ion spray voltage (IS) 2300V; curtain gas (CUR) 30; gas 1, 35; gas 2, 0; declustering potential (DP), 60V; focusing potential (FP), 265V; declustering potential (DP2), 15V. Isolation of peptide ions of interest was performed in the quadrupole Q1 set to low resolution.

2.1.5.2 Two-Dimensional Liquid Chromatography (SCX-RP-HPLC)

Mass spectrometric experiments were performed using a QSTAR Elite LC/MS/MS (Applied Biosystems), interfaced with an Agilent 1100 micro-pump and an autosampler. iTRAQ-labelled peptides were fractionated by strong cationic exchange HPLC (SCX) using a Zorbax SCX Bioseries II column: 50 × 0.8 mm (Agilent Technologies). Buffer A was 2% ACN/0.1% FA, pH 2.9 and Buffer B was 5% ACN/0.05% FA, pH 2.9. The dried iTRAQ-labelled sample was resuspended in 50 µl of 0.5% ACN/0.5% FA. The sample containing 50 µg of protein was injected on a SCX column, which was tandem with a reverse phase column (RP), through a six-port valve. The sample was then separated using SCX in the first dimension, followed by reverse phase RP in the second dimension. The flow rate was 15 µl/min. The peptides bound onto the SCX column were then eluted with step-wise salt injections of 0-1 M KCl, totalling 14 salt injections, resulting in a total run time of 152 min. The peptides were then captured onto a 100 × 50 mm trap column (3.5 µm C18, Zorbax 300 SB). Buffers for RP HPLC column were A (0.5% ACN and 0.5% FA) and B (95% ACN and 0.1% FA). Typical acetonitrile gradient elution for the second

dimension was: 0-30% ACN in 90 min, 30-50% ACN in 20 min, 50-90% ACN in 1 min. The LC eluent was then subjected to positive ion nano-flow electrospray MS/MS analysis in an information dependent acquisition mode (IDA). In IDA mode, a TOF-MS survey scan was acquired (*m/z* 400-1500, 60 s), with the three most intense multiple charged ions (counts=10) in the survey scan sequentially subjected to product ion analysis. Product ion scan was accumulated for 2 s in the mass range 100-2000 with a modified enhance all mode Q2 transition setting favouring low mass ions so that the reporting iTRAQ ion (114, 115, 116, 117 *m/z*) intensities were enhanced for quantification. The mass spectrometric parameters were as follows: ion spray voltage (IS) 3000V; curtain gas (CUR) 30; gas 1, 35; gas 2, 0; declustering potential (DP), 60V; focusing potential (FP), 265V; declustering potential (DP2), 15V. Isolation of peptide ions of interest was performed in the quadrupole Q1 set to low resolution.

2.1.5.3 Valve Switching Technique (SCX-RP-HPLC)

In an attempt to increase the total number of peptides eluted from SCX prior to RP-HPLC, peptide elution was also investigated using a buffer containing KH_2PO_4 as an alternative to formic acid. The choice of KH_2PO_4 allows the establishment of a fast pH equilibrium on the SCX column, and provides a stable pH environment throughout the separation (Mitulovic *et al.*, 2004). The method for the (SCX-RP-HPLC) valve switching technique was similar to that outlined in Section 2.1.4.2, with the following modifications. The sample was resuspended in 2.5 mM KH_2PO_4 / 2% ACN, pH 2.5 and the SCX column equilibrated in the same solution (2.5 mM KH_2PO_4 / 2% ACN, pH 2.5). Sample load time was 30 min with valve switching after 20 min, allowing a 10-min trap wash by the nano-pump to remove KH_2PO_4 before peptide elution. When phosphate buffer was used, an additional wash step of the RP trap column was performed using 0.1% FA. After sample

loading the trap was washed again to remove salts that remained on the column and in the tubing to prevent them from entering the ion source of the mass spectrometer. The sample was then transferred onto the nano-separation RP column.

2.1.5.4 MS data analysis

Data analysis for the iTRAQ experiments was performed with Protein Pilot 2.0 (Applied Biosystems). The cut-off for the confidence settings for all analyses was at 95%. Database searching for each sample was done on predicted tryptic peptide sequence data using the MSDB database (Mass Spectrometry protein sequence database). The MSDB database is a non-identical protein sequence database maintained by the Proteomics Department at the Hammersmith Campus of Imperial College London (ftp://ftp.ncbi.nih.gov/repository/MSDB/msdb.nam). This database contains protein sequences derived from organisms whose genomes have been sequenced. Due to the identical homology often observed between highly conserved sequences, peptide sequence hits from *P. falciparum* were occasionally incorrectly ascribed to other species of *Plasmodium.* This was also observed for a number of proteins originating from human erythrocytes that were incorrectly ascribed to other mammalian species due to identical peptide sequences. As a consequence, such hits had to be manually verified in detail. The MSDB database was used initially because we did not have access to the PlasmoDB database, which is a database that is used to specifically search for *Plasmodium* proteins.

2.2 RESULTS

2.2.1 Protein separation by SDS-PAGE

SDS-PAGE was used as a first dimension to separate proteins according to their apparent molecular weights prior to LC/MS/MS analysis. Protein separation was performed to facilitate increased coverage of the number of proteins identified by subsequent MS analysis, and to improve the accuracy (confidence) of these identifications. The sample used for iTRAQ-labelling optimisation was obtained from a synchronised culture of *P. falciparum*, extracted at randomly selected timepoints during the parasite's intraerythrocytic development cycle. The sample was reduced, blocked and then labelled with iTRAQ reagents prior to tryptic digestion (described previously in Sections 2.1.1 and 2.1.2) Figure 3 illustrates the separation of the iTRAQ-labelled versus unlabelled samples by SDS-PAGE.

Visual analysis of the unlabelled and the iTRAQ-labelled samples showed no evidence of the binding of the tag. In 2007, Wiese conducted a similar experiment using myoglobin and BSA, demonstrating that iTRAQ-labelled proteins migrated slower than their native counterparts during SDS-PAGE due to the additional mass of the bound tag. In this present study; however, both samples (proteins) migrated in a similar pattern (Figure 3). Further analysis, as detailed below in Section 2.2.2, showed that the peptide reactive group failed to covalently link the iTRAQ reagent tag to lysine side-chains and the N-terminus group of the sample peptides. As a result, an alternative labelling method was investigated (as described in Section 2.1.3), the results of which are detailed in section 2.2.3.

In the meantime, in order to determine the number and type of proteins that could be identified through MS analysis, lanes II and III in Figure 3 were each divided into 15 equal

bands, and these bands were excised from the gel as described previously (Section 2.1.2). All 30 bands were subjected to in-gel digestion, and the resulting peptides were subsequently combined into a single sample for analysis by LC/MS/MS as discussed below.

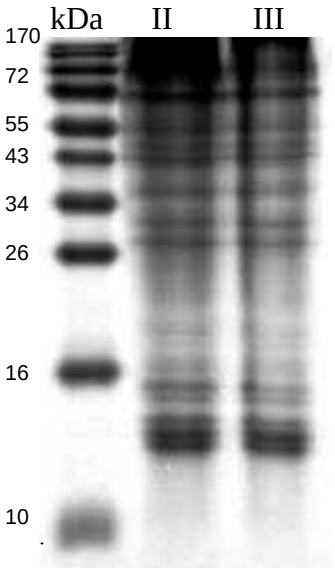


Figure 3: SDS-PAGE of iTRAQ-labelled samples. I: MW Marker, II (Unlabelled, control), III iTRAQ (labelled). Represented in Figure 3 are Lane I, molecular marker (ranging from 10-170kDa), Lane II, unlabelled (control) and Lane III, iTRAQ-labelled samples (combined T 114- and T 115-tagged samples). Samples represent parasite total protein extracts labelled with iTRAQ.

2.2.2 One-Dimensional Liquid Chromatography (RP-HPLC)

Digested peptides were analysed on the QSTAR Elite running Protein Pilot 2.0 software. Initial analysis using complex protein digests extracted from SDS-PAGE bands (Section 2.1.1) and separated by 1-dimensional Reverse Phase (i.e. 1-D RP-HPLC) chromatography revealed that this high-throughput approach resulted in relatively poor chromatographic separation. Only 11 proteins were identified from the combined samples with the majority being highly abundant *H. sapiens* haemoglobin. None of the proteins identified belonged to *P. falciparum* (see Table 1). Although several proteins listed in Table 1 do not appear to belong to either P. *falciparum* or *H. sapiens* but to other species, closer inspection revealed that identical homology exists between the underlying peptide sequences identified for

these species (e.g. Orangutan) and that of *H. sapiens*. Thus, the identified proteins should have been ascribed to *H. sapiens* by the automated software, rather than to other related species. In such cases manual curation of the correct peptide-species match had to be performed.

Due to the low number of proteins identified with very high confidence (95% or greater), combining the labelled and unlabelled samples, and running this combined sample through LC/MS/MS was deemed to be a suboptimal approach to profiling the *P. falciparum* proteome, particularly since the presence of highly abundant proteins appeared to mask the detection of the more low abundant *Plasmodium* proteins. Thus, an alternative approach was therefore investigated whereby the excised bands were run individually in order to yield more protein identifications (detailed in Section 2.2.3).

The MS/MS data generated for the 11 identified proteins confirmed that the iTRAQ reagent tags 114 and 115 did not bind to the lysine and the N-terminal residues of peptides. Thus, in the MS spectra of the iTRAQ-labelled and unlabelled sample shown in Figure 4 (human haemoglobin, beta-chain), no evidence could be seen for the presence of the reporter ions in the corresponding 114-115 *m/z* region. In addition to requiring the presence of lysine and N-terminal residues, the technology is also significantly influenced, for example, by sample pH where extremes of pH will lead to poor or non-binding of tags. Similarly the presence of residual substances from protein sample preparation, in particular primary amines that react with iTRAQ reagents, may also impair labelling. Presence of urea, DTT as well as amine could hinder the binding of iTRAQ tags. Given this result, and in order to reduce sample handling and chemical processing prior to MS analysis, an on-line SCX HPLC separation method was subsequently explored as an alternative to reducing proteome

complexity, in an attempt to improve protein yield, increase binding capacity of iTRAQ tags, and to improve the identification of *Plasmodium* proteins.

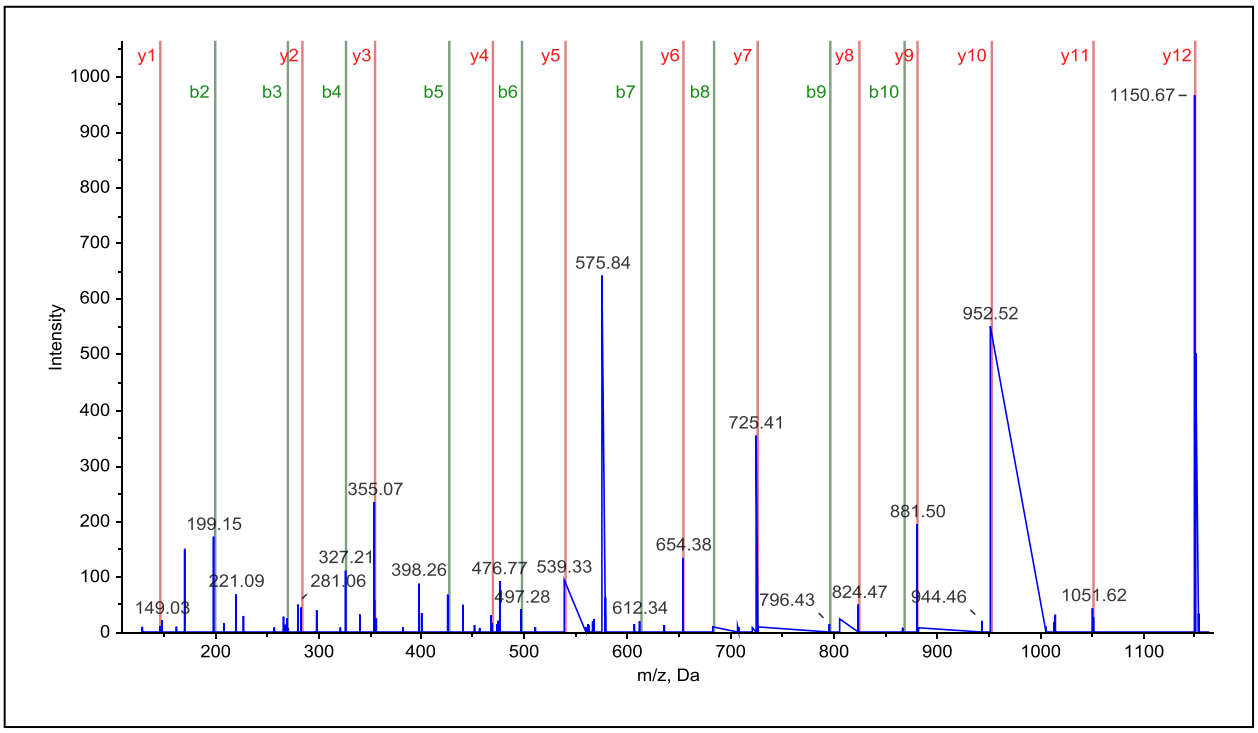


Figure 4: Nano-ESI-MS/MS spectra of the proteolytic peptide FFESFGDLSTPDAVMGNPK (Haemoglobin beta chain, *H.sapiens*) generated by in-gel tryptic digestion of iTRAQ-labelled peptides. The spectra show no evidence of the reporter ions being present, which would be expected to appear at the 114-117 *m/z* region of the MS spectrum.

Table 1: Summary of proteins identified following separation by SDS-PAGE and 1D-RP-HPLC/MS/MS

Unused	N	Total	% Cov	Accession #	Name	Species
55.54	1	55.54	93.9	Q549N7|Q549N7_HUMAN	Mutant beta-globin – *H. sapiens* (Human)	HUMAN
39.91	2	39.91	79.6	Q5R9M5|Q5R9M5_PONPY	Putative uncharacterized protein DKFZp468J1717 - *Pongo pygmaeus* (Orangutan)	ORANGOTANG
15.2	3	36.87	96.6	A0N071|A0N071_HUMAN	Delta-globin chain – *H. sapiens* (Human)	HUMAN
12.02	4	12.02	18.4	A7ZWZ1|A7ZWZ1_ECOHS	Beta-galactosidase - *Escherichia coli* O9:H4 (strain HS)	Escherichia coli
5.61	5	5.61	14.9	A5A6M6|A5A6M6_PANTR	Keratin 1 - *Pan troglodytes verus*	PANTER
4.1	6	8.55	73.2	A8DUV3|A8DUV3_MOUSE	Alpha-globin - *Mus musculus* (Mouse)	MOUSE
3.7	7	3.7	7.6	A8K2I0|A8K2I0_HUMAN	cDNA FLJ78504, highly similar to H. sapiens keratin 6A – *H. sapiens* (Human)	HUMAN
2.52	8	2.52	19.1	Q9Z1R9|Q9Z1R9_MOUSE	Trypsinogen 16 - *Mus musculus* (Mouse)	MOUSE
2.28	9	38.97	89.1	Q6R7N2|Q6R7N2_HUMAN	Hemoglobin beta – *H. sapiens* (Human)	HUMAN
2.01	10	4.16	42.2	Q9GLX7|Q9GLX7_SAISC	Hybrid gamma1/gamma2-globin - *Saimiri sciureus* (Common squirrel monkey)	MONKEY
2	11	2	21.9	Q9D7Y7|Q9D7Y7_MOUSE	Adult male stomach cDNA, RIKEN full-length enriched library, clone:2210010C04TRYPSIN III, CATIONIC - (Mouse)	MOUSE

Key: Unused – A measure of the protein confidence for a detected protein; **N** – The rank of the specified protein relative to all other proteins in the list of detected proteins; **Total** – A measure of the total amount of evidence for a detected protein calculated using all of the peptides detected for the protein; **% Cov** – The percentage of matching amino acids from identified peptides having confidence greater than or equal to 95%, divided by the total number of amino acids in the sequence; **Accession number** – MSDB accession number assigned to protein; **Name** – MSDB name assigned to protein; **Species** – Species origin of protein based on peptide sequences identified

2.2.3 Two-Dimensional Liquid Chromatography (SCX-RP-HPLC)

A 50-µg *P. falciparum* culture extracted randomly at different time points was isolated and then digested in solution to run on the SCX-RP-HPLC. The sample was not labelled with iTRAQ. The incorporation of the SCX column as the first separation dimension into the nano-HPLC system improved the MS performance and increased the amount of relevant *Plasmodium* data obtained. In the approach, the peptides are trapped on a SCX column, and are then eluted into the RP column using salt injections. An acetonitrile gradient was used to elute peptides from the RP column onto the mass spectrometer. Peptides were eluted in a 5-35% ACN window as described previously (see Section 2.1.5.2).

In terms of protein yield, a total of 20 proteins were identified with high confidence (95% or greater), 15 belonged to *Plasmodium* and the rest were *H. sapiens* or *Toxoplasma gondii* (*T. gondii*) related (Table 2). On closer inspection, the peptides identified as belonging to the *T. gondii* heat shock protein 70 (UniProtKB accession number AAC72001.1) were found to share identical homology to those belonging to *P. falciparum*. Thus, this protein most likely belonged to *P. falciparum*. Interestingly, BLAST alignment between these two species of the full heat shock protein 70 sequence revealed high overall amino acid identity (Identities = 498/609 (81%), Positives = 555/609 (91%), Gaps = 0/609 (0%), E (expect) value 0.0). The non-*Plasmodium* proteins that were identified were the result of using the MSDB database through which to perform the peptide searches. The database is not specific for *Plasmodium*, but rather is a composite non-identical protein sequence database built from a number of primary source databases such as TREMPL, SWISSPOT and GENBANK (Rohila *et al.*, 2009). Thus, in order to limit the proteins that are identified through MS to just those specific to *Plasmodium*, it would be necessary to obtain a *Plasmodium* protein sequence database, and to perform the peptide sequence searches

against this database. This optimised strategy was pursued using the *Plasmodium*-specific database PlasmoDB 5.5, as detailed in Chapter 3.

Table 2: Summary of proteins identified following separation by 2-D SXC-RP HLPC MS/MS

N	Unused	Total	% Cov	Accession #	Name	Species
1	15.79	15.79	20.2	Q4KKW9_HUMAN	Solute carrier family 4, anion exchanger, member 1(Erythrocyte membrane protein band 3, Diego blood group)- *H. sapiens* (Human).	*H. sapiens*
2	8.07	8.07	35.2	pfa3D7predicted|chr13|chr13.phat_570	coding) 100% identity to 100% of EAA15302.1: translation elongation factor EF-1, subunit alpha	***Plasmodium yoelii yoelii***
3	7.67	7.67	5.6	SPTB1_HUMAN	Spectrin beta chain, erythrocyte (Beta-I spectrin).- *H. sapiens* (Human).	*H.sapiens*
4	6.3	6.3	22.6	pfa3D7|Plasmodium_falciparum	glyceraldehyde-3-phosphate dehydrogenase, putative	***Plasmodium berghei***
5	6	6	11.6	Q8T6B1_PLAFA	Glyceraldehyde-3-phosphate dehydrogenase.-*Plasmodium falciparum*.	***Plasmodium falciparum***
6	4.32	4.32	13.5	Q8I0P6_PLAF7	Elongation factor 1-alpha.- *Plasmodium falciparum* (isolate 3D7).	***Plasmodium falciparum***
7	4.18	4.18	4.6	Q8IB24_PLAF7	Heat shock 70 kDa protein.- *Plasmodium falciparum* (isolate 3D7).	***Plasmodium falciparum***
8	3.64	3.64	3.2	BAD92655	AB209418 NID: - *H. sapiens*	*H. sapiens*
9	3.22	3.22	25.2	Q7JSX6_PLAFA	Histone H4 (Fragment).- *Plasmodium falciparum*.	***Plasmodium falciparum***
10	2.85	2.85	11.7	pfa3D7predicted|chr13|chr13.phat_270	coding) 100% identity to 100% of 1T26.A:Chain A, *Plasmodium falciparum* Lactate Dehydrogenase	**270|FullPhat|*Plasmodium_falciparum*_Sanger|**
11	2.74	2.74	71.8	pfa3D7predicted|chr11|chr11.phat_78	coding) 100% identity to 100% of AAC46612.1: histone H2B	**78|FullPhat|*Plasmodium_falciparum*_TIGR**
12	2.66	2.66	16.2	Q8IIV1_PLAF7	Histone H2B.- *Plasmodium falciparum* (isolate 3D7).	***Plasmodium falciparum***
13	2.57	2.57	14.9	pfa3D7predicted|chr8|chr8.phat_205	coding) 81% identity to 90% of AAC72001.1:heat shock protein 70 [*Toxoplasma gondii*]	*Toxoplasma gondii*
14	2.39	2.39	11.7	Q71T02_PLAFA	L-lactate dehydrogenase (EC 1.1.1.27).- *Plasmodium falciparum*.	***Plasmodium falciparum***
15	2.24	2.24	17.5	pfa3D7predicted|chr11|chr11.phat_77	coding) 96% identity to 100% of Q8T7J8:Histone 4 Location= chr11 complement	**77|FullPhat|*Plasmodium_falciparum*_TIGR**

16	2.01	2.01	2	SPTA1_HUMAN	Spectrin alpha chain, erythrocyte (Erythroid alpha-spectrin).- *H. sapiens* (Human).	*H. sapiens*
17	2	2	29.8	pfa3D7predicted|chr8|chr8.phat_157	coding) 84% identity to 90% of CAH94808.1:conserved hypothetical protein	***Plasmodium berghei***
18	2	2	5.3	pfa3D7predicted|chr6|chr6.phat_90	coding) 100% identity to 100% of Q07805:Ornithine aminotransferase	**90|FullPhat|*Plasmodium_falciparum*_Sanger**
19	1.52	1.52	51.2	pfa3D7predicted|chr7|chr7.phat_166	coding) 100% identity to 100% of CAH98717.1: histone h2b, putative	***Plasmodium berghei***
20	1.4	1.4	25.9	pfa3D7predicted|chr9|chr9.phat_173	coding) 100% identity to 100% of AAA93010.1:PBGRP	**173|FullPhat|*Plasmodium_falciparum*_Sanger|**

Key: **Unused** – A measure of the protein confidence for a detected protein; a good protein confident is 95% and above. **N** – The rank of the specified protein relative to all other proteins in the list of detected proteins; **Total** – A measure of the total amount of evidence for a detected protein calculated using all of the peptides detected for the protein; **% Cov** – The percentage of matching amino acids from identified peptides having confidence greater than or equal to 95%, divided by the total number of amino acids in the sequence; **Accession number** – MSDB accession number assigned to protein; **Name** – MSDB name assigned to protein; **Species** – Species origin of protein based on peptide sequences identified

Figure 5 overleaf are the results of an attempt to optimise the iTRAQ protocol. The iTRAQ tagging protocol was seen to work, as the peptides were covalently linked to isobaric tag 115. The inset shows a zoom into the *m/z* region of the MS/MS spectrum where reporter ions appear.

