Proteomic Analysis Delineates the Signaling Networks of Plasmodium falciparum

Brittany Pease
University of Central Florida

Part of the Medical Sciences Commons
Find similar works at: https://stars.library.ucf.edu/etd

This Doctoral Dissertation (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations, 2004-2019 by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

STARS Citation
https://stars.library.ucf.edu/etd/1243
PROTEOMIC ANALYSIS DELINEATES THE SIGNALING NETWORKS OF *PLASMODIUM FALCIPARUM*

by

BRITTANY NICOLE PEASE
M.S. University of Central Florida, 2011
B.S. Rollins College, 2008

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Biomedical Sciences in the Burnett School of Biomedical Sciences in the College of Medicine at the University of Central Florida Orlando, Florida

Summer Term
2015

Major Professor: Debopam Chakrabarti
ABSTRACT

Malaria is a life-threatening disease caused by Plasmodium parasites that are spread through the bites of infected mosquito vectors. It is a worldwide pandemic that threatens 3.4 billion people annually. Currently, there are only a few validated Plasmodium drug targets, while drug resistance continues to rise. This marks the urgency for the development of novel parasite-specific therapeutics. Plasmodium falciparum diverges from the paradigm of the eukaryotic cell cycle by undergoing multiple rounds of DNA replication and nuclear division without cytokinesis. A better understanding of the molecular switches that coordinate the progression of the parasite through the intraerythrocytic developmental stages will be of fundamental importance for the design of rational intervention strategies.

To achieve this goal, we performed an isobaric tag-based approach for a system-wide quantitative analysis of protein expression and site-specific phosphorylation events of the Plasmodium asexual developmental cycle in the red blood cells. This study identified 2,767 proteins, 1,337 phosphoproteins, and 6,293 phosphorylation sites. Approximately 34% of identified proteins and 75% of phosphorylation sites exhibit changes in abundance as the intraerythrocytic cycle progresses.

Because the links between Plasmodium protein kinases as key cell cycle regulators to cellular events are largely unknown, it is of importance to define their cognate physiological substrates. To test the hypothesis that genetic screening would be a useful approach for discovery of candidate substrates of a protein kinase, we used the orphan kinase PfPK7 as a model. Our comparison of the phosphoproteome profiles between the wild-type 3D7 and PfPK7− parasites identified 146 proteins with 239 phosphorylation sites exhibiting decreased phosphorylation in the
absence of PfPK7 at the developmental stages where nuclear division and merozoite formation occur. Further analysis of the decreased phosphorylated events revealed three motifs that are enriched among phosphorylated sites in proteins that are down regulated. In vitro kinase assays were done to validate the potential substrates of PfPK7 and to elucidate the signaling events that are regulated by PfPK7.

In parallel to our experimental analysis, we used a computational approach for substrate prediction from our phosphoproteome dataset. This analysis identified 43 distinct phosphorylation motifs and a range of proline-directed potential MAPK/CDK substrates. To identify substrates/interactors of Plasmodium CDK-like kinases, we also used HA-tagged CDK-like kinases, PfPK6 and Pfmrk lines. Co-immunoprecipitation of the HA-tagged PfPK6 and Pfmrk baits, followed by mass spectrometric analyses, identified the components of the protein interaction complexes of these kinases. Our analyses of HA-PfPK6 and HA-Pfmrk immunoprecipitates identified 15 and 21 proteins in the interaction complex, respectively. The ability of recombinant PfPK6 and Pfmrk to interact and/or utilize any of the proteins identified in the interaction complex as substrates was verified through in vitro kinase assays and pull-down analysis.

This study is the most comprehensive definition of the constitutive and regulated expression of the Plasmodium proteome during the intraerythrocytic developmental cycle, and offered an insight into the dynamics of phosphorylation during the asexual cycle progression [1]. In summary, this study has 1) defined the constitutive and regulated expression of the Plasmodium proteome during its asexual life cycle, 2) demonstrated that fluctuation and reversible phosphorylation is important for the regulation of P. falciparum’s unique cell cycle, 3) provided the foundation for quantitative phosphoproteomic analysis of kinase negative mutants to understand their function, 4)
provided a major step towards defining kinase-substrate pairs operative within parasite’s signaling networks, and 5) generated a preliminary interactome for PfPK6.
ACKNOWLEDGMENTS

First and foremost I would like to thank my advisor Dr. Debopam Chakrabarti for his guidance and constant support during my training. His determination and drive for excellence and perfection was a huge motivating factor behind my performance in the laboratory. He has contributed greatly to my professional development and molded me into the researcher I am today.

Additionally I would like to thank all of my committee members, Dr. Alex Cole, Dr. Annette Khaled, Dr. Mollie Jewett, and Dr. Ratna Chakrabarti. The insightful comments, advice, and suggestions provided by my committee were critical in influencing the work presented in my dissertation.

Furthermore I am grateful to all of the current and past members of Dr. Debopam Chakrabarti’s lab, whose support helped guide me through all of my lab work.

Special thanks goes to Dr. Christian Doerig of Monash University for material support during this study and the generation of transfected cell lines. I would also like to thank Dr. Steven Gygi and Dr. Edward Huttlin of Harvard University for their technical support during this study and their expertise in mass spectrometric based proteomics analyses. I would also like to thank the American Chemical Society for publishing my work Global Analysis of Protein Expression and Phosphorylation of Three Stages of Plasmodium falciparum Intraerythrocytic Development in the Journal of Proteome Research 2013 12(9): p. 4028-4045.

I am also very thankful to my family and friends for their love, encouragement, and support throughout my academic career. Specifically I would like to thank my mother Kimberly Cox who provided unconditional support and encouragement, which kept me going throughout my dissertation work.
TABLE OF CONTENTS

LIST OF FIGURES ............................................................................................................................. xii

LIST OF TABLES ................................................................................................................................. xiv

LIST OF ACRONYMS ............................................................................................................................ xv

CHAPTER 1: INTRODUCTION .................................................................................................................. 1

1.1 Hypothesis and Specific Objectives ............................................................................................... 1

1.2 Literature Review ............................................................................................................................ 2

1.2.1 Plasmodium falciparum: The Most Virulent Form of Malaria ................................................. 2

1.2.2 Unique Life Cycles of Malaria Parasites .................................................................................. 4

1.2.3 Identification of the P. falciparum Kinome .............................................................................. 7

1.2.4 P. falciparum Protein Kinase 6 .................................................................................................. 11

1.2.5 P. falciparum MO15-related Protein Kinase Pfmrk .................................................