Further experimentation was conducted to improve the protein yield prior to LC-MS/MS, as well as to optimise the parameters used for MS/MS analysis. This time, no iTRAQ tag was used because this work was done to specifically to try increase the number of proteins identified by LC-MS/MS, where similar studies have been able to identify up to 780 proteins using a similar approach (Aggarwal *et al.*, 2006). It was, however, acknowledged that significant differences between research groups in factors such as the sensitivity of the MS equipment used, the proprietary algorithms used by the MS to identify protein peptide sequences, and the minimum cut-off confidence level used to define a successful protein hit, all contribute to the final number of proteins identified. Table 2 contains the results of the re-optimised in-solution digest coupled to 2-D chromatography. It shows protein yields obtained when the gradient was increased from 55 min to 111 min, resulting in a total run of 152 min. An improvement in the number of proteins identified was observed, with a total of 158 proteins being identified. Of these, 24 were unique *Plasmodium* proteins, 80 were deemed to be non-*Plasmodium* related proteins (proteins from other species), 17 were related to haemoglobin and 33 were additional human proteins (Figure 6). Thus, from an initial 11 then 20 proteins, a total of 158 proteins were now identified. Spectrin and haemoglobin were amongst the highest scores for total percentage protein covered by the positively identified peptide sequences. Even with the use of a SCX column to fractionate proteins, a large amount of contaminating proteins (primarily isoforms of human

haemoglobin) was still being identified. Due to this, a haemoglobin removal strategy was explored, and the results of this approach are detailed in Section 2.2.5 below.

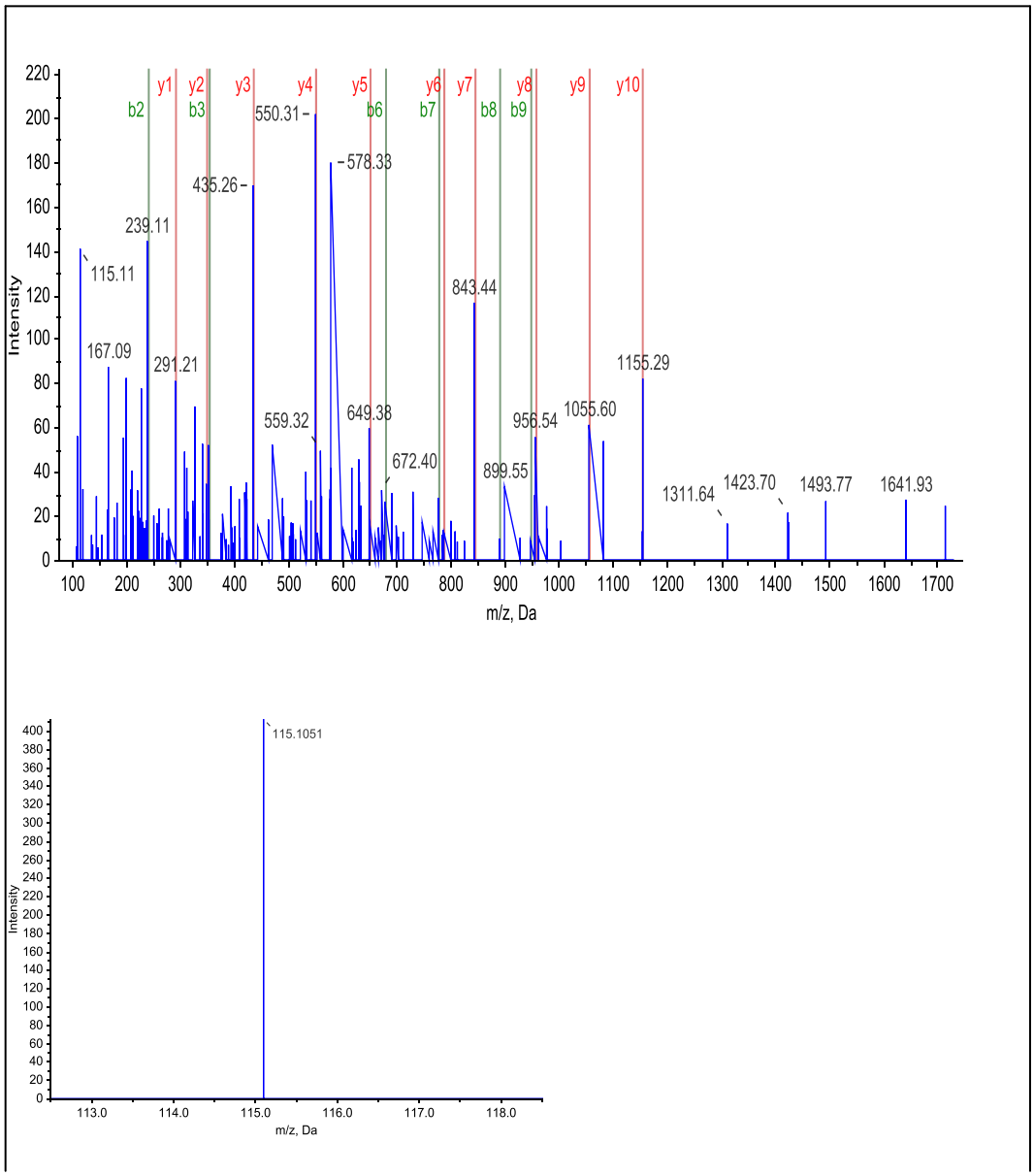


Figure 5: Nano-ESI-MS/MS spectra of the proteolytic peptide THINLVVIGHVDSGK$^{(iTRAQ)}$ (elongation factor 1 alpha, *P. falciparum*) generated by in-solution digestion of iTRAQ-labelled peptides. The inset shows a zoom into the *m/z* region of the MS/MS spectrum where reporter ions appear.

A second important point from these results was that a large number of proteins identified by MS/MS appeared to belong to other species. However, as before, and on closer inspection of the individual peptides, it was clear that the majority of these in fact belonged to *Plasmodium*. However, due to the high sequence identity between conserved peptides belonging to *Plasmodium* and those of other species, the use of the default database MSDB meant that successfully identified proteins were wrongly ascribed to other species. The MSDB database therefore proved not to be ideal for *Plasmodium* searches, providing further support for the use of a *Plasmodium* specific protein database.

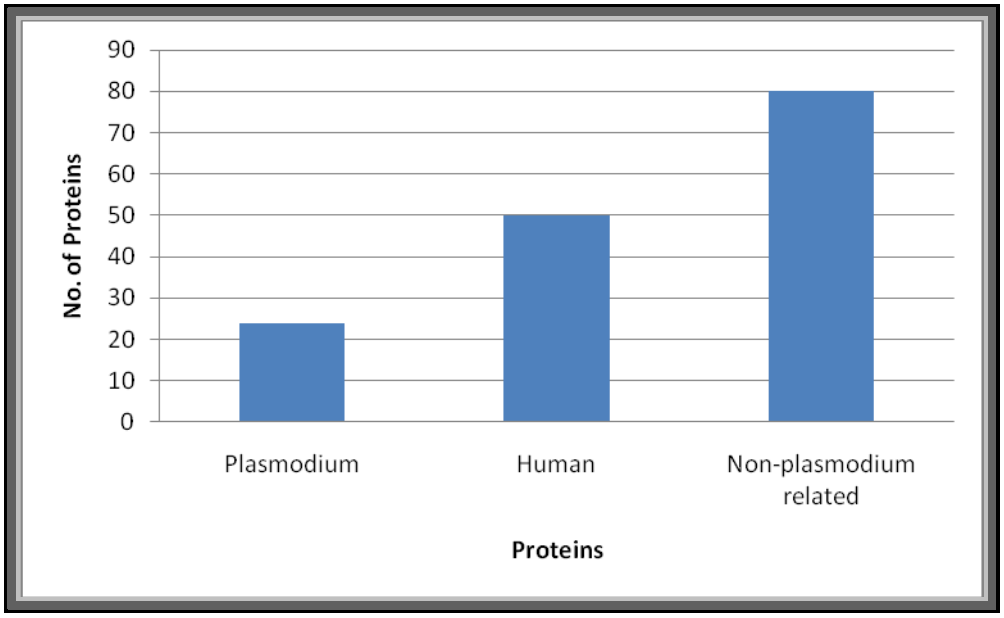


Figure 6: An illustration of the number of proteins identified by 2-D SCX-RP chromatography coupled to MS/MS.

2.2.4 Depletion of haemoglobin to enhance detection of *Plasmodium* proteins

The main constituent of the erythrocyte soluble protein fraction is haemoglobin (Hb), as illustrated by a very abundant band representing the Hb monomer at approximately 16 in Figure 7 (lane I). The effectiveness of depletion using Ni-NTA superflow is shown in Figure 7. The majority of the Hb was depleted; however, non-targeted bands also appear in the bound fraction (Lane III). Hundred (100) µg of protein was depleted using the Ni-NTA superflow. After depletion the concentration was measured again. The concentration was measured to get 50 µg proteins. The proteins bound on the superflow were eluted using 100 mM imidazole. These proteins were also quantified to obtain 50 µg of protein to run on the gel. Due to time constraints, these bands were not analysed on the RP-HPLC. Therefore, more work still needs to be done in order to know exactly the nature of these bands.

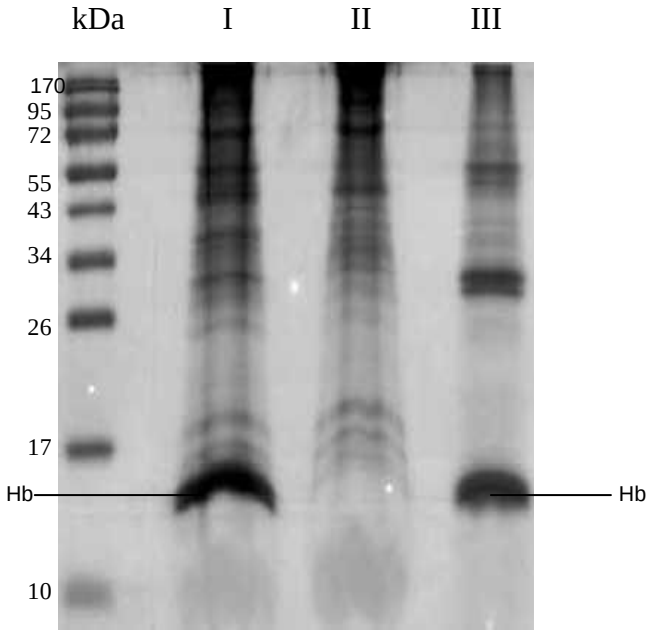


Figure 7: Protein page-stained SDS-PAGE analysis of the depletion of Hb (haemoglobin) from the *Plasmodium* protein fraction. Lane I, non-depleted protein sample containing 50 µg of total protein lysate; Lane II, Hb-depleted fraction, Lane III, Ni (II) bound fraction, eluted with 100 mM imidazole.

2.3 DISCUSSION

This work set out to establish a method for fractionating, and identifying through mass spectrometry, proteins isolated from *P. falciparum* cultures in a relatively high-throughput manner. Coupled with iTRAQ protein labelling, establishment of this multidimensional protein identification technology (MudPIT) would enable the approach to be used to profile the effects of drugs and other perturbations on the parasite's proteome. Two main approaches were investigated to achieve reliable fractionation of the parasite's proteome prior to MS analysis, namely gel-based and HPLC-based separation. The separation of *P. falciparum* proteins by SDS-PAGE initially resulted in relatively low protein coverage. At the same time, no protein labelling was observed using iTRAQ due to inefficient binding of the iTRAQ tag. A comparison between the unlabelled and the labelled sample (Section 2.1.1) did not show a characteristic shift in molecular weight of the proteins (Figure 3), which is used as an initial indicator of successful protein labelling. Further analysis of the same samples by LC/MS/MS exhibited no evidence of the tag being present in the expected *m/z* region (114-117) of the MS/MS ion spectra, confirming that the protein peptides had not been labelled. This lack of labelling could potentially be attributed to a number of reasons. iTRAQ is an amine and pH-dependent technology, where extremes of pH may result in non- or inefficient binding. It is therefore critical to check the pH of all solutions and solubilisation of the sample after acetone precipitation. However, the pH of the samples used in this study was 8.5, which is the specified pH required to facilitate optimal iTRAQ tag labelling. It is therefore unlikely that the pH played a role in the non-binding of the tags. Another possible reason for the observed lack of protein labelling could have been the presence of primary amines in chemicals used for protein sample preparation, which are able to react with the iTRAQ reagents, thereby competing with the peptide derivitization reaction. The amount of water present in the solution also plays a big role in terms of

binding, where too much water interferes with the binding capacity of the tags. Specifically, iTRAQ reagent hydrolysis competes with the amine-labelling reaction, therefore the efficiency of the iTRAQ reagent labelling is highly dependent on the environment reagents and the aqueous environment. Lastly, chemicals such as urea and DTT are some of the reagents that are also know to potentially interfere with the iTRAQ labelling reaction. However, since the protein samples investigated in this study were subjected to acetone precipitation to remove such reagents, it is unlikely that these could have played a significant role.

A total of 11 proteins were identified with very high confidence using SDS-PAGE as the first dimension. None of the proteins identified belonged to *Plasmodium* (Table 1). A possible explanation for the low protein yield and the absence of *Plasmodium* proteins is due to the fact that all the excised bands were combined and run as one sample on the MS/MS. Given that the primary point of SDS-PAGE is to fractionate the proteins into peptides and run them separately on the MS/MS, combining the protein extracts was therefore not the best approach to use. At the same time, the results from the SDS-PAGE approach also highlighted the fact that the low number of proteins that were identified may also have been due to the very high abundance of haemoglobin identified in the samples. This is despite removing large amounts of this protein through repetitive wash steps during protein processing. The highly abundant haemoglobin proteins could have masked the presence of other proteins excised from SDS-PAGE, resulting in the lack of identification of relatively low abundance proteins derived from *Plasmodium* (Echan *et al.*, 2005). The presence of several non-*Plasmodium* proteins in the list of proteins identified by MS, particularly haemoglobin, suggests that the saponin lysis extraction protocol and the washing steps used to remove most of the haemoglobin might not have been optimal,

particularly given the high sensitivity of MS to detect the presence of relatively highly abundant proteins. Saponins are widely used to make the plasma membrane of cells permeable, allowing small molecules and macromolecules to enter or leave the cytoplasm. When erythrocytes are incubated with saponin in isotonic buffer, the membrane becomes permeable to hemoglobin (Baumann *et al.*, 2000), therefore in this study saponin was used to lyse RBC leaving behind the *Plasmodium* proteins. Since an abundance of haemoglobin was observed in the proteins identified using SCX-RP-HPLC–MS/MS, it is possible that saponin treatment did not lyse the cells properly.

Another factor that could have influenced the low number of proteins initially identified is the extraction protocol used to extract proteins from SDS-PAGE. Although the protocol used was according to the manufacturer's recommendations, it is possible that this may also have not been optimal for the samples used in this study, resulting in low protein yield. As a result, further work was performed, as outlined in Chapter 3, to optimise the SDS-PAGE extraction protocol for *Plasmodium* proteins. SDS-PAGE has been used as a preferred method of separation for proteins, but in this study it has proven to be less efficient when separating complex protein mixtures. In 1994, Hellman and colleagues compared the in-gel and in-solution digestion of proteins, and found that the yield of peptides after proteolysis in the gel was more than that obtained by digestion in solution. In contrast, in-solution digest of samples in this present study resulted in an increase in the observed total number (20 as opposed to 11 by SDS-PAGE) of proteins identified, of which 15 were *Plasmodium* and 5 were proteins from other species (Table 2). In addition, efficient iTRAQ tag binding was observed using this latter approach, as confirmed by the differentially labelled peptides in the MS scan (Figure 5). The incorporation of an SCX-

HPLC dimension to the standard RP-HPLC MS/MS therefore proved to be more efficient in protein fractionation compared to SDS-PAGE.

Prior to further extensive experimentation to increase the number of *Plasmodium* proteins identified through MS/MS, and as part of the optimisation process, a separate experiment was conducted using a standards kit comprising a six-protein mix (Applied Biosystems) as a control to validate the protein processing protocols and operation of the MS/MS. All six proteins present in the mixture were successfully identified with high confidence (see Appendix 1, Supplementary CD, Table 2) highlighting that, while the system was correctly setup to identify relatively less complex protein mixtures, the complexity of the *Plasmodium* proteome necessitated more optimisation.

Further experimentation was therefore carried out in extensive detail to try to increase protein yield and to optimise the RP-HPLC MS/MS. These included the use of a bigger RP column (C18), longer gradient, increasing the concentration of the starting material, the use of phosphate as a loading buffer, and the use of C8-cartridge and C18 spin clean-up columns. A total of 18 proteins were obtained with a bigger RP column (0.1 × 50 mm C18 column versus the previous 0.075 × 150 mm C18 column), of which 10 were *Plasmodium* hits, while 6 were haemoglobin and 2 were spectrin proteins of human origin. In an RP phase, peptides are eluted according to hydrophobicity. Hydrophobic and hydrophilic peptides carrying the same charge will differ in their retention on the RP column. Hydrophilic peptides will be eluted first and hydrophobic ones later (Mitulovic *et al.*, 2004). A bigger RP column was used because the initial smaller column resulted in more peptides going to waste and not being analysed due to column overloading. The use of a longer C18 gradient (111 mins versus 52 mins) was also seen to improve protein yield by

providing sufficient binding capacity and ample time for peptide binding onto the column. A total of 74 proteins were identified, 18 were *Plasmodium* and 56 were haemoglobin- and spectrin-related. In contrast, the subsequent use of clean-up columns prior to employing the above approach was seen to reduce the total number of proteins detected to 35. Another clean-up method that would be worth exploring, but was not conducted in this study due to time constraints, is the BioRad clean-up kit, which is suggested to work better for complex protein mixtures than acetone precipitation (BioRad). In an attempt to increase the number of proteins identified, the concentration of the starting material was increased from 50 µg to 70 µg. This, however, resulted in more haemoglobin hits being produced than *Plasmodium* hits.

The use of phosphate as a loading buffer was not practical for our HPLC set-up since our LC/MS/MS only uses a 6-port valve. The use of phosphate required the column switching technique (Mutulovic *et al*., 2004), which ideally uses a 10-port valve. Loading buffer is one of the most important parameters used for SCX column separation; it should contain a low elution potential for the RP trap column to ensure that peptides will bind a low concentration of organic solvents such as ACN (Mutulovic *et al*., 2004). The use of phosphate resulted in 51 identified *Plasmodium* proteins; however, the problem with phosphate was its compatibility with online ESI-MS. The use of phosphate as a loading buffer should therefore be considered for future offline SCX runs. In-solution digest coupled to 2-D chromatography was further demonstrated to be a better option in terms of protein coverage. A total of 158 proteins were identified (Figure 6: Table 1, Appendix 1 Supplementary CD). The remaining proteins were ascribed to other species. On closer inspection it was again clear that although some of these other proteins were ascribed to, for example, *Toxoplasma gondii*, the peptide sequences identified for this species were

identically homologous to those of *P. falciparum*. This discrepancy was probably again due to the use of the conventional MSDB database against which the peptide sequence searches were performed, which contains peptide sequences from the genomes of all species sequenced to date. In order to correct such descrepancies, it would be necessary to manually BLAST the peptide sets representing each protein identified for the additional species against the *P. falciparum* genome to confirm that the peptides/protein belongs to this parasite. In the case of the haemoglobin isoforms ascribed to non-human species such as the Brown lemur for example, it would be necessary to perform this BLAST procedure to confirm that the haemoglobin isoform(s) were in fact human. Given the difficulty and time required to manually BLAST large numbers of peptide sequences, an alternative approach was investigated using a *Plasmodium* specific protein peptide database (see Chapter 3).

An important method which was not investigated in this study, due to a lack of the prerequisite hardware, is the incorporation of a protein fraction collector prior to MS as a way of improving fractionation of the *Plasmodium* proteome, and increasing the amount and quality of data obtained. In a mass spectrometer fitted with a protein fraction collector, fractions are typically collected into a 96-well plate and then each fraction is loaded onto a RP column automatically via an autosampler and analysed by LC/MS/MS. The advantage of a protein fraction collector is that more peptide fractions can be collected and a decision made as to which fractions are to be analysed. Interesting fractions can also be revisited for analysis if needed (Peng *et al.*, 2003).

In all these attempts to get maximum protein coverage, human haemoglobin and human spectrin were still the dominating proteins. Spectrin is a contractile protein of high

molecular weight that is a component of a network in the membrane of red blood cells, giving the cells flexibility. Haemoglobin is an important nutrient source for intraerythrocytic malaria parasites and a major contaminant in malaria cultures. The presence of haemoglobin has previously been seen to hinder the analysis of the *Plasmodium* proteome (Nirmalan *et al.,* 2004). Haemoglobin accounts for approximately 97% of the erythrocyte proteome (Ringrose *et al.,* 2008).

A strategy that can deplete most of the haemoglobin from *Plasmodium* cultures while leaving behind relatively low abundance *Plasmodium*-specific proteins is greatly required. The basis of this work allowed me through collaboration to explore the development of a possible depletion strategy aimed at removing most of the haemoglobin, leaving mostly *Plasmodium*-specific proteins behind. To enable this depletion, a strategy incorporating Ni-NTA agarose was used. As outlined in the results (see Figure 7), Ni-NTA was indeed able to bind to haemoglobin, and it was subsequently possible to significantly reduce the presence of this highly abundant interfering protein in the total *Plasmodium* protein lysate, as shown in Lane II of Figure 7. It was also clear from Lane III that other proteins besides haemoglobin exhibit affinity for Ni (II). Due to time constraints, it was unfortunately not possible to determine what these proteins were and what proportion of these belonged to *Plasmodium*. The results did, however, demonstrate what could be a very useful approach for depleting haemoglobin from *Plasmodium* cultures, particularly if a high-throughput online SCX/RP/MS/MS approach is employed to profile this parasite.

This work has provided the initial effort to set up a functional MS laboratory for the high-throughput analysis and quantification of the *P. falciparum* proteome; however, more work remains to be done in terms of increasing the number of proteins identified with high

confidence (95% or greater). Incorporation of a protein fraction collector is greatly required to further decrease proteome complexity. Replacing the C18 column with a hydrophilic column is another future possibility. The C18 column used for this study works on hydrophobicity; hydrophobic peptides bind more strongly to the column and hydrophilic peptides are more likely to go to waste while the hydrophobic ones are retained. Thus, a hydrophilic column might be better suited to analysing the *P. falciparum* proteome, whose 80% A+T genome content is well known to limit the use of conventional biochemical approaches to analyse this organism (LaCount *et al.*, 2005).

In traditional proteomics, including studies of the malaria parasite, analysis is frequently performed with mass spectrometers functioning in positive-ion mode. Analysis of this type allows the detection of peptides that are rich in negatively-charged amino acids residues, or post-translationally modified with phosphate. Another approach for future purposes is the use of negative ions. Thus, a combination of negative and positive ion data collected from the same sample should significantly improve sequence coverage and database search scores (Sims and Hyde, 2006).

CHAPTER 3: INVESTIGATING THE MORPHOLOGICAL EFFECT OF A NOVEL ANTIMALARIAL COMPOUND REF#31 ON THE INTRAERYTHROCYTIC DEVELOPMENT CYCLE OF *PLASMODIUM FALCIPARUM*, AND IDENTIFICATION OF PROTEINS DIFFERENTIALLY EXPRESSED BY *PLASMODIUM FALCIPARUM* AFTER TREATMENT WITH REF#31

ABSTRACT

Plasmodium falciparum (*P. falciparum*) has proven itself to be extremely resilient and highly resistant to any form of immunization. This realization has intensified efforts to identify new drugs and new drug targets that can be exploited to combat the parasite. *Dicoma anomala* (*D. anomala*) ssp. *gerardii*, a highly variable indigenous plant characterized by its stiff, sharp-pointed bracts, has previously been shown to possess some antimalarial activity *in vitro*. Exploration of this plant and its antimalarial properties could therefore offer new opportunities for the discovery and development of new compounds against malaria. Currently, no knowledge exists regarding the molecular effects of *D. anomala* on *P. falciparum*, nor how these effects inhibit the growth of this parasite. In order to better understand the possible mode of action of an antimalarial candidate derived from this plant (Reference #31 (Ref#31)), high-throughput proteomics approaches developed earlier were employed. To guide the use of this approach, an initial morphology study was undertaken to examine the effect of Ref#31 on the 48-hr intraerythrocytic cycle of *P. falciparum*. The changes to *P. falciparum* morphology in synchronised cultures treated with Ref#31 were observed in thin blood smears fixed and stained with Giemsa and examined by light microscopy. Cultures treated in the mid-ring stage were sampled every 6 hr up to 48 hr from three flasks; one was a control culture, one was treated with an inhibitory drug concentration of IC_{50}, and one with an inhibitory drug concentration of IC_{99}. Within 18 hrs post-treatment (hpt), the control cultures already showed signs of development to the mid trophozoite stage, whereas Ref#31-treated cultures exhibited the late ring to early trophozoite stage of parasite growth. Thirty hours into the study control cultures were predominantly in the early schizont stage, whereas treated cultures appeared to be arrested in the early trophozoite stage of growth. Ref# 31 therefore exhibited a quick onset of action and high efficacy against the blood stages of *P. falciparum.*

Based on the results of the morphology study, three different time points were selected, i.e. 6 hrs, 12 hrs and 18 hrs, to enable proteome analysis of *P. falciparum* following Ref#31 treatment. Using ProteinPilot software, a total of 91 proteins were identified with high confidence through protein peptide searches for the 6-hr control sample, of which 80 were *Plasmodium* proteins and 11 were non-*Plasmodium* related proteins, while 150 proteins were recovered for the 6-hr treated sample, of which 128 were *Plasmodium* and 22 were non-*Plasmodium* related proteins. A total of 119 proteins were identified for the 12-hr control sample and a total of 126 proteins for the 12-hr treated sample (Appendix 2, Supplementary CD Tables 1-6). A total of 226 proteins were recovered for the 18-hr control sample, of which 205 were *Plasmodium* proteins and 22 were non-*Plasmodium* related proteins, while 127 proteins were recovered for the 18-hr treated sample, of which 115 were *Plasmodium* and 12 were non-*Plasmodium* related proteins. Using a recently developed label-free quantitation approach, based on mass spectrometer spectral counts, it was possible to further investigate these proteins to determine those differentially expressed after treatment. In addition, the use of MS/MS also enabled simultaneous characterisation of post-translation modification (PTM) events that occurred in the *Plasmodium* proteome following treatment. Interestingly, several protein modifications were identified that have recently been implicated to play a role in the cell growth inhibition properties of other compounds belonging to this drug class. Using existing bioinformatics tools, interpretation of these post-translational events proved, however, to be very tedious. Given their widespread nature and potential biological importance, development of high-throughput bioinformatics tools are thus required to enable accurate meaningful interpretation of PTM events identified during high-throughput mass spectrometry analysis. At the same time, new bioinformatics tools are needed to enable the integration of such data with that from

other molecular technologies, such as microarrays. In this way, a more comprehensive understanding of the effects of antimalarial drugs such as Ref#31 could be achieved**.**

3 INTRODUCTION

Humans have suffered greatly through the centuries because of parasites. Parasites are organisms that derive nourishment and protection from other living organisms, known as hosts. They may be transmitted from animals to humans, from humans to humans, or from humans to animals (Gerald *et al.*, 1996). *P. falciparum* is a protozoan parasite, one of the species of *Plasmodium* that cause malaria in humans. Most protozoa are too small to be seen with the naked eye, most are around 10–50 μm, but can easily be found under a microscope. The *Plasmodium* parasite was first seen by Laveran on November 6 1880 at a military hospital in Constantine, Algeria, when he discovered a microgametocyte exflagellating. He then called the organism after himself, Laverania falcipa, which later became known as *Plasmodium falciparum* (*P. falciparum*) (Cox, 2002; Garnham, 1966). Today, *P. falciparum* is placed in the kingdom Protista, phylum Apicomplexa, class Sporozoea, subclass Coccidia, order Ercoccidiida, suborder Haemosporina, genus *Plasmodium* and species *falciparum* (Karapelou, 1987).

P. falciparum has proven itself to be extremely resilient and highly resistant to any form of immunization. Resistance of malaria parasites has risen from several factors, including overuse of antimalarial drugs for prophylaxis, inadequate or incomplete therapeutic treatments of active infections, a high level of parasite adaptability at the genetic and metabolic levels, and a massive proliferation rate that permits selected populations to emerge relatively rapidly (Hyde, 2007). By means of an elaborate and complex life cycle, this protist has managed to wade all attempts at eradication and continues to thrive and

threaten the lives of millions of people each year (Schaechter, 1993). Although a number of advances have been made towards the understanding of the disease, relatively few antimalarial drugs have been developed in the last 30 years (Clarkson *et al.,* 2004; Ridley, 2002). The range of current antimalarials is narrow. These include those based on quinine (chloroquine, mefloquine, amodiaquine and halofantrine) or other aminoquinolines (primaquine and tafenoquine), the antifolate compounds (pyramethamine, proguanil, cyocloguanil, dapsome and sulfadoxine), the artemisinin derivatives (artesunate, artemether/arteether, Co-artem and others), and, most recently, the hydroxynapthoquinone atovaquone (Baniecki *et al.,* 2007).

Since the treatment and control of malaria depends largely on a limited number of chemoprophylactic and chemotherapeutic agents, there is an urgent need to develop novel, affordable antimalarial treatments. This urgency has been further highlighted by the increasing prevalence of drug-resistant strains of *P. falciparum*, which have contributed to the escalating disease burden (Clarkson *et al.,* 2004).

Historically, the majority of antimalarial drugs have been derived from medicinal plants or from structures modelled on plant lead compounds. These include the quinoline-based antimalarials, as well as artemisinin and its derivatives. Medicinal plants are commonly used in South African traditional healthcare to treat a range of ailments, including malaria and its associated symptoms (Clarkson *et al.,* 2004; Watt and Breyer-Brandwijk, 1962). The recognition and validation of traditional medical practices and the search for plant-derived drugs could lead to new strategies in malaria control. It is necessary to obtain more scientific information concerning the efficacy and safety of the remedies in use, because many people in third world countries already use and depend on herbal medicines for the

treatment of malaria (Prozesky *et al.*, 2001). South Africa boasts remarkable biodiversity and rich cultural traditions of plant use. Scientific understanding of medicinal plants is, however, largely unexplored and pharmacological investigation of the South African flora only gained momentum recently (Van Wyk, 2002).

The genus *Dicoma* has approximately 35 species distributed over tropical South Africa with one species in Asia. It is placed in the tribe Mutisieae, subtribe Gochnatiinae. Most species are herbs or small shrubs, but even trees are known. The genus has been divided into eight sections, two of them with heterogam and six with homogam heads. So far the chemistry of seven species does not show a clear division; most common are germacranolides (Zdero *et al.*, 1990). *Dicoma anomala* (*D. anomala*) is an indigenous plant growing in sub-Saharan Africa. Prickly, mauve and green, it grows flat along the ground. It is often found on rocky hillsides and open grass veld, preferring full sun and well-drained soil, and it propagates only from seed. In South Africa, it is widely distributed in Limpopo, the North-West, Gauteng, Mpumalanga, the Free State, the Northern Cape and KwaZulu-Natal. As a medicinal plant, it is famous and much-respected, and is used for a wide variety of ailments by several South African tribes. The Xhosas and the Sotho use the dried powdered stem, leaves and especially the root for treating sores, scabs, wounds and ringworms. It is used by South African tribes for coughs and colds, to clear blood disorders, as a tonic and a heart herb (Roberts, 1990). Compounds isolated from this plant include four germacranolides, namely 14-acetoxydicomanolide, 14-oxodicomanolides, eudesmanolide and guaianolide (Hutchings *et al.*, 1996; Zdero *et al.*, 1990). It is widely utilized by people, who should be informed against over-utilization since this plant has demonstrated to have all types of activities including antimalarial activities (Mnengwane *et al.*, 2007).

The aim of the current study was to establish the morphological effects of the novel plant-derived antimalarial lead Ref#31, on the intraerythrocytic development cycle of *P. falciparum*. This information was then used to determine quantitative and qualitative protein changes in *P. falciparum* following Ref#31 treatment, in an attempt to better understand the molecular effects of this compound on the parasite.

3.1 MATERIALS AND METHODS

3.1.1 Drugs

Ref#31 was obtained from the Bioprospecting Research Platform (CSIR, courtesy of Dr. Vinesh Maharaj). The compound is a single chemical entity obtained from *D. anomala*, and has been demonstrated to possess significant antimalarial activity *in vitro*. Two drug concentrations were investigated, based on previously determined IC_{50} and IC_{99} values for this compound (Dr. Linda Mtwisha, CSIR Systems Biology, personal communication). Stock solutions of Ref#31 were dissolved in dimethyl sulphoxide (DMSO). The final concentrations used were 0.42 µg/ml for IC_{50} and 3.0 µg/ml for IC_{99}.

3.1.2 Morphological study of the drug (Ref#31)

The 3D7 strain of *P. falciparum* was cultured in 25 × 50-ml flasks, containing total culture volumes of 10 ml. A total of three flasks were used. Two flasks were treated with Ref#31 but with different concentrations IC_{50} (1.75 µM) and IC_{99} (12.5 µM), and the third flask was used as the untreated control. The parasites were allowed to invade new red blood cells (RBC) and then treated in the mid-ring phase (8 hrs post-invasion). The cultures were maintained in culture flasks at 37ºC with a gaseous phase of 90% N_2, 5% O_2, and 5% CO_2 (Trager and Jensen, 1976). Parasites were cultured in human erythrocytes (blood group A+)

in RPMI 1640 medium; Sigma, supplemented with 25 mM HEPES, 20 mM sodium bicarbonate and 40 µg/ml gentamicin. Albumax II (0.5%; Invitrogen) replaced serum in the culture medium. The culture was maintained between 2-5% parasitemia. To establish synchrony, each culture was centrifuged for 5 min at 2000×g (room temperature), the supernatant discarded and the pellet resuspended in 6 ml of aqueous 15% D-Sorbitol for 5 min at 37°C. An additional 12 ml of D-Glucose was added and the culture incubated at 37°C for 5 min. Sorbitol synchronization is based on the increased permeability of infected erythrocytes to the sugar alcohol. In an isotonic sorbitol solution, mature trophozoite-infected erythrocytes import sorbitol via the PSAC (Plasmodial Surface Anion Channel) and/or other parasite-induced channels; continued uptake produces an osmotic imbalance which leads to cell lysis. Immature ring stage forms survive this treatment because they do not exhibit increased sorbitol permeability (Hill *et al.*, 2007).

Drug morphology assays were performed on synchronized parasites by treating 10 ml of *P. falciparum* cultures in the mid-ring phase of the intraerythrocytic life cycle with a final concentration of 3.0 µg/ml Ref#31 (in DMSO) for IC_{99} and 0.42 µg/ml for IC_{50}. The same volume of DMSO was added to cultures representing control treatments. Thin smears were prepared from the cultures every 6 hrs over a 48-hr period. Morphological forms and parasitemia were detected and estimated by Giemsa-stained smears and light microscopy magnified at 1000X.

3.1.3 Culturing of parasites and drug treatment (Ref#31)

The 3D7 strain of *P. falciparum* was cultured in 75 × 250-ml flasks, containing total culture volumes of 30 ml for drug treatments and subsequent protein extractions. A total of six flasks were used of which three flasks contained cultures treated with DMSO (control) and

three flasks contained cultures treated with Ref#31 (IC_{99}). This concentration was selected based on the morphology results, where no viable forms of the parasite were seen after 48 hrs at this drug concentration compared to IC_{50}. Parasites were cultured using the conditions previously described in Section 3.1.2. The parasites were transferred into Falcon tubes and then centrifuged to precipitate the pellet at the bottom. After centrifugation for 5 min at 2000×g (room temperature), the supernatant was removed and the pellet resuspended in 30 ml of culture medium. For each replicate sample, the culture was then transferred to a new culture flask and fresh erythrocytes added. Seven millilitres of culture was harvested in triplicate at the late ring stage (6 hrs post-treatment (hpt)), the early trophozoite stage (12 hpt), and the mid/late trophozoite stage (18 hpt). Similar volumes were also harvested in triplicate from the control DMSO-treated culture at the same time points. All drug treatment experiments were performed in technical duplicate.

3.1.4 Protein extraction

Total protein was extracted from harvested cultures using a method developed by Nirmalan and colleagues (2004). Parasites were released from erythrocytes by saponin lysis (10% saponin in PBS) for 10 min on ice and thereafter washed five times in ice-cold wash medium (RPMI 1640 medium (Sigma) supplemented with 25 mM HEPES, 20 mM sodium bicarbonate and 40 mg/ml gentamicin. Incubation of cells and tissues with saponin makes the lipid bilayer permeable to macromolecules. Saponin-lysed erythrocytes do not reseal, thus indicating an irreversible damage of the lipid bilayer (Baumann *et al.*, 2000).

Saponin pellets containing parasites were resuspended in 200 ml of lysis buffer (8 M Urea, 2 M thiourea, 2% CHAPS). The sample was freeze-thawed five times in liquid nitrogen and then sonicated using five 10-sec bursts on ice, interspersed with 10 sec of cooling

(power at 65%) (Sonicator 4000, ThermoFischer Scientific). The supernatant was clarified by ultracentrifugation for 30 min at 30 000×g at 4°C. Care was required with sonification to prevent warming of the sample, which introduces artifacts resulting from urea breakdown and protein carbamylation into the sample.

3.1.5 Determination of protein concentration

Protein concentration was measured using a RC-DC quantification kit from BioRad, following the manufacturer's procedure, using bovine serum albumin (BSA) as the standard. The RC-DC Assay can be used for samples containing both reducing agents as well as detergents, which are the major interfering agents in protein estimation. Hence, this assay is ideal for protein samples for 2-dimensional electrophoresis (2DE) (BioRad).

The RC-DC Assay is based on the principle of Lowry estimation, where protein reacts with alkaline copper and subsequently reduces the folin reagent and leads to colour development. This colour development is due to changes to tyrosine and tryptophan and, to a lesser extent, cysteine and histidine. The absorbance of the final product can be measured at 750 nm.

3.1.6 Depletion of haemoglobin to enhance protein identification

Typically, 1 ml of Ni-NTA superflow in ethanol (Qiagen) was diluted with 1 ml of distilled water. This was then centrifuged for 1 min in a desktop centrifuge (14 000×g at room temperature) to remove the ethanol. This step was repeated twice to ensure proper removal of ethanol. After centrifuging, an equal amount of water was added to the superflow to make a 100% solution (50% superflow + 50% water). A total of 100 µg/µl protein (170 µl) was added to an equal amount of superflow (protocol courtesy of Dr. T. Tsekoa). The mixture was vortexed gently and incubated on ice for 30 min. The mixture was then centrifuged for 2 min in a desktop centrifuge (14 000×g at room temperature). The

supernatant was collected in a clean tube for further analysis. The remaining superflow was washed twice with distilled water to remove any excess proteins that did not bind to Ni-NTA. Hundred (100) mM imidazole was passed over a 1-ml Ni-NTA superflow. The mixture was mixed and incubated on ice for 30 min. It was then centrifuged for 2 min in a desktop centrifuge (14 000×g at room temperature). The protein-containing supernatant was transferred into a clean test tube and the pellet discarded.

3.1.7 Separation of protein mixtures by gel electrophoresis (SDS-PAGE) and modified in-gel digestion

The decision to employ SDS-PAGE for sample fractionation, rather than SCX-HPLC-based fractionation, was governed by the unavailability of a replacement specialised proteomics SCX column. Following sample haemoglobin depletion using Ni-NTA, as described above, 6 volumes of acetone (at -20^{o}C) were added to 150 µl of protein sample (at 4^{o}C) in a 1-ml Eppendorf, and the tube inverted three times. The tube was incubated at -20^{o}C for 5 hrs until a precipitate formed, then spun at 6000×g for 10 min. The acetone was decanted and the precipitated samples dissolved in 50 µl of dissolution buffer (0.5 M triethylammonium bicarbonate (0.0197 g/ml). A BSA protein assay was carried out for the determination of the protein concentration (detailed in Section 3.1.5), and then 50 µg of the protein was aliquoted into 1.5-ml microfuge tubes. The samples were mixed with SDS-PAGE sample buffer (1:1), and loaded and run in a single lane on a 15% gradient SDS-PAGE. Protein molecular weight marker (PageBlue^{TM}, prestained protein ladder, Fermentas) was used. After SDS-PAGE separation, the proteins were visualized by protein page staining. A total of 19 gel bands were excised and washed with 40% w/v NH_4HCO_3, and the gel pieces were then dehydrated for 15 min in 100% v/v ACN. The gel bands were excised at random equally interspersed lengths because the gels did not show any clear

visible bands, therefore there was no specific place on the gel to indicate where the bands should be excised. The gel pieces were incubated at room temperature until they turned white. The pieces were then dried in a speed vacuum concentrator for 10 min. A solution of dithiothreitol (50 mM dithiothreitol in 25 mM NH_4HCO_3) was added to the dried gel pieces and incubated for 1 hr at 60°C. After 1 hr, the supernatant was discarded. Iodoacetamide (55 mM iodoadetamide in 25mM NH_4HCO_3) was then added to the gel pieces, incubated in the dark for 1 hr at room temperature and the supernatant was then discarded. The gel pieces were then dehydrated for 15 min in 100% acetonitrile, this step was repeated twice. The pieces were then dried in a speed vacuum concentrator for 10 min. The gel pieces were placed on ice and 20 µl of trypsin was added. A vial of trypsin (20 µg, Whitehead Scientific) was reconstituted with 200 µl of trypsin buffer, and 180 µl of 25 mM NH_4HCO_3 was added to 20 µl trypsin. A total of 20 µl (0.025 µg/µl) trypsin solution was added to the dehydrated gel pieces and incubated on ice for 1 hr. The supernatant was removed and 20 µl of 25 mM NH_4HCO_3 was added to the gel pieces, followed by incubation overnight at 37°C. The digested gel pieces were then vortexed and centrifuged. The tubes were sonicated, using the conditions described in Section 3.1.4, for 10 min and the supernatant collected and dried down for further analysis.

3.1.8 One-Dimensional Liquid Chromatography (RP-HPLC)

Protein identification was performed using a QSTAR Elite LC/MS/MS (Applied Biosystems), interfaced with an Agilent 1100 micro-pump and an autosampler. Prior to RP nano-LC MS/MS analysis, the samples were resuspended in 100 µl of 2% ACN/0.1% FA (pH 2.9). Tryptic peptides were separated by RP-HPLC on a C18 Zorbax 300SB nanocapillary column (0.1 × 50 mm, 3.5 µm, Agilent 1100 series). A solvent system consisting of solvent A (0.5% ACN and 0.5% formic acid) and solvent B (95% ACN and

0.1% formic acid) was used. Peptides were eluted into the mass spectrometer using a gradient of 12 to 35% solution B in 10 min, 35 to 100% solution B in 50 min. The mass spectrometric parameters were as follows: ion spray voltage (IS) 2300V; curtain gas (CUR) 30; gas 1, 35; gas 2, 0; declustering potential (DP), 60V; focusing potential (FP), 265V; declustering potential (DP2), 15V. Isolation of peptide ions of interest was performed in the quadrupole Q1 set to low resolution.

3.1.9 MS data analysis

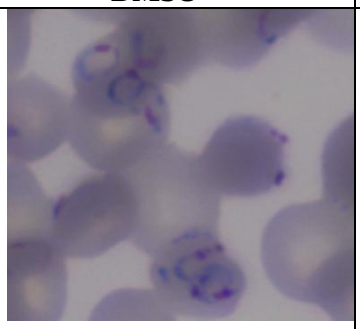
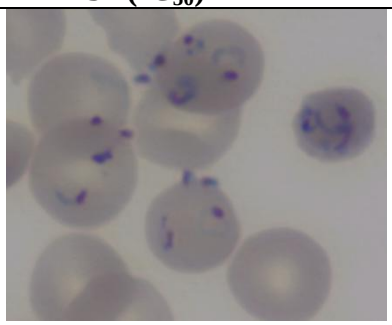
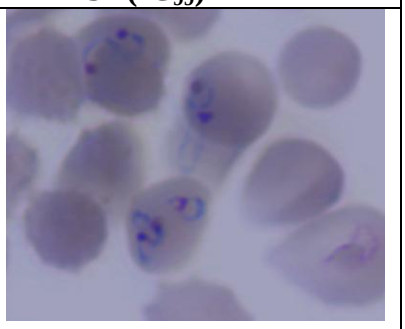
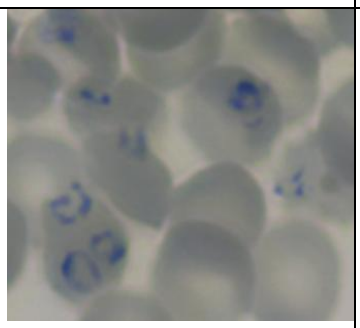
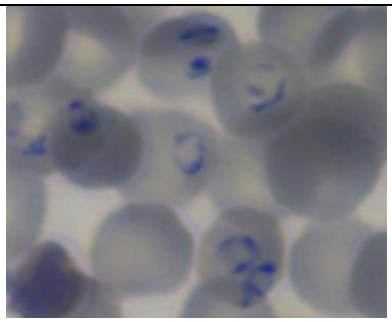
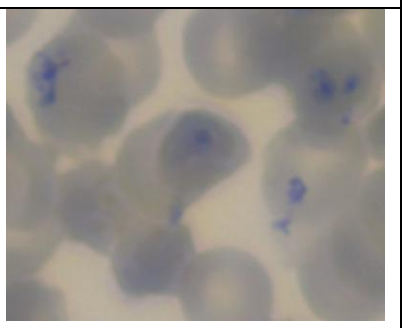
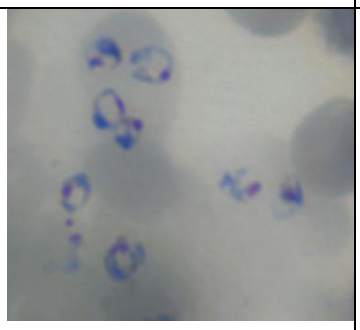
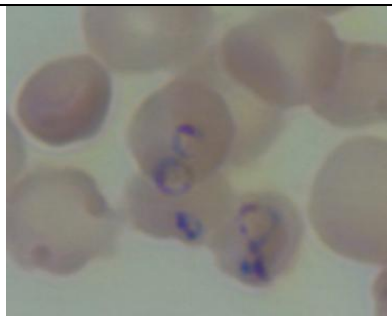
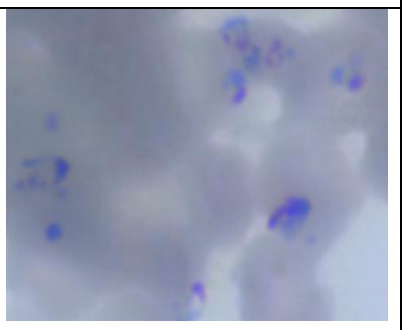
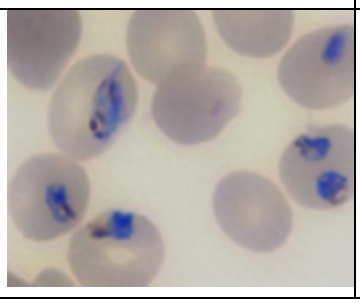
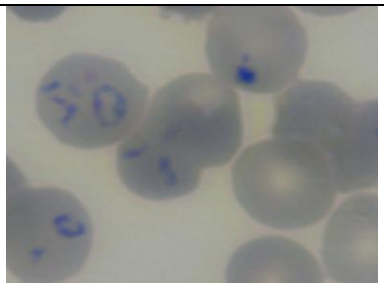
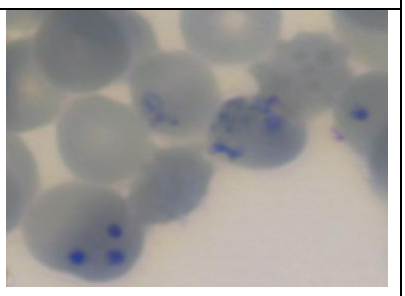
Protein identification was performed with Protein Pilot 2.0 software (Applied Biosystems). The cut-off for the confidence settings for all analyses was set at 95%. Database searching was done on predicted tryptic peptide sequence data using the PlasmoDB v5.5 database (Stoeckert *et al.*, 2006), which is specific for proteins from *Plasmodium* species. To enable label-free quantitation, raw MS spectra were converted from Applied Biosystems' WIFF file format into the open mzML and protXML file formats using the respective software packages ProteoWizard v1.6.0 (http://proteowizard.sourceforge.net) and the Trans Proteomic Pipeline (TPP) (v4.3 JETSTREAM rev 1, Build 200909091257 (MinGW)) (Keller *et al.*, 2005; http://tools.proteomecenter.org). All proteins meeting the minimum recommended TPP probability cut-off for quantitation of 0.2 (20%) were included for analysis. Absolute protein abundance values were obtained from protXML spectra and normalised based on original MS protein loading concentration using the software Absolute Protein Expression (APEX), which quantifies protein abundance by counting the number of detected tryptic peptides (versus expected tryptic peptides for each peptide) and their corresponding MS spectra (Braisted *et al.*, 2008). A 2-factor (treatment and time) Analysis of Variance (ANOVA) model, provided as part of the software package Multiple experiment Viewer (MeV v4.5 (www.jcvi.org), followed by post-hoc two-class unpaired

Rank Products analysis was used to estimate proteins that were differentially expressed following Ref#31 treatment.

3.2 RESULTS

3.2.1 Effect of Ref#31 on the life cycle of *P. falciparum* strain 3D7

The morphological study was performed in order to determine at which stage of the life cycle Ref#31 affects or inhibits the growth of the parasite (Figure 1). This information enabled a decision to be taken on harvesting time points to be used for the parasite following drug treatments for proteomic studies. The morphology study compared the growth of control (DMSO) versus Ref#31-treated parasites over a time period of 48 hrs. The culture was subjected to two different drug concentrations, IC_{50} and IC_{99}. The parasites were allowed to invade new RBC and were then treated in the mid-ring phase (6 hrs post-invasion). Twelve hours after the addition of the drug, changes were seen in the morphology of the ring parasite forms. A large proportion of the rings were abnormal in shape and size compared to the control cultures. Both IC_{50}- and IC_{99}-treated cultures remained in the late ring phase throughout the 48-hr intraerythrocytic life cycle. Between 18 hr and 48 hr in the treated cultures the parasites showed major morphological malformation. At the end of 48 hr, almost all the treated parasites were disintegrated, shrivelled or compact.

HPT	DMSO	REF#31 (IC_{50})	REF#31 (IC_{99})
0 h			
6 h			
12 h			
18 h			

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

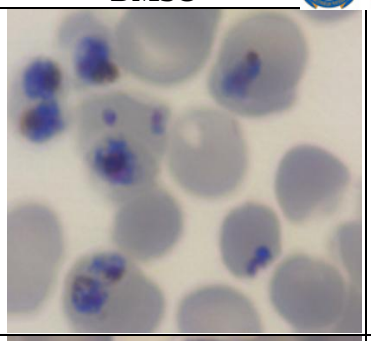
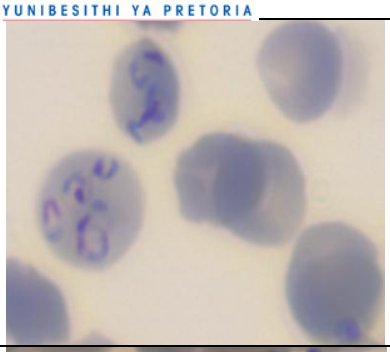
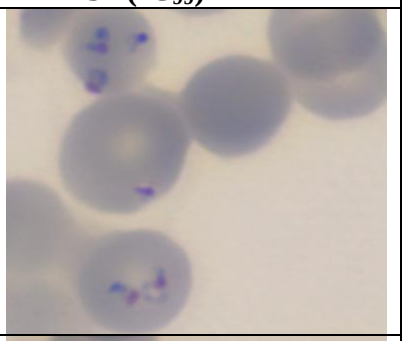
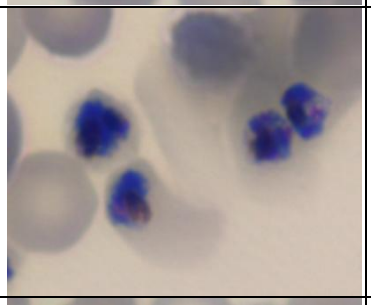
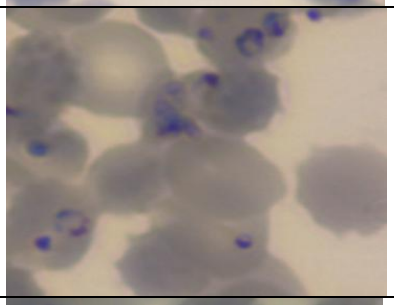
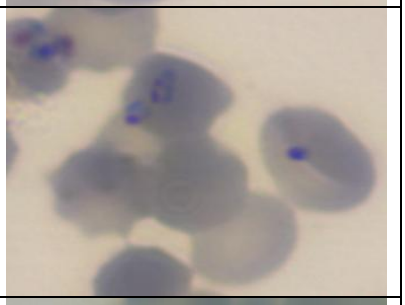
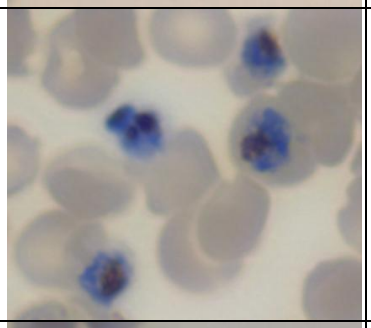
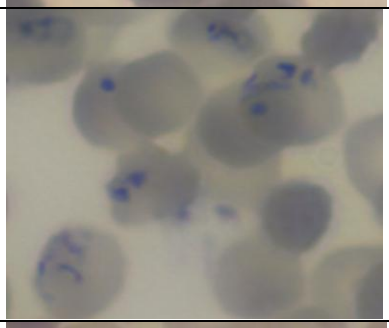
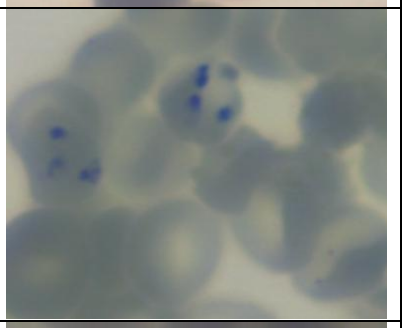
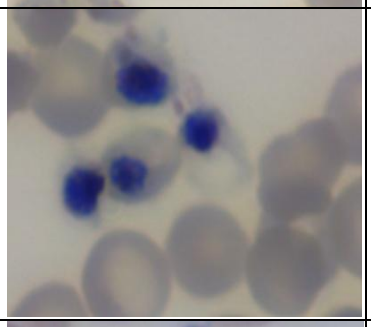
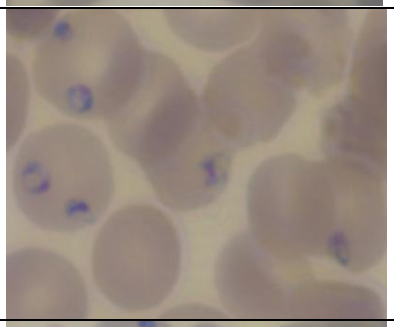
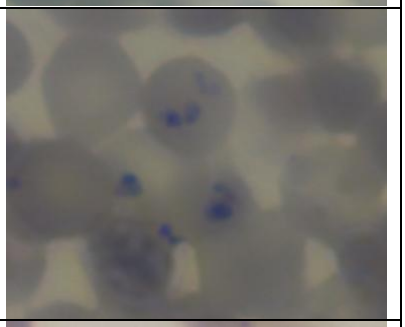
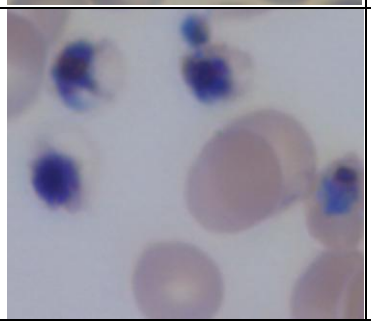
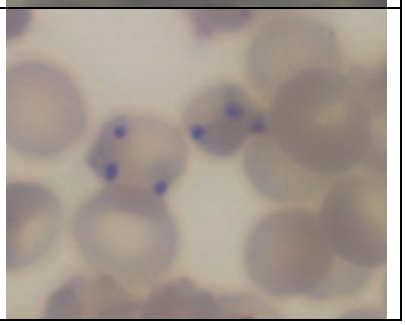
HPT	DMSO		REF#31 (IC_{99})
24 h			
30 h			
36 h			
42 h			
48 h			

Figure 1: Morphology study of *Plasmodium falciparum* 3D7 growth over time following treatment with DMSO (control) and Reference #31 (drug lead). **HPT-** Hours Post Treatment. Treatment was performed approximately 8 hrs post RBC invasion.

.

3.2.2 Protein Separation by SDS-PAGE

The results of the morphology study led to the selection of time points 6, 12 and 18 hrs upon which to investigate differential changes to the *P. falciparum* proteome following Ref#31 treatment. Selection of these time points was also influenced by *Plasmodium* microarray data obtained following Ref#31 treatment, which indicated the likely arrest of parasite growth during early trophozoite development (Drs. John Becker, Dalu Mancama, CSIR Systems Biology, personal communication). Samples from each time point and treatment condition were prepared, as described in Sections 3.2.4-3.2.5, and separated to reduce proteome complexity using SDS-PAGE, followed by in-gel trypsin digestion prior to LC/MS/MS analysis. This time the individual bands constituting each lane on the SDS-PAGE were analysed separately. Each SDS-PAGE well contained 50 µg of the total protein lysate, and prior to digestion using a modified protocol (see Section 3.1.7) each well lane was separated into 19 gel slices. Figures 2-4 illustrate SDS-PAGE gels containing total protein lysate at time points 6 hrs, 12 hrs, and 18 hrs following haemoglobin depletion. Since the concentration of the protein samples was severely limited, insufficient sample was available to run corresponding controls (depicting protein lysate samples prior to haemoglobin depletion).

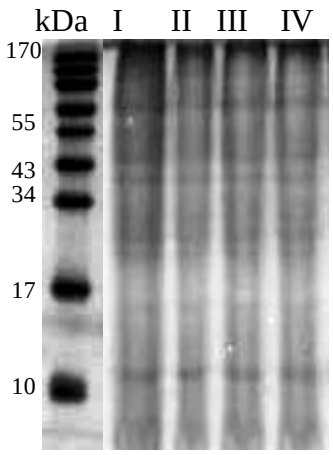


Figure 2: Protein page-stained SDS-PAGE analysis of 6-hr control versus drug-treated samples. Lane kDa, Molecular Marker (ranging from 10-170 kDa), Lanes I and II, 6-hr control, Lanes III and IV, 6-hr drug treated samples.

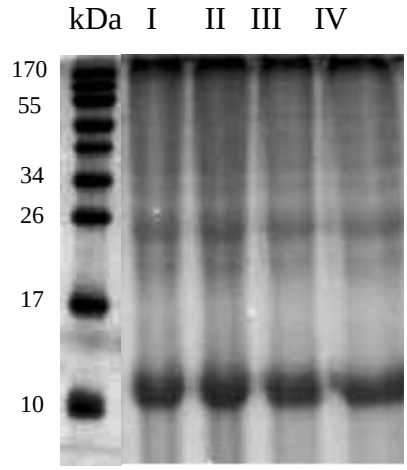


Figure 3: Protein page-stained SDS-PAGE analysis of 12-hr control versus drug-treated samples. Lane kDa, Molecular Marker (ranging from 10-170 kDa), Lanes I and II, 12-hr control, Lanes III and IV, 12-hr drug treated samples.

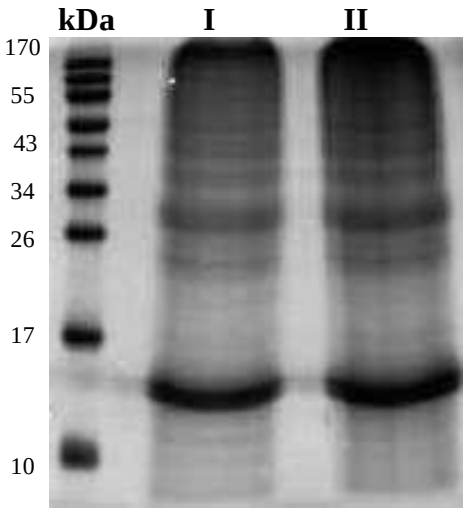


Figure 4: Protein page-stained SDS-PAGE analysis of 18-hr control versus drug-treated samples. Lane kDa, Molecular Marker (ranging from 10-170 kDa), Lane I, 18-hr control sample, Lane II, 18-hr drug treated sample. The 18-hr samples were not run in duplicate since, after extraction, there was not enough sample to run in duplicate.

3.2.3 One-Dimensional Liquid Chromatography (RP-HPLC)

Digested peptides were analysed on a QSTAR Elite running Protein Pilot v2.0 software. Peptide sequence data from Protein Pilot was then searched against the PlasmoDB v5.5 database in order to obtain the identities of the corresponding proteins. PlasmoDB v5.5 is an official database of the *P. falciparum* genome sequencing consortium (Bahl *et al.*, 2002). PlasmoDB v5.5 was used because it is specific for *Plasmodium*. Therefore the use of this database makes sure that none of the non-*Plasmodium* hits from other species show up during analysis. Likewise, no *Plasmodium* hits could be wrongly ascribed to other species containing identical peptide sequences. Results from the initial analysis of complex protein digests extracted from SDS-PAGE bands (see Section 2.2.1) and separated by 1-Dimensional Reverse Phase (i.e. 1-D RP-HPLC) chromatography had previously resulted

in poor chromatographic separation. This outcome was primarily because the SDS-PAGE lanes had been combined after fractionation. However, a different and less time-consuming in-gel digestion protocol developed during this project (detailed in Section 3.1.7), resulted in better RP-HPLC separation. In this protocol, instead of combining the fractions after digestion, the lanes were analysed separately using RP-HPLC coupled to MS/MS.

The database MSDB was used previously to enable the identification of peptides, and subsequently proteins, detected during MS analysis. This is because at the time, this was the default and only searchable database available. Substituting the MSDB database with PlasmoDB v5.5 was subsequently seen to significantly improve the number of *Plasmodium* hits obtained, as illustrated in Figure 5 (see also Appendix 2, Supplementary CD,Tables 1 and 2), which compares the number of proteins identified using PlasmoDB v5.5 and the MSDB database. Although the MSDB database enabled identification of more proteins than PlasmoDB, less were *Plasmodium* hit identities. As shown in Appendix 2, Supplementary CD Tables 3-8, a total of 91 proteins were successfully identified for the 6-hr control sample, of which 80 were *Plasmodium* proteins and 11 were non-*Plasmodium* related proteins; a total of 150 proteins were successfully identified for the 6-hr treated sample, of which 128 were *Plasmodium* and 22 were non-*Plasmodium* related; a total of 119 proteins were successfully identified for the 12-hr control sample and 126 proteins for the 12-hr treated sample; and a total of 226 proteins were successfully identified for the 18-hr control sample, of which 205 were *Plasmodium* proteins and 22 were non-*Plasmodium* related proteins, while 127 proteins were recovered for the 18-hr treated sample, of which 115 were *Plasmodium* and 12 were non-*Plasmodium* related proteins (all peptide identifications were detected at ≥95% confidence).

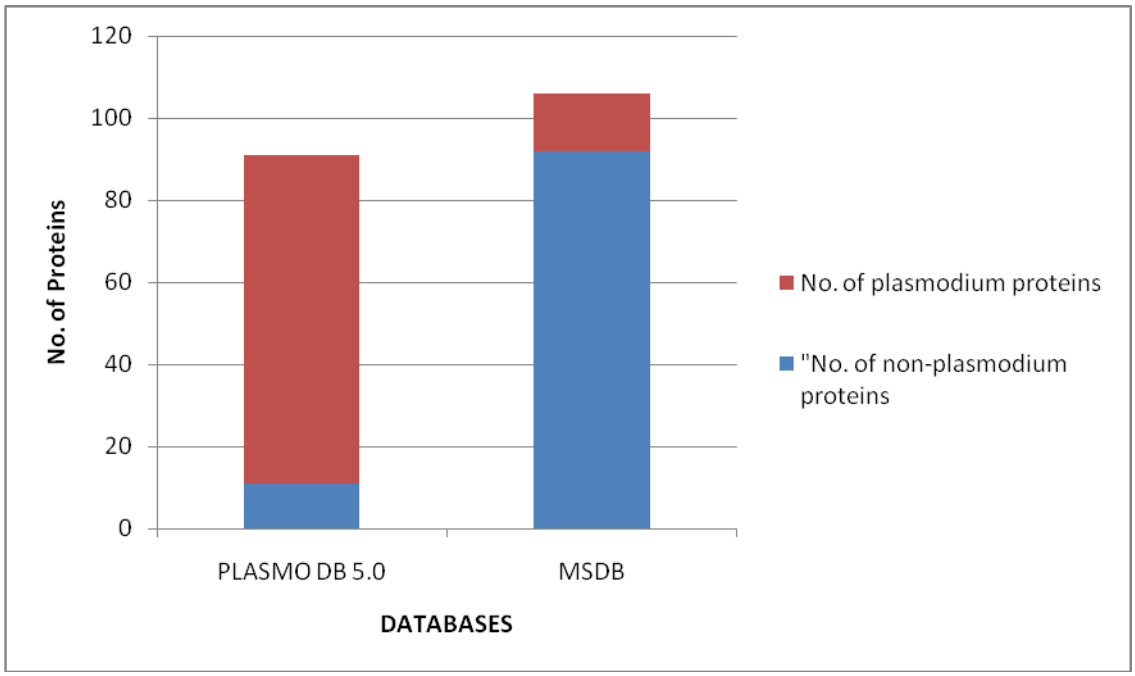


Figure 5: Comparison of two databases (PlasmoDB v5.5 and MSDB database) used for protein identification. PlasmoDB v5.5 identified more *Plasmodium* proteins than the MSDB database. The data was taken from the 6-hr control sample (Appendix 2, Supplementary CD Tables 1 and 2).

The protein identification software Protein Pilot (Applied Biosystems) provides a generalised overview of data obtained from a mass spectrometer. The information obtained by Protein Pilot includes the sequence of each successfully identified peptide and the name of the corresponding parent protein, the confidence level associated with this identification, post-translational modification events observed on each peptide, and the total percentage of the protein covered by the identified peptides. In order to obtain more in-depth information on the *P. falciparum* proteins identified in the six replicate Ref#31 samples, it was necessary to analyse the MS/MS spectra using a combination of downstream software tools. The mass spectra for each sample was converted into the recently developed platform-independent mzML spectra format using the software ProteoWizard (v1.6.0), and then exported into the Trans Proteomic Pipeline (TPP) (v4.3) suite of analytical tools

developed by the Institute of Systems Biology/Seattle Proteome Centre, Seattle, USA. Summary statistics were then first generated for all the peptides identified in each sample. An example of this summary is shown in Figure 6 below for the protein actin I (PFL2215w), obtained from the data sample 6 hrs Control Duplicate (SDS-PAGE fraction 7).

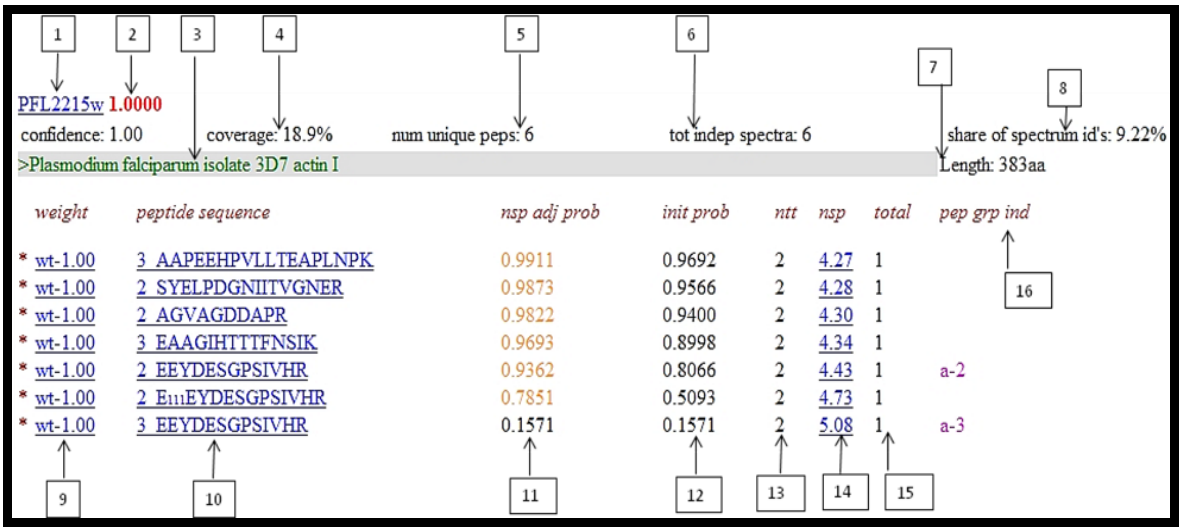


Figure 6: Illustration of the summary statistics generated for each protein in TPP.

Key: (1) PlasmoDB ID – in TPP it was possible to hyperlink this information to the PlasmoDB database, or to other databases containing relevant *P. falciparum* data; (2) confidence level of protein identification with 1.00 being the maximum; (3) Protein annotation; (4) percentage of total protein sequence covered by identified peptides; (5) number of peptides unique to the protein; (6) total number of peptides identified; (7) protein length; (8) percentage contribution of peptides in this protein to the spectrum of peptides indentified for the sample 6 hrs Control Duplicate SDS-PAGE fraction 7; (9) if a peptide is present in multiple sequence entries, this weight indicates the contribution of this peptide among each of those sequence entries. An asterisk indicates that this peptide is only found in (or is unique to) this protein entry; (10) peptide sequence and number of isoforms present; (11) nsp adj prob (number of sibling peptides adjusted probability): Proteins with multiple peptide identifications are likely more correct than proteins with single peptide identifications. Different peptides from the same protein are termed sibling peptides. The number under this column is the adjusted peptide probability after taking into account its

number of sibling peptides in this dataset. These adjusted peptide probabilities are used to calculate the protein probability; (12) original or initial probability assigned to this peptide; (13) number of tolerable (tryptic) termini (NTT) indicates whether the peptide sequence exhibits zero, one, or two expected cleavage termini; (14) this bin number (ranging from 0 to 7) reflects the number of sibling peptides after discretization; (15) number of instances that this peptide was identified in the underlying dataset(s); (16) (peptide group indicator): if a peptide was identified from spectra acquired on ions of different precursor charge states, these different identifications of the same peptide are considered independent evidence for that peptide identity. The same peptide is listed multiple times for each charge state and the different charge states are indicated in this column

This information is useful for providing an in-depth understanding of the nature of the peptides identified in each sample. The processing of tandem mass spectra data in the Trans Proteomic Pipeline enabled samples to be analysed further by label-free absolute protein quantitation. In order to perform expression analysis the PlasmoDB v5.5 database and a protein accession list file comprising common proteins identified across the Ref#31 sample dataset were imported into APEX (v1.0.0; www.jcvi.org). This software improves on basic spectral counting methods by including a correction factor for each protein (called an *Oi* value) that accounts for variable peptide detection by MS techniques. The technique uses machine learning classification to derive peptide detection probabilities that are used to predict the number of tryptic peptides expected to be detected for one molecule of a particular protein (*Oi*). This predicted spectral count is compared to the protein's observed MS total spectral count during APEX computation of protein abundances (Braisted *et al.*, 2008).

3.2.4 Differential protein expression analysis following Ref#31 treatment

APEX absolute protein abundance values from Ref#31 treated and untreated *P. falciparum* replicate groups were compared using two-way analysis of variance (ANOVA), provided

as part of the software package MeV v4.5 (www.jcvi.org), to identify proteins that are differentially expressed following drug treatment. Two-way ANOVA was also chosen to factor into the analysis the effects of sampling across three different time points. To increase the pool of proteins used to estimate differential expression between the treated and untreated gorups, all proteins meeting the recommended minimum TPP probability cut-off of 0.2 (i.e. 20% confidence, in contrast to 95%) were included in the analysis dataset. As a result, a total of 2087 unique proteins were included in the dataset. Given the inclusion in this dataset of proteins identified with relatively low confidence, independent validation of those found to be differentially expressed would be required to conclusively confirm their identity. This can in part be achieved by the manual but tedious examination of the ion spectra evidence, or through downstream targeted validation strategies using high-resolution mass spectrometry (e.g. through the use of a QTrap if available). Ideally, ultimate validation of differential expression could be achieved through western blot analysis of total protein extracts from treated and untreated parasite samples, using antibodies that are specific to the predicted protein(s) of interest. Two-way ANOVA was chosen to estimate the effects of the variables treatment and time on protein expression, based on the balanced design of the dataset (two treatment groups sampled in duplicate at time points 6 hrs, 12 hrs and 18 hrs). In order to improve statistical validity, expression analysis should, however, have been performed in sample groups of triplicate or more. However, this was not possible, due primarily to the unavailability of time and equipment needed to repeat the experiment using triplicate samples. Therefore, results observed from the quantitative analysis were considered to be tentative, and useful for establishing the methodology for quantitative analysis. The results require future confirmation through experiment repetition using triplicate samples and a combination of the targeted approach highlighted above. To account for missing protein expression values in the dataset, random

permutation tests set to a stringent level of 10^6 were applied to the data. Two-way ANOVA was also chosen since significantly fewer replicates are required with this approach compared to non-parametric methods to achieve the same level of statistical power, as demonstrated on data derived from single-colour oligonucleotide microarrays (Barrera *et al*., 2004). Two-way ANOVA was performed at a significance level of $p \leq 0.05$ and results adjusted for multiple testing using relatively conservative Bonferroni's correction, which multiplies the significance p statistic by the number of results generated. The number of proteins estimated to be differentially expressed is summarised in Figure 7. Forty-seven proteins were found to be uniquely affected by treatment, while three were affected independently by treatment and time. A synergistic effect of treatment and time was observed on 37 proteins. In total, 1892 (91%) proteins were unaffected by treatment, time or potential interactions between these test variables. A full list of the differentially expressed proteins is given in Appendix 2, Supplementary CD Tables 9-11. Post-hoc analysis to determine fold changes to proteins estimated by two-way ANOVA to be significantly affected by treatment was performed in MeV using the two-class unpaired Rank Products method.

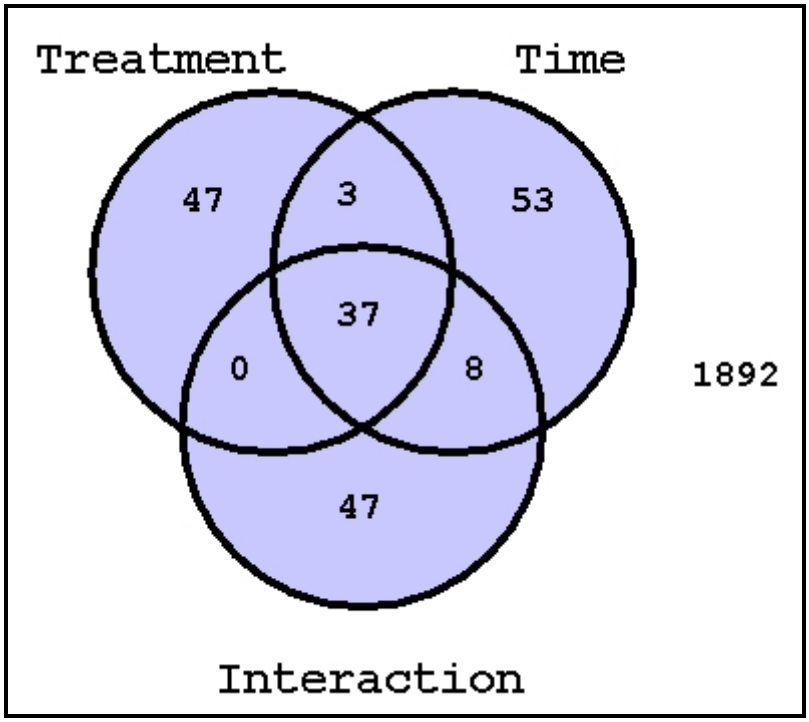


Figure 7: Venn diagram summarising the proteins estimated through two-way ANOVA to be differentially expressed in *P. falciparum* following treatment with Ref#31. Comparisons were performed as detailed above in section 3.2.4. 'Treatment' denotes the number of proteins that were differentially expressed between the treated and untreated sample groups due to the effects of Ref#31 alone. None of these changes were significantly influenced by any underlying changes that occur over time during the parasite development cycle. 'Time' denotes the number of proteins whose expression was significantly affected over time as a consequence of parasite development. None of these changes were significantly influenced by parasite treatment with Ref#31. 'Interaction' denotes the number of proteins whose expression was found to be significantly influenced both by Ref#31 treatment and by changes that occur over time in the parasite during development.

Comparisons of each treated sample (i.e. at 6 hrs, 12 hrs, 18 hrs) were performed relative to the 6-hr control sample at a significance level of $p \leq 0.05$, and results given for proteins differentially expressed at the default software setting of 1.4-fold higher or lower. These proteins are listed in Table 1.

Table 1: *P. falciparum* proteins found to be differentially expressed greater or lower than 1.4-fold, following treatment with Ref#31

6 hrs post- treatment			
PlasmoDB ID	**Annotation**	**significance p**	**Fold change**
PF11_0354	conserved Plasmodium protein, unknown function	0.021264369	1.998902
PFC0335c	conserved Plasmodium protein, unknown function	0.03666667	1.4140837
MAL7P1.147	ubiquitin carboxyl-terminal hydrolase, putative	0.001034483	1.9994781
MAL13P1.256	phosphatidylinositol transfer protein, putative	0.026551723	0.7076364
PF11_0242	calcium-dependent protein kinase, putative	0.03988506	0.7072286
MAL13P1.120	splicing factor, putative	0.03229885	0.70722395
PF14_0425	fructose-bisphosphate aldolase	0.012068965	0.70717466
12 hrs post-treatment			
PFL0590c	non-SERCA-type Ca2+ -transporting P-ATPase	0.017126437	1.9995828
MAL13P1.13	conserved Plasmodium protein, unknown function	0.03678161	1.9995177
PF07_0118	conserved Plasmodium membrane protein, unknown function	0.003678161	1.9997005
PFC0335c	conserved Plasmodium protein, unknown function	0	1.9998239
MAL13P1.296	conserved Plasmodium protein, unknown function	0.012068965	1.9996133
PFL0115w	dynein heavy chain, putative	6.90E-04	1.9997894
PF14_0379	conserved Plasmodium protein, unknown function	0.04781609	1.9993541
MAL13P1.256	phosphatidylinositol transfer protein, putative	0.011149426	0.7076364
PFF1300w	pyruvate kinase	0.04505747	0.70769674
MAL7P1.126	conserved Plasmodium protein, unknown function	0.04954023	0.70784533
PF11_0242	calcium-dependent protein kinase, putative	0.0408046	0.7072286
MAL13P1.120	splicing factor, putative	0.017126437	0.70722395
PF14_0425	fructose-bisphosphate aldolase	0.008850575	0.70717466
18 hrs post-treatment			
MAL13P1.296	conserved Plasmodium protein, unknown function	0.03735632	1.414172
PFD0462w	DNAJ protein	0.049195405	1.4141091
MAL13P1.120	splicing factor, putative	0.049770113	0.70722395
PF14_0425	fructose-bisphosphate aldolase	0.03321839	0.70717466

Based on the preliminary evidence provided from the expression analyses, a diverse range of *Plasmodium* proteins were observed to be potentially affected by Ref#31 treatment, including those involved in protein degradation, phosphorylation and transcription, amongst others. The potential relevance of these changes is discussed, together with changes observed in protein post-translational modification, in Section 3.2.5 below. Given the inclusion in the dataset, however, of proteins identified with relatively low confidence, confirmation of the identities of some of these proteins will be required at higher confidence levels. One approach would be to perform targeted proteomics using high-resolution mass spectrometry, to identify and compare levels of specific proteins of interest before and following Ref#31 treatment. At the same time, acquisition by the CSIR of new proteome fractionation equipment previously not available during this study now means that significantly more proteins can in future be identified and characterised with relatively higher confidence.

3.2.5 Post-translational analysis of protein modifications following Ref#31 treatment

It was originally planned that the protein expression results be subsequently repeated, by labelling each of the samples in the dataset in triplicate using iTRAQ labels, then repeating the mass spectra acquisition and downstream analysis. The primary advantage of this approach is that it would enable each of the samples in the dataset to be pooled and run in parallel, thereby accelerating the data acquisition process and minimising any potential variability that might otherwise occur between samples due to sample processing and handling. During the course of the project, new research findings emerged indicating that compounds containing an alpha-methylene-gamma-lactone moiety, such as Ref#31, inhibit eukaryotic cell growth through mechanisms involving changes to protein post-translational modification. Given this new evidence, combined with time-constraints that limited

repeated mass spectra acquisition due to decommissioning of the mass spectrometer to replace the vacuum pump system, a decision was taken to focus the remainder of the data analysis on investigating post-translational changes to the *P. falciparum* proteome following Ref#31 exposure as described below. For future high-throughput analyses of the *P. falciparum* proteome, establishment of the iTRAQ approach means, however, that this is now available for routine use, with the primary advantages being the ability to run multiple samples in parallel.

Mass spectrometry is a powerful technology that enables the identification and quantitation of proteins in biological systems. Recently, developments in the technology such as increased detection resolution and sensitivity now make it possible to also accurately establish the post-translational modification status of a protein's amino acid residues. Post-translational modifications of proteins control many biological processes, and examining their diversity is critical for understanding mechanisms of cell regulation. This information is becoming increasingly important to understanding biological processes in organisms of interest, and has received growing attention in *Plasmodium* research (Witze *et al.*, 2007). Researchers at CSIR Biosciences recently established that the alpha-methylene-gamma-lactone group in Ref#31 is essential for the compound's antiplasmodial activity, as modification of this group significantly reduces this activity (Dr. Vinesh Maharaj, CSIR Bioprospecting, personal communication). This moiety has also been found to be critical in other structurally related compounds, such as parthenolide and arglabin, for cell growth inhibition. Furthermore, it has been proposed that the mode of action of these compounds involves alterations to the post-translational modification process, leading eventually to cell cycle arrest (Gopal *et al.*, 2009). Interestingly, parthenolide has also been shown to exhibit antiplasmodial activity (Gopal *et al.*, 2009). To investigate the possible influence on *P.*

falciparum of post-translational modification events mediated by Ref#31, the post-translational modification status of peptide sequences acquired in the six experimental groups was investigated. Given the vast post-translational changes that occur in the parasite, close attention was paid to proteins in pathways implicated in eliciting the effects of the alpha-methylene-gamma-lactone group, namely proteins involved in histone acetylation and deacetylation, and those involved in protein ubiquitination (Gopal *et al.*, 2007, 2009). Other modifications of related potential interest included proteins involved in methylation and phosphorylation. A summary of the proportion of *P. falciparum* peptide sequences found to contain post-translational modifications is given in Figures 8-10.

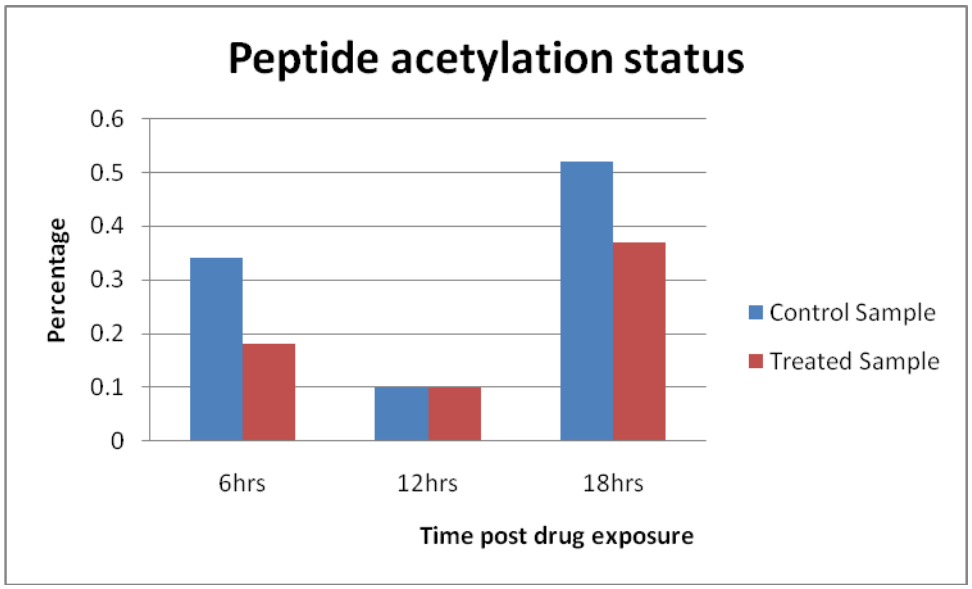


Figure 8: Proportion of *P. falciparum* peptide sequences observed to be acetylated.

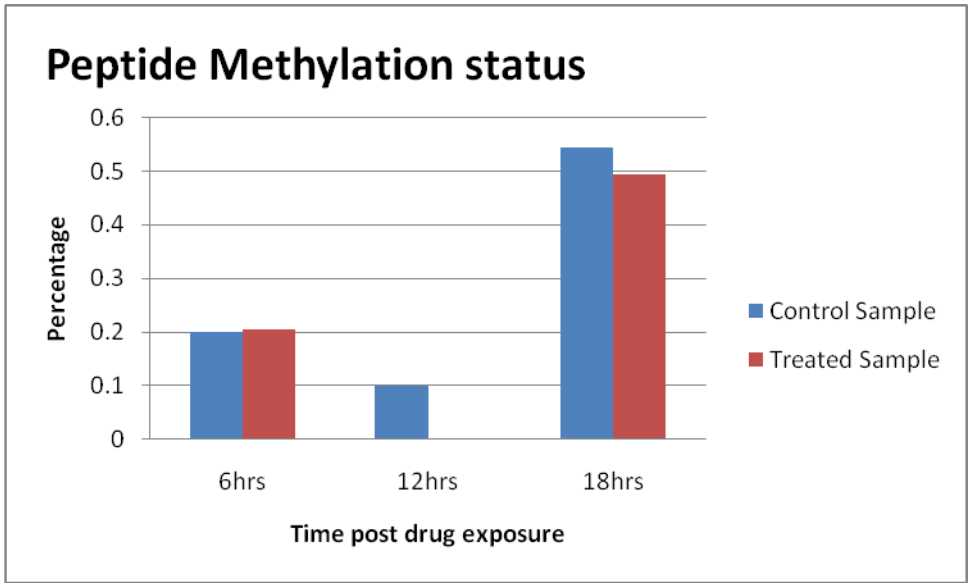


Figure 9: Proportion of *P. falciparum* peptide sequences observed to be methylated.

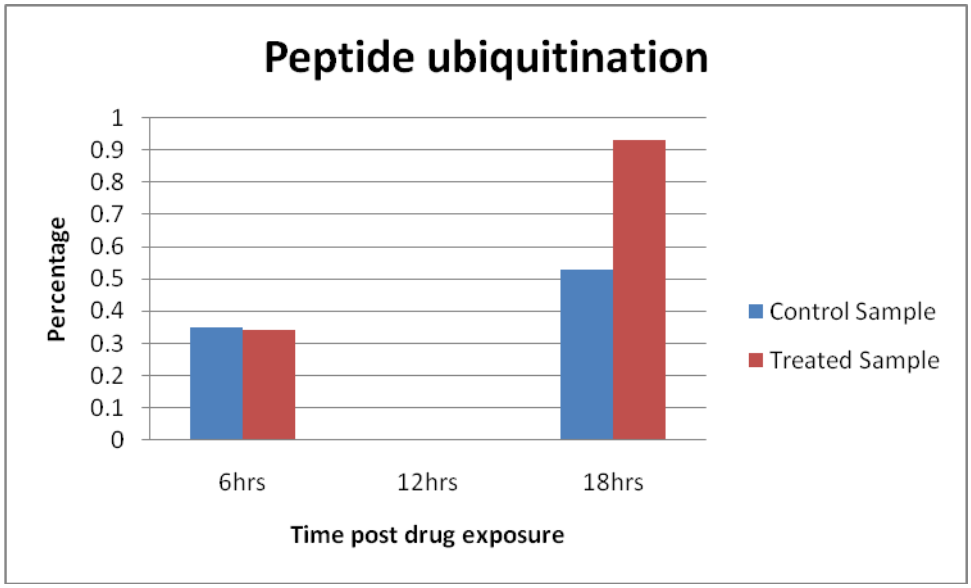


Figure 10: Proportion of *P. falciparum* peptide sequences observed to be ubiquitinated.

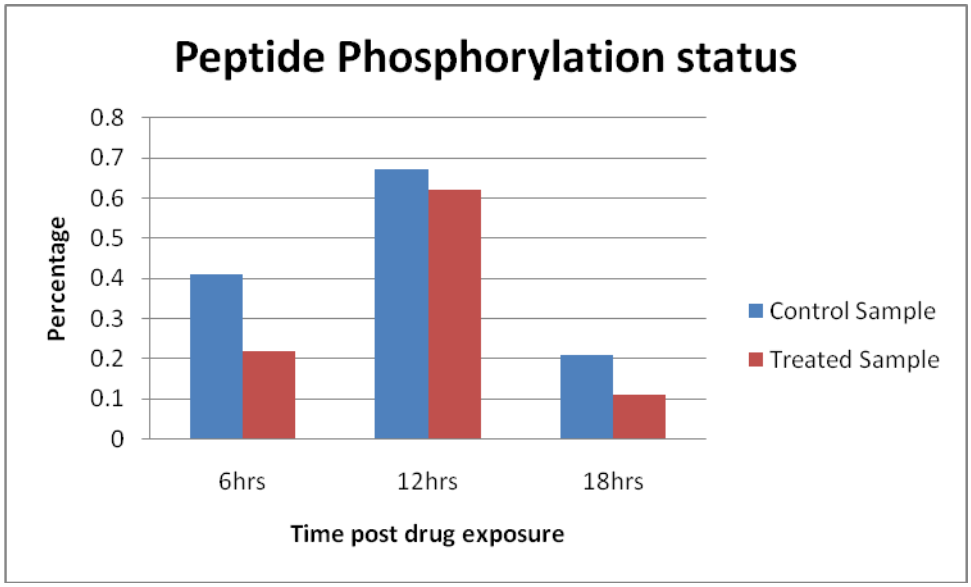


Figure 11: Proportion of *P. falciparum* peptide sequences observed to be phosphorylated.

High-throughput mass spectrometry profiling of the *P. falciparum* proteome following Ref#31 exposure was found to generate a huge amount of data (approximately > 125Gb of computer memory), where at present no software is available to comprehensively investigate all the post-translational modification events that are observed in such large datasets. Therefore, to make analysis of this data possible, only proteins in pathways implicated in the inhibitory effects of the alpha-methylene-gamma-lactone group were selected, and amino acid residues examined to map post-translational changes of potential importance. An example of this mapping is illustrated in the figure below (Figure 12) for the histone H2A protein (PFF0860c), where post-translational data was obtained in the control sample at time points 6 and 12 hrs post-drug exposure, and at 12 hrs in the treated sample. Since peptide sequences for histone H2A were not detected with high statistical confidence (<95%) at 18 hrs, no post-translational data is shown at this time point. Similarly, no data was available for the treated sample at 6 hrs and 18 hrs post-drug exposure.

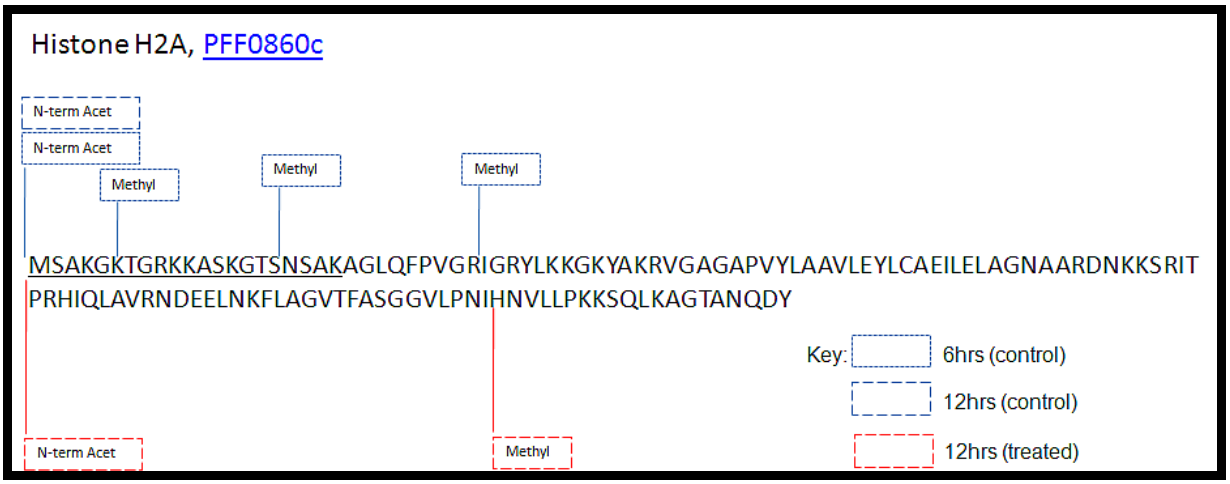


Figure 12: Illustration of the acetylation and methylation status of amino acid residues belonging to histone H2A. Since not all peptide sequences were detected at a high confidence level ≥95%, information from lower-confidence identifications was not included.

As shown in Appendix 2, Supplementary CD Tables 12-15, the *Plasmodium falciparum* proteome undergoes extensive post-translational modification during the course of development, with acetylation, methylation, ubiquitination and phophorylation being widely observed in protein peptide sequences originating from both untreated and treated sample groups. The most extensive modification observed in both cases was protein phosphorylation where, for example, at 6 hr post-treatment a total of 40 325 and 45 454 phosphorylated peptides (which include isoform variants) were identified in the untreated and treated samples, respectively, while at 18 hr 39 437 and 923 phosphorylated peptides were identified, respectively. Due to the extensive nature of this PTM event, this data has not been included for the sake of brevity. During the course of analysing the large amount of PTM data it became clear that without the presence of an *a priori* hypothesis which enables investigation of specific proteins of interest, limitations in current mass spectrometry software make accurate and meaningful interpretation of this information highly complex and challenging. For this project, however, *a priori* selection of histone

deacetylase 1 (PlasmoDB ID PFI1260c), a putative ubiquitin (E3) ligase (PlasmoDB ID PFF1365c) and alpha tubulin (PlasmoDB ID PFI0180w) enabled PTM events to be specifically examined in these proteins in light of recent findings for this class of drugs. Thus, compounds in this drug class are believed to elicit eukaryotic cell growth inhibition through mechanisms elicited by the depletion of histone deacetylase activity, an enzyme involved in chromatin remodelling, and depletion of ubiquitin E3 ligase, involved in cell cycle progression. In both instances, drug treatment has been observed to elicit ubiquitination of these proteins, and their subsequent degradation via the proteasome machinery (Gopal *et al.*, 2007, 2009). Compounds belonging to this class have also been shown to disrupt the alpha tubulin tyrosination cycle, required for the control and mechanics of chromatin remodelling and cell division, through inhibition of tubulin carboxypeptidase (TCP) activity (Fonrose *et al.*, 2007). This enzyme is normally responsible for enzymatic removal of the COOH-terminal tyrosine residue on alpha-tubulin (Fonrose *et al.*, 2007).

In this present study the presence of peptide sequences originating from histone deacetylase was not detected in either the treated or untreated samples at 6 hr, 12 hr and 18 hr, and it was therefore not possible to establish the effect, if any, of ubiquitination on this protein following drug treatment. In the case of alpha tubulin, the detection of carboxy terminal tyrosine residue cleavage indicating tubulin carboxypeptidase enzyme function suggests that the effects of Ref#31 in *P. falciprum* are unlikely to be mediated through direct inhibition of this enzyme as proposed earlier for parthenolide (Fonrose *et al.*, 2007). However, more data points sampled from across the parasite's 48-hr development cycle following Ref#31 treatment would be required to confirm this finding. More interestingly,

several peptide sequences comprising a putative ubiquitin E3 ligase, believed to contribute to controlling cell cycle progression, were found to be ubiquitinated following treatment, while this was not observed in corresponding sequences and all variants (protein isoforms) isolated from the untreated samples (highlighted in Appendix 2, Supplementary CD), Table 14). At the same time, a significant increase in the level of a putative ubiquitin carboxyl-terminal hydrolase (MAL7P1.147) was observed 6 hrs after Ref#31 exposure (see Table 1). During proteosome-mediated protein degradation, C-terminal polyubiquitin is hydrolysed by ubiquitin carboxyl-terminal hydrolase to monomeric ubiquitin, which is conjugated through a series of ubiquitin-activating and -conjugating enzymes, including ubiquitin E3 ligase, to the target protein. Closer analysis of this process following Ref#31 treatment, in combination with other functional data obtained through microarray analysis for example, may therefore be warranted.

3.3 DISCUSSION

Ever-increasing resistance of *Plasmodium falciparum* to the current range of antimalarials has led to intensified efforts to identify new drugs and new drug targets against this parasite. Recently, *Dicoma anomala* (*D. anomala*) ssp. *gerardii*, a highly variable indigenous plant, has been shown to possess good antimalarial activity *in vitro*. However, currently, no knowledge exists regarding the molecular effects of *D. anomala* on *P. falciparum*, nor how these effects inhibit the growth of this parasite. In order to better understand the possible molecular effects of an antimalarial lead derived from this plant (Reference #31 (Ref#31)), a high-throughput approach was employed to investigate the effect of this compound on the parasite's proteome.

To guide the use of this approach, a morphology study was initially undertaken. Morphological studies indicated that Ref#31 affects the parasite's growth between the late ring phase and early trophozoite stage of the life cycle (Figure 1). Within 18 hrs of drug treatment, the control DMSO-treated cultures were already showing signs of early trophozoite development; whereas Ref#31-treated cultures still only exhibited the late ring phase of parasite growth. At this stage, for cultures treated with an IC_{99} drug concentration, almost all parasites were disintegrated and the rings were recognised with difficulty. Thirty hours post-treatment, clear differences existed between the control and drug-treated cultures, where control cultures were predominantly in the schizont phase, whereas treated cultures were still in the early trophozoite phase of growth. Within 48 hrs, parasites in the control DMSO-treated culture had gone through trophozoite and schizont phases of their life cycle, and, in some instances, very early rings structures were once again visible. However, cultures treated with Ref#31 did not appear to develop past the early trophozoite stage throughout the 48-hr period. The morphological changes observed with Ref#31 were

similar to those observed by Kombila in 1997, who treated four Gabonese children infected with malaria with pure artemether. Changes to the parasite occurred between 5-24 hr after adding artemether (Kombila *et al.,* 1997). The few remaining rings were shrivelled and compact. Based on the morphology study, Ref #31 exhibits a quick onset of action and high efficacy against the blood stages of *Plasmodium*.

Based on the outcome of the morphology studies, it was possible to subsequently use this information to design a study to investigate the effects of Ref#31 on the *P. falciparum* proteome. This was designed to initially involve identifying and quantifying differentially expressed proteins using a combination of iTRAQ and a high-throughput proteomics approach employing online 2D HPLC (strong cationic exchange / reverse phase HPLC) coupled with mass spectrometry. Although this approach has been successfully used to study the proteome of organisms such as yeast and bacteria (Washburn *et al.,* 2001; Vollmer *et al*., 2003), results from this present study suggest that the approach was not optimal for studying the *Plasmodium* proteome. Reasons for this include the potentially more complex proteome of *Plasmodium*, and the presence of highly abundant proteins in the *Plasmodium* protein lysate, in particular haemoglobin, which was observed to mask detection of lower abundance proteins. A recently developed alternative to high-throughput online SCX/RP mass spectrometry proteomics involves HPLC-based fractionation of complex proteome lysates, and the subsequent individual analysis of these fractions using reverse phase HPLC coupled to mass spectrometry (Peng *et al.,* 2003). A similar alternative approach, substituting SDS-PAGE for HPLC-based fractionation, has also been developed. Due to the unavailability of suitable HPLC fractionating equipment, the SDS-PAGE approach was selected to fractionate the *Plasmodium* proteome, prior to mass spectrometry analysis. Additional changes that improved proteome yield included the

incorporation of a bigger C18 HPLC column, amending the original SDS-PAGE protocol (Section 3.1.7), and the use of a depletion strategy to remove highly abundant haemoglobin. An attempt to further simplify the *Plasmodium* proteome by depleting haemoglobin resulted in a two-fold increase in *Plasmodium* proteins. This strategy would also be useful if the SDS-PAGE results in Figure 1 (Section 2.2.4) were analysed further. Using SDS-PAGE, another option is to simply not cut out the bands containing haemoglobin; however, the downside is that some *Plasmodium* proteins may also be lost.

Protein extracts from 6-, 12- and 18-hr samples were subjected to haemoglobin depletion and then separated with SDS-PAGE. Analysis of 6-hr samples (control and drug-treated) with 1D-RP-HPLC resulted in the identification of 91 proteins for the 6-hr control sample, 80 *Plasmodium* proteins and 11 non-*Plasmodium* related proteins; 150 proteins recovered for the 6-hr treated sample, 128 *Plasmodium* and 22 non-*Plasmodium* related proteins (Appendix 2,Supplementary CD), Tables 3- 8). A total of 119 proteins were identified for the 12-hr control sample and 126 proteins identified for the 12-hr treated sample. A total of 226 proteins were recovered for the 18-hr control sample, 205 *Plasmodium* proteins and 22 non-*Plasmodium* related proteins; 127 proteins were recovered for the 18-hr treated sample, 115 *Plasmodium* and 12 non-*Plasmodium* related proteins. Hypothetical and proteins of unknown function made up the largest category of proteins that were identified. Using a recently developed label-free approach, it was possible to estimate the relative abundance of these proteins, and to determine those likely to be differentially expressed in *P. falciparum* following Ref#31 treatment. However, since differential expression analysis was performed on samples in duplicate, and it was not possible to repeat these analyses in triplicate due to the unavailability of time and equipment, further statistical analysis using adequately replicated sample sets will be required to confirm the preliminary data observed

in this work. At the same time, validation of potential differentially expressed proteins could be achieved using a targeted mass spectrometry approach, where ions representing peptides for these proteins are specifically targeted in the mass spectra, and expression levels measured in triplicate samples. In total, preliminary analysis indicated that the expression of less than 10% of the proteins detected for the parasite were found to be affected; however, the processes affected appeared to be diverse. Since the identities of some of these proteins were achieved at relatively low confidence, further work through a targeted approach, together with in-depth analysis of the ion spectra evidence obtained for each protein of interest, will be required. This is particularly the case for a putative ubiquitin carboxyl-terminal hydrolase (MAL7P1.147), believed to be involved in protein post-translational modification.

Recently emerged evidence implicates the involvement of post-translational modification (PTM) events in the cell growth inhibition efficacy of alpha-methylene-gamma-lactone ring-containing compounds, a class to which Ref#31 belongs. In particular, structurally related compounds in this drug class such as parthenolide (PN) have been investigated and are believed to elicit eukaryotic cell growth inhibition through mechanisms elicited by the depletion of histone deacetylase activity, an enzyme involved in chromatin remodelling, and depletion of ubiquitin E3 ligase, involved in cell cycle progression(Gopal *et al.*, 2007, 2009). To determine whether these events might also be relevant to *P. falciparum* following Ref#31 treatment, data from these events were collated for all the proteins profiled at time points 6 hrs, 12 hrs and 18 hrs. Of particular interest were PTM changes involving histone acetylation and methylation, and protein ubiquitination associated proteosome-mediated degradation. Although changes in protein phosphorylation were also potentially relevant, the extensive changes observed for this PTM in *P. falciparum*

necessitates the development of high-throughput bioinformatics tools to enable accurate meaningful interpretation of this event. Although a number of peptide sequences encoding *P. falciparum* histones (H2A/H2B/H3/H4) were detected, those corresponding to lysine residues on histone tails were not picked up. It was therefore not possible to determine the PTM status of these residues following treatment and, in particular, whether there was an overrepresentation of acetylated histone tail lysine residues in this sample across time that would suggest impaired histone deacetylase activity. Detection of histone deacetylase (PlasmoDB ID PFI1260c) was not observed at a confidence level ≥95%, making it not possible to establish the PTM status, specifically ubiquitination, of this protein following treatment. The targeted profiling of *P. falciparum* histone deacetylase and histones following Ref#31 treatment, and subsequent PTM analysis of histone tail and histone deacetylase lysine residues, may shed more light on this. Isolation could be done through, for example, the isolation of histone- and histone deacetylase-containing fractions on SDS-PAGE (Beck *et al*., 2006), or through the use of alternative MS technology such as an Ion-Trap, programmed to specifically characterise with high confidence peptides derived from these proteins. Investigation of a putative *P. falciparum* ubiquitin E3 ligase (PlasmoDB ID PFF1365c), involved in protein degradation and cell cycle control, did, however, establish differences in the ubiquitination status of this protein in treated cultures compared to untreated controls, where no ubiquinated peptide sequences (or isoforms of these) were detected in the latter at the time points investigated. Although this mirrors similar findings for other compounds in this drug class such as parthenolide and arglabin, the full implications of these changes are presently unclear, particularly regarding time-related changes during development. Extensive work is required to establish the chain of events leading to the ubiquitination of this protein, and to demonstrate in *P. falciparum* how this in turn might lead to cell cycle arrest. The inclusion of data arising from parallel studies of *P.*

falciparum, where for example significant changes in expression have been observed for genes involved in histone and ubiquitin cellular function following Ref#31 treatment (Drs. John Becker / Dalu Mancama, CSIR Systems Biology, personal communication), may thus prove useful to establishing the role of these systems, if any, in the mode of action of this compound.

4 SUMMARY

The lethal species of malaria parasite, *P. falciparum*, continues to exact a huge toll of mortality and morbidity, particularly in sub-Saharan Africa (Sims *et al.*, 2006). In light of this there is an urgent need to develop novel, affordable antimalarials. This urgency has been further highlighted by the increasing prevalence of drug-resistant strains of the malaria parasite *P. falciparum*, which have contributed to the escalating disease burden (Clarkson *et al.*, 2003). Advances in proteomics and mass spectrometry have opened up unprecedented opportunities for understanding the complex biology of this parasite and how it responds to drug challenge and interventions (Sims *et al.*, 2006). Proteomic analysis of *Plasmodium* proteins began nearly two decades ago with investigations of 2D gel patterns of parasite proteins before and after drug treatments (Sims *et al.*, 2006). Application of proteomics technologies in the investigation of biological systems creates new possibilities in the elucidation and discovery of novel drug targets. A proteomic analysis involves protein separation and protein identification, as well as quantitation and most recently characterization of post-translational modifications. Proteomics has been applied in the investigation of various disorders, and the application has resulted in the detection of a large number of differences in the levels and the modifications of proteins between healthy and diseased states (Garbis *et al.*, 2005).

In this project, a novel antimalarial lead Ref#31 was tested *in vitro* for its efficacy against the blood stages of *P. falciparum*. This compound was previously shown to have antimalarial activity *in vitro* (Dr. Vinesh Maharaj, CSIR Bioprospecting, personal communication). In this present study, *in vitro* results of the morphology of the parasite following treatment looked promising. The drug demonstrated that it affects the parasite at the ring stage of its life cycle prior to transition through the trophozoite stage. Within 6 hrs

of drug treatment (performed at mid ring stage), the control DMSO-treated cultures were already showing signs of early trophozoite development, whereas Ref#31-treated cultures still only exhibited the ring phase of parasite growth. In order to better understand the molecular mechanisms involved in mediating this effect, proteomic analysis of Ref#31-treated samples versus untreated samples was undertaken.

Having established the morphological effects of Ref#31 on *P. falciparum*, the focus of the project was to identify and quantify differentially expressed proteins by developing and using a combination of iTRAQ and a high-throughput proteomics approach employing 2D HPLC (strong cationic exchange/reverse phase HPLC) coupled with mass spectrometry. Although this technique has been used successfully in other eukaryotic systems (Aggarwal *et al* 2006), results from this study suggest that the approach was not optimal for studying the *Plasmodium* proteome. A comprehensive series of protocols to optimise iTRAQ and mass spectrometry therefore had to be undertaken to tailor the approach to investigating *Plasmodium*, and to maximise the amount of proteins identified. Due to MS equipment unavailability and time constraints, quantitation was, however, eventually performed using a recently developed label-free approach.

Initial MS analysis demonstrated that the *Plasmodium* proteome is complicated by the presence of haemoglobin. Haemoglobin masks the presence of low abundance proteins. A depletion strategy that removes most of the haemoglobin and leaves behind mostly specific proteins (i.e. *Plasmodium* proteins) was therefore explored. 1D-RP-HPLC was eventually the method chosen for identifying the proteins expressed by the *Plasmodium* parasite before and after drug treatment. The protein lysates were first depleted of haemoglobin and fractionated with SDS-PAGE before running the sample on 1D-RP-HPLC. Compared to

2D HPLC, the amount of proteins identified increased dramatically using 1D-RP-HPLC and the depletion strategy. In spite of these significant improvements, the lack of suitable machinery in the form of an HPLC capable of sample fractionation and collection, prior to MS, prevented the technique being applied to quantitatively profile the *Plasmodium* proteome following drug perturbation in a truly comprehensive manner. However, on the basis of the number of proteins identified at high confidence (≥95%), the optimised SDS-PAGE / 1D-RP-HPLC protocol is ideally suited to rapidly profile proteins isolated from *Plasmodium* organelles or protein complexes of interest.

During the development of the optimised protocol, evidence emerged implicating a putative role of post-translationally (PTM) mediated mechanisms in the ability of compounds containing an alpha-methylene-gamma-lactone moiety, such as Ref#31, to inhibit eukaryotic cell growth. In particular, these mechanisms are proposed to involve ubiquitination and subsequent depletion of histone deacetylase 1 and the E3 ubiquitin-type ligases (Gopal *et al*., 2009). To determine whether these mechanisms may potentially contribute to Ref#31-induced inhibition of *P. falciparum*, the optimised SDS-PAGE / 1D-RP-HPLC protocol was applied to explore PTMs that occurred in the parasite following treatment. Analysis revealed that the *Plasmodium* proteome undergoes extensive PTM changes during the course of the parasite's development, with changes in acetylation, methylation, ubiquitination and phosphorylation being frequently observed, even during normal development. Among these changes, modifications to a putative ubiquitin E3 ligase were identified post-treatment in the form of ubiquitination of isoforms of this protein, together with expressional changes to a putative ubiquitin carboxyl-terminal hydrolase involved in this cellular process. Similar changes at the mRNA level are observed using microarray analysis in this system, together with changes associated with histone activity

(Drs. John Becker / Dalu Mancama, CSIR Systems Biology, personal communication). Thus, further investigation of these processes could potentially prove useful in establishing the molecular effects of Ref#31 in *P. falciparum*.

A significantly large amount of data was generated during this work, and the large amount of PTM data that was generated highlights the urgent need for further development of high-throughput bioinformatics tools that will enable accurate meaningful interpretation of PTM events, as these appear to be increasingly important to understanding molecular function. Combined with improvements to fractionate protein lysates of interest prior to MS/MS, advances in bioinformatics would significantly improve both the quality and quantity of the data acquired by this approach. Ultimately, this will lead to a more comprehensive and better understanding of the effects of drug perturbation on the *Plasmodium* proteome, and could contribute to the development of more effective compounds against this parasite.

5 REFERENCES

Aebersold, R. and Mann, M. 2003. Mass spectrometry-based proteomics**.** *Nature,* 422: 198-207.

Aggarwal, K., Choe, L.H., Lee, K.H. 2006. Shotgun proteomics using iTRAQ isobaric tags. *Briefings in Functional Genomics and Proteomics*, 5 (2): 112-120.

Anderson, N.L. and Anderson, N.G. 2002. The human plasma proteome: history, character, and diagnostic prospects. *Molecular and Cellular Proteomics*, 1 (11): 845-867.

Alban, A., David, S.O., Bjorkesten, L., Andersson, C., Sloge, E., Lewis, S., Currie, I. 2003. A novel experimental design for comparative two-dimensional gel analysis: two dimensional difference gel electrophoresis incorporating a pooled internal standard. *Proteomics,* 3: 36-44.

Artesunate Malaria Trial (SEAQUAMAT) group. 2005**.** Artesunate versus quinine for treatment of severe falciparum malaria: a randomised trial. *Lancet,* 366: 717-725.

Bahl, A., Brunk, B., Coppel, R.L., Crabtree, J., Diskin, S.J., Fraunholz, M.J., Grant, G.R., Gupta, D., Huestis, R.L., Kissinger, J.C., Labo, P., Li, L., McWeeney, S.K., Milgram A.J., Ross, D.S., Schug, J., Stoeckert Jr, C.J. 2002. PlasmoDB: *The Plasmodium* genome resource. An intergrated database providing tools for accessing, analysing and mapping expression and sequence data (both finished and unfinished). *Nucleic Acids Research*, 30(1): 87-90.

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

Ball, E.G., McKee, R.W., Anfinsen, C.B., Cruz, W.O., Geiman Q.M. 1948. Studies on malarial parasites: ix. Chemical and metabolic changes during growth and multiplication *in vivo* and *in vitro*. *The Journal of Biological Chemistry*, 175: 547-71

Baniecki, M.L., Wirth, D.F., Clardy, J. 2007. High-throughput *Plasmodium falciparum* growth assay for malaria drug discovery. *Antimicrobial Agents and Chemotherapy*,51 (2): 716-723.

Barrera, L., Benner, C., Tao, YC., Winzeler, E., Zhou, Y. 2004. Leveraging two-way probe-level block design for identifying differential gene expression with high-density oligonucleotide arrays. *BMC Bioinformatics*, 5: 42.

Baumann, E., Stoya, G., VdroÈlkner, A., Richter, W., Lemke, C., Linss, W. 2000. Hemolysis of human erythrocytes with saponin affects the membrane structure. *Acta Histochemica*, 102: 21-35.

Beck, H.C., Nielsen, E.C., Matthiesen, R., Jensen, L.H., Sehested, M., Fine, P., Grauslund, M., Hansen, A.M., Jensen, O.N. 2006. Quatitative proteomic analysis of post-translational modifications of human histones. *Molecular and Cellular Proteomics*, 5: 1314-1325.

Beeson, J.G., **Amin, N., Kanjala, M., Rogerson, S.J.** 2002. Selective accumulation of mature asexual stages of *Plasmodium falciparum*-infected erythrocytes in the placenta. *Infection and Immunity*, 70: 5412-5415.

Birkholtz, L., van Brummelen, A.C., Clark, K., Niemand, J., Marechal, E., Llinas, M., Louw, A.I. 2008. Exploring functional genomics for drug target and therapeutics discovery in *Plasmodia*. *Acta Tropica*, 105: 113-123.

Bloland, P.B., Kachur, S.P., Williams, H.A. 2003. Trends in antimalarial drug deployment in sub-Saharan Africa. *Journal of Experimental Biology*, 206: 3761-3769.

Braisted, J.C., Kuntumalla, S., Vogel, C., Marcotte, E.M., Rodrigues, A.R., Wang, R., Huang, S., Ferlanti, Saeed, A.I., Fleischmann, R.D., Peterson, S.N., Pieper, R. 2008. The APEX quantitative proteomics tool: Generating protein quantitation estimates from LC-MS/MS proteomics results. *Bioinformatics*, 9:529, 1-11.

Boehm, A.M., Galvin, R.P., Sickmann, A. 2004. Extractor for ESI quadrupole TOF tandem MS data enabled for high-throughput batch processing. *BMC Bioinformatics*, 5 (162): 1-5.

Boehm, A.M., Pütz, S., Altenhöfer, D., Sickmann, A., Falk, M. 2007. Precise protein quantitation based on peptide quantification using iTRAQ™. *BMC Boinformatics,* 8 (214): 1-18.

Breman, J.G., Egan, A., Keusch, G.T. 2001. Introduction and summary: The intolerable burden of malaria: a new look at the numbers. *American Journal of Tropical Medicine and Hygiene*; 64 (Suppl 1-2): iv - vii.

Calvosa, V.S., Adagu, I.S., Povoa, M.M. 2001. *Plasmodium falciparum*: emerging mefolquine resistance *in vitro* in Para State North of Brazil. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 95: 330-331**.**

Caniato, R. and Puricelli, S. 2003. Natural antimalarial agents. *Critical Reviews I Plant Sciences*, 22(1): 79- 105.

Clarkson, C., Maharaj V.J., Neil R.C., Olwen M.G. , Pillay, P., ,Motlalepula G.M., Bhagwandin, N., Smith, P.J., Folb, I.P. 2004. *In vitro* antiplasmodial activity of medicinal plants native to or naturalised in South Africa. *Journal of Ethnopharmacology*. 92: 177-191.

Collins, W.E., Jeffery, G.M., Roberts, J.M. 2003. A retrospective examination of anaemia during infection of humans with *Plasmodium vivax. American Journal of Tropical Medicine and Hygiene*, 68(4): 410-412.

Cooper, R. and Carucci, D. 2004. Proteomic approaches to studying drug targets and resistance in *Plasmodium*. *Current Drug Targets – Infectious Disorders*, 4: 41-51.

Cox, F.E.G. 2002. History of human parasitology. *Clinical Microbiology Reviews*, 15 (4): 595-612.

Crawley, J., Smith, S., Muthinji, P., Marsh, K., Kirkham, F. 2001. Electroencephalographic and clinical features of cerebral malaria. *Archives of Disease in Childhood*, 84: 247-253.

Dhingra, V., Rao K.W., Narasu, M.L. 2000. Current status of artemisinin and its derivates as antimalarial drugs. *Life Sciences*, 66: 279-300.

Dirksen, E.H.C., Kolkman, A., Slijper, M., Heck, A.J.R. 2005. Double standards in quantitative proteomics: direct comparative assessment of difference in gel electrophoresis (DiGE) and metabolic stable isotope labelling. *Molecular and Cellular Proteomics*, 4 (3): 255-266.

Djimde, A., Doumbo, O.K., Cortese, J.F., Kayentao, K., Doumbo, S., Diourte, Y., Dikco, A., Su, X-Z., Nomura, T., Fidock, D.A., Wellems, T.E., Plowe, C.V. 2001. A molecular marker for chloroquine-resistant falciparum malaria. *The New England Journal of Medicine*, 344 (4): 257-263.

Echan, L.A., Tang, H., Ali-Khan, N., Lee, K., Speicher, D.W. 2005. Depletion of multiple high-abundance proteins improves protein profiling capacities of human serum and plasma. *Proteomics*, 5: 3292-3303.

Eckstein-Ludwig, U., Webb, R.J, Van Goethem, I.D., East, J.M., Lee, A.G., Kimura, M. O'Neill, P.M., Bray, P.G., Ward, S.A., Krishna, S. 2003. Artemisinins target the SERCA of *Plasmodium falciparum*. *Nature*, 424: 887–889.

Egan, T.J. 2009. Artemisinin-resistant *Plasmodium falciparum*: can the genie be put back in the bottle. *Future Microbiology*, 4 (6): 637-639.

Elrick, M.M., Walgren, J.L., Mitchell, M.D., Thompson, D.C. 2006. Proteomics: Recent applications and new technologies. *Basic and Clinical Pharmacology and Toxicology*, 98: 432-441.

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

Foth, B.J., Zhang, N., Mok, S., Preiser, P.R., Bozdech, Z. 2008. Quantitative protein expression profiling reveals extensive post-transcriptional regulation and post-translational modifications in schizont-stage malaria parasites.*Genome Biology*, 9 (12): 1-18.

Fonrose, X., Ausseil, F., Soleilhac, E., Masson, V., David, B., Pouny, I., Cintrat, J., Rousseau, B., Barette, C., Massiot, G., Lafanechère, L. 2007. Parthenolide inhibits tubulin carboxypeptidase activity. *Cancer Research*. 67 (7): 3371-3378.

Francis, S.E., Sullivan, Jr., D.J, Goldberg, D.E. 1997. Haemoglobin metabolism in the malaria parasite *Plasmodium falciparum*. *Annual Review of Microbiology*, 51: 97-123.

Garbis, S., Lubec, G., Fountoulakis, M. 2005. Limitations of current proteomics technologies. *Journal of Chromatography A*, 1077: 1-18.

Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A., Nelson, K.E., Bowman, S., Paulsen, I.T., James, K., Eisen, J.A., Rutherford, K., Salzberg, S.L., Craig, A., Kyes, S., Chan, M., Nene, V., Shallom, S.J., Suh, B., Peterson, J., Angiuoli1, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M.W., Vaidya, A.B., Martin, D.M.A., Fairlamb, A.H., Fraunholz, M.J., Roos, D.S., Ralph, S.A., McFadden, G.I., Cummings, L.M., Subramanian, G.M., Mungall, C., Venter, J.C., Carucci, D.J., Hoffman, S.L., Newbold, C., Davis, R.W., Fraser,

C.M., Barrell, B. 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum. Nature,* 419: 498-511.

Garnham, P. 1966, Malaria parasites and other haemosporidia, Blackwell Scientific publications, Oxford, pp 1114.

Gerald, D., Schmidt, S., Larry, S., Roberts, J. 1996. Foundations of Parasitology. Fifth edition. Wm. C. Brown Publishers.

Gluckmann, M., Fella, K., Waidelich, D., Merkel, D., Kruft, V., Kramer, P., Walter, Y., Hellmana, J., Karas, M., Kroeger, M. 2007. Prevalidation of potential protein biomarkers in toxicology using iTRAQ™ reagent technology. *Proteomics*, 7, 1564–1574.

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

Goldberg, D.E., Slater, A.F.G., Beavis, R., Chait, B., Cerami, A., Henderson, B.G. 1991. Hemoglobin degradation in the human malaria pathogen *Plasmodium falciparum*: A catabolic pathway initiated by a specific aspartic protease. *Journal of Experimental Medicine*. 173: 961-969.

Gopal, Y.N.V., Arora, T.S., Van Dyke, M.W. 2007. Parthenolide specifically depletes histone deacetylase 1 protein and induces cell death through ataxia telangiectasia mutated. *Chemistry and Biology*, 14: 813-823.

Gopal, Y.N.V., Chanchorn, E., Van Dyke, M.W. 2009. Parthenolide promotes ubiquitination of MDM2 and activates p53 cellular functions. *Molecular Cancer Therapy*, 8 (3): 552-561.

Govorukhina, N.I., Keizer-Gunnink, A., Van der Zee, A.G.J., de Jong, S., de Bruijn, H.W.A., Bischoff, R. 2003. Sample preparation of human serum for the analysis of tumor markers: Comparison of different approaches for albumin and g-globulin depletion. *Journal of Chromatography A.* 1009: 171-178.

Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., Aebersold, R. 1999. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology,* 17: 994-999.

Hayton, K. and Su, X.Z. 2004. Genetic and biochemical aspects of drug resistance in malaria. *Current Drug Targets- Infectious Disorders*, 4: 1-10.

Hellman, U., Wernstedt, C., Gonez, J., Heldin C.H. 1995. Improving of an "in-gel" digestion procedure for the micropreparation of internal protein fragments for amino acid sequencing. *Analytical Biochemistry,* 224: 451-455.

Hill, D.A., Pillai, A.D., Nawaz, F., Hayton, K., Doan, L., Lisk, G., Desai, S.A. 2007. A blasticidin S-resistant *Plasmodium falciparum* mutant with a defective plasmodial surface anion channel. *PNAS,* 104 (3): 1063-1068.

Hood, B.L., Veestra, T.D., Conrads, T.P. 2004. Mass spectrometry-based proteomics. *Internation Congress Series*, 1266: 375-380.

http://www.AppliedBiosystems.com

Hunter, T.C., Andon, N.L., Koller, A., Yates, J.R., Haynes, P.A. 2002**.** The functional proteomics toolbox: methods and applications .*Journal of Chromatography. B*, 782: 165.

Hutchings, A., Scott A.H., Lewis G., Cunningham, A.B. 1996. Zulu medicinal plants, an inventory. Kohler Carton and Print. University of Natal Press Pietermaritzburg and National Botanical Institute Cape Town.

Hyde, J.E. 2007. Drug-resistant malaria- an insight. *The FEBS Journal*. 274: 4688–4698.

Jambou, R., Legrand, E., Niang, M. 2005**.** Resistance of *Plasmodium falciparum* field isolates to *in vitro* artemether and point mutations of the SERCA-type PfATPase6. *Lancet*, 366: 1960-1963.

Johnson, J.R., Florens, L., Carucci, D.J., Yates J.R. 2004. Proteomics in malaria. *Journal of Proteomic Research,* 33: 296-306.

Karapelou, J. 1987. Parasite Life Cycles, Springer Verlag, New York.

Karou, D., Dicko, M.H., Sanon, S., Simpore, J., Traore, A.S. 2003. Antimalarial activity of *Sida acuta* Burm. f. (Malvaceae) and *Pterocarpus erinaceus* Poir. (Fabaceae). *Journal of Ethnophamacology*, 89: 291-294.

Keller, A., Eng, J., Zhang, N., Li, X.J., Aebersold, R. 2005. A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Molecular Systems Biology*, 1:.1-8.

Kombila, M., Duong, T.H., Dufillot, D. 1997. Light microscopic changes in *Plasmodium falciparum* from Gabonese children treated with Artemether. *American Journal of Tropical Medicine and Hygiene*, 57 (6): 643-645.

Kremsner, P.G. and Krishna, **S.** 2004. Antimalarial combinations. *Lancet*, 364: 285-294.

Krettli, A.U., Valter, F., Andrade-Neto, V.F., Brandão, M.G.L., Ferrari, W.M.S. 2001. The search for new antimalarial drugs from plants used to treat fever and malaria or plants ramdomly selected: a Review. *Memorias do Instituto Oswaldo Cruz, Rio de Janeiro*, 96(8): 1033-1042.

LaCount, D.J., Vignali, M., Chettier, R., Phansalkar, A., Bell, R., Hesselberth, J.R., Schoenfeld, L.W., Ota, I., Sahasrabudhe, S., Kurschner, C., Fields, S., Hughes, R.E. 2005. A protein interaction network of the malaria parasite *Plasmodium falciparum. Nature*, 438: 103-229.

Lamarque, M., Tastet, C., Poncet, J., Demettre, E., Jouin, P., Vial1, H., Dubremetz1, J. 2008. Food vacuole proteome of the malarial parasite *Plasmodium falciparum. Journal of Clinical Application of Proteomics*, 2: 1361-1374.

Levine, J., Weickert, M., Pagratis, M., Etter, J., Mathews, A., Fattor, T., Lippincott, J., Apostol, I. 1998. Identification of a nickel (II) binding site on hemoglobin which confers susceptibility to oxidative deamination and intramolecular cross-linking. *Journal of Biological Chemistry*. 273 (21): 13037-13046.

Makanga, M., Bray1, P.G., Horrocks, P., Ward, S.A. 2005. Towards a proteomic definition of CoArtem action in *Plasmodium falciparum* malaria. *Proteomics*, 5: 1849-1858.

Mariño-Ramírez, L., Kann, M.G., Shoemaker, B.A., Landsman, D. 2005. Histone structure and nucleosome stability. *Expert Review of Proteomics*, 2(5): 719-729.

Marouga, R., David, S., Hawkins, E. 2005.The development of the DIGE system: 2D fluorescence difference gel analysis technology. *Analytical and Bioanalytical Chemistry*, 382: 669-678.

Medana, I.M. and Turner, G.D.H. 2006. Human cerebral malaria and the blood–brain barrier. *International Journal for Parasitology*, 36: 555-568.

Menendez, C., Fleming, A.F., Alonso, P.L. 2000. Malaria-related anaemia. *Parasitology Today*, 16 (11): 469-476.

Milijaona, R., Jambou, R., Raharimalala, L., Ranaivo, L., Ranson, M.A., Roux, J. 2000. Mefloquine-resistant strains of *Plasmodium falciparum* in Madagascar: impact on tourists and public health. *Annals of Tropical Medicine and Parasitology*, 94: 313-317.

Miller, L.H., Good, M.F., Milon, G. 1994. Malaria pathogenesis. *Science*, 264: 1878-1883.

Mishra, S. K., Mohanty, S., Mohanty, A., Das, D.B. 2006. Management of severe and complicated malaria. *Journal of Postgraduate Medicine*,52 (4): 281-287.

Mitulovic, G., Stingl, C., Smoluch, M., Swart, R., Chervet, J., Steinmacher, I., Gerner, C., Mechtler, K. 2004. Automated, on-line two-dimensional nano-liquid chromatography tandem mass spectrometry for rapid analysis of complex protein digests. *Proteomics*, 4: 2545-2557.

Mnengwane, J. and Koekemoer, M. 2007. National Herbarium, Pretoria.

Nirmalan, N., Sims, P.F.G., Hyde, J.E. 2004. Quantitative proteomics of the human malaria parasite *Plasmodium falciparum* and its application to studies of development and inhibition. *Molecular Microbiology*. 52 (4): 1187-1199.

Nosten, F. and Brasseur, P. 2002. Combination therapy for malaria: the way forward? *Drugs*, 62: 1315-1329.

Ogutu, B.R., Smoak, B.L., Nduati, R.W. 2000. The efficacy of pyrimethane-sulfadoxine (Fansidar) in the treatment of uncomplicated *Plasmodium falciparum* malaria in Kenyan children. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 94(1): 83-94.

Olliaro, P.L and Taylor, W.R. 2003. Antimalarial compounds: from bench to bedside *Journal of Experimental Biology*, 206: 3753-3759.

Payne, D. 1987. Spread of chloroquine resistance in *Plasmodium falciparum*.*Parasitology Today*, 3: 241-246.

Peng, Z., Elias, J.E., Thoreen, C.C., Licklider, L.J., Gygi, S.P. 2003. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: The yeast proteome. *Journal of Proteome Research*, 2: 43-50.

Petri, W.A. Jr. 2003. Can a proteomics strategy be used to identify the anti-malarial activity of chloroquine? *Trends in Pharmacological Sciences*, 24 (5): 210-211.

Philip, J. and Rosenthal, M.D. 2008. Artesunate for the treatment of severe falciparum malaria. *The New England Journa l of Medicine*, 358: 1829-1836.

Pink, R., Hudson, A., Mouries, M., Bendig, M. 2005. Opportunies and challenges in antiparasitic drug discovery. *Nature Reviews Drug Discovery*, 4: 727-740.

Plowe, C.V. 2003. Monitoring antimalarial drug resistance: making the most of the tools at hand. *Journal of Experimental Biology*, 206: 3745-3752.

Price, R.N., Uhlemann, A.C., Brockman, A., McGready, R., Ashley, E., Phaipun, L. 2004. Mefloquine resistance in *Plasmodium falciparum* and increased *pfmdr 1*. *Lancet,* 364: 438-447.

Prieto, J.H., Koncarevic, S., Park, S.K., Yates, J., Becker, K. 2008. Large-scale differential proteome analysis in *Plasmodium falciparum* under drug treatment. *PLoS ONE,* 3 (12): 1-9.

Prozesky, E.A., Meyer, J.J.M., Louw, A.I. 2001. *In vitro* antiplasmodial activity and cytotoxicity of ethnobotanically selected South African plants. *Journal of Ethnopharmacology*, 76: 239-245.

Reacher, M., Campbell, C.C., Freeman, J., Doberstyn, E.B., Brandling-Bennet, A.D. 1981. Drug therapy for *Plasmodium falciparum* malaria resistant topyrimethamine-sulfadoxine (Fansidar). A study of alternate regimens in Eastern Thailand. *Lancet,* 2: 1066-1069.

Rief, M., Pascual, J., Saraste, M., Gaub, H.E. 1999. Single molecule force spectroscopy of spectrin repeats: Low unfolding forces in helix bundles. *Journal of Molecular Biology,* 286: 553-561.

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

Righetti, P.G., Campostrini, N., Pascali, J., Hamdan, M. Astner, H. 2004.Quantitative proteomics: a review of different methodologies. *European Journal of Mass Spectrometry,* 10: 335-348.

Ringrose, J.H., Van Solinge, W.W., Mohammed, S., O'Flaherty, M.C., Van Wijk, R. Heck, A.J.R., Slijper, M. 2008. Highly efficient depletion strategy for the two most abundant erythrocyte soluble proteins improves proteome coverage dramatically. *Journal of Proteome Research,* 7: 3060-3063.

Roberts, M. 1990. Indigenous healing plants. Southern Book Publishers. 1st Ed. Cape Town.

Rohila, J.S., Chen, M., Chen, S., Chen, J., Cerny, R.L., Dardick, C., Canlas, P., Fujii, H., Gribskov, M., Kanrar, S., Knoflicek, L., Stevenson, B., Xie, M., Xia Xu, X., Zheng, X., Zhu, J., Ronald, P., Fromm, M.E. 2009. Protein-protein interactions of tandem affinity purified protein kinases from rice. *PLoS ONE*, 4: 1-8

Rosenthal, P.J. and Miller, L.H. In *Antimalarial Chemotherapy. Mechanisms of Action, Resistance and New Direction in Drug Discovery*; Rosenthal, P.J., Ed Humana Press Inc.: Totowa NJ, 2001, pp 3-13.

Sachs, J. and Malaney, P. 2002. Economic and social burden of malaria. *Nature*, 415: 680-684.

Schaechter, M. 1993. Mechanisms of Microbial Disease, 2nd edition, Williams & Wilkins, London, pp 973.

Schofield, L. and Grau, G.E. 2005. Immunological processes in malaria pathogenesis. *Nature*, 5: 722-735.

Sibley, C.H., Hyde, J.E., Sims, P.F.G., Plowe, C.V., Kublin, J.G., Mberu, E.K., Cowman, A.F., Winstanley, P.A., Watkins, W.M., Nzila, A.M. 2001. Pyrimethamine–sulfadoxine resistance in *Plasmodium falciparum*: what next? *Trends in Parasitology*, 17 (12): 582-587.

Shadforth, I.P., Dunkley, T.P.J., Lilley, K.S., Bessant, C. 2005. i-Tracker: For quantitative proteomics using iTRAQ™. *BMC Genomics*, 6 (145): 1-6.

Sims, P.F.G. and Hyde, J.E. 2006. Proteomics of the human malaria parasite *Plasmodium falciparum*. *Expert Reviews of Proteomics*. 3(1): 87-95.

Song, X., Bandow, J., Sherman, J., Baker, J.D., Brown, P.W., McDowell, M.T., Molloy, M.P. 2008. iTRAQ experimental design for plasma biomarker discovery. *Journal of Proteome Research*, 7: 2952-2958.

Stevenson, M., Zhong, S., Hakeem, S., Karkada, M. 2001. Modulation of host responses to blood-stage malaria by interleukin-12: from therapy to adjuvant activity. *Journal of Microbes and Infection*, 3: 49-59.

Stoeckert, C.J., Fischer, J.S., Kissinger, J.C., Heiges, M., Aurrecoechea, C., Gajria, C., Roos, D.S.,Aurrecoechea, C., Gajria, B. 2006. PlasmoDB v5: new looks, new genome. *Trends in Parasitology*, 22(12): 543-546.

Thelen, J.J. 2007. Quantitative proteomics in plants: Choices in abundance. *The Plant Cell*, 19: 3339-3346.

Thien, H.V., Piet, A. Kager, P.A., Sauerwein, **H.P.** 2006. Hypoglycemia in falciparum malaria: is fasting an unrecognized and insufficiently emphasized risk factor? *Trends in Parasitology*, 22 (9): 4120-4125.

Tonge, R., Shaw, J., Middleton, B., Rowlinson, R., Rayner, S., Young, J., Pognan, F., Hawkins, E., Currie, I., and Davison, M. 2001. Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics*, 1: 377-396.

Toovey, S. and Jamieson, A. 2004. Audiometric changes associated with the treatment of uncomplicated falciparum malaria with co-artemether. *Transactions of the Royal Society of Tropical Medicine and Hygiene.* 98: 261-267.

Tran, Q.L., Tezuka, Y., Ueda, J., Nguyen, N.T., Maruyama, Y., Begum, K., Kim, H., Wataya, Y., Tran, Q.K., Kadota, S. 2003. *In vitro* antiplasmodial activity of antimalarial medicinal plants used in Vietnamese traditional medicine. *Journal of Ethnopharmacology* 86: 249-252.

Unlu, M., Morgan, M.E., Minden, J.S. 1997. Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis*, 18: 2071-2077.

Unwin, R.D. 2006. Quantitative proteomics reveals posttranslational control as a regulatory factor in primary hematopoietic stem cells. *Blood*, 107: 4687-4694.

Unwin, R.D., Pierce, A., Watson, R.B., Sternberg, D.W., Whetton, A.D. 2005. Quantitative proteomic analysis using isobaric protein tags enables rapid comparison of changes in transcript and protein levels in transformed cells. *Molecular and Cellular Proteomics*, 4: 924–935.

Van Wyk, B.E. 2002. A review of ethnobotanical research in Southern Africa. South African Journal of Botany. 68: 1–13.

van Agtmael, M.A., Eggelte, T.A., van Boxtel, C.J. 1999. Artemisinin drugs in the treatment of malaria: from medicinal herb to registered medication. *Trends in Pharmacological Sciences*, 20: 199-204.

Vedadi, M., Lew, J., Artz, J., Amani, M., Zhao, Y., Dong, A., Wasney, G.A., Gao, M., Hills, T. 2007. Genome-scale protein expression and structural biology of *Plasmodium falciparun* and related Apicomplexan organisms. *Molecular and Biochemical Parasitology*, 151: 100-110.

Vollmer, M., Nägele, E., Hörth, P. 2003. Differential proteome analysis: Two-dimensional nano-LC/MS of *E. coli* proteome grown on different carbon sources. *Journal of Biomolecular Techniques*, 14: 128-135.

Washburn, M.P., Wolters, D., Yates, J.R., 3rd. 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology*, 19: 242–247.

Watt, J.M. and Breyer-Brandwijk, M.G. 1962. The Medicinal and Poisonous Plants of Southern and Eastern Africa. E. & S. Livingstone Ltd., Edinburgh, UK.

Wiese, S., Reidegeld, K.A., Meyer, H.E., Warscheid, B. 2007. Protein labelling by iTraq: A new tool for quantitative mass spectrometry in proteome research. *Proteomics*, 7: 340-350.

Wellems, T.E. and Plowe, C.V. 2001. Chloroquine-resistance malaria. *The Journal of Infectious Diseases*, 184: 770-776.

White, N.J. 2004. Antimalarial drug resistance. *The Journal of Clinical Investigation*, 8 (113): 1084-1092.

White, N.J. 2003. Malaria, In Manson's Tropical Diseases (21st edn) (Cook, G.C. and Zumla, A., eds), pp. 1205–1295, W.B. Saunders.

Witze, E.S., Old, W.M., Resing, K.A., Ahn, N.G. 2007. Mapping protein post-translational modifications with mass spectrometry. *Nature*, 410, 798-806.

Wongsrichanalai, C., and Meshnick, S.R. 2008. Declining artesunate-mefloquine efficacy against falciparum malaria on the Cambodia–Thailand border. *Emerging Infectious Diseases*, 14(5): 716-719.

World Health Organisation. 2008. Malaria world report.

World Health Organisation. Malaria Fact Sheet. 2007. World Health Organisation Fact Sheet: Number 94.

World Health Organization. 2000. Severe falciparum malaria. *Transactions of the Royal Society of Tropical Medicine and Hygiene*.94 (Suppl. 1), S1–S90.

World Health Report. 1993-2000. Implementation of the global malaria control strategy.

Woodrow, C.J., Haynes, R.K., Krishna, S. 2005. Artemisinins. *Journal of Postgraduate Medicine*, 81: 71-78.

Wu, K.P., Lin, W.T., Hung, W.N., Yian, Y.H., Chen, Y.R., Chen. Y.J., Sung, T.Y., Hsu, W.L. 2005. MassTRAQ: A fully automated tool for iTRAQ labeled protein quantification. Stanford, USA. Edited by: Martin DC. IEEE Computer Society, pp 157-158.

Yan, J. X., Harry, R. A., Spibey, C., Dunn, M. J. 2000. Post electrophoretic staining of proteins separated by two-dimensional gel electrophoresis using SYPRO dyes. *Electrophoresis*,21:3657-3665.

Zieske, L.R. 2006. A perspective on the use of iTRAQ reagent technology for protein complex and profiling studies. *Journal of Experimental Botany*, 57 (7): 1501-1508.

Zolotarjova, N., Martosella, J., Nicol, G., Bailey, J., Boyes, B.E., Barrett, W.C. 2005. Differences among techniques for high-abundant protein depletion. *Proteomics*, 5: 3304-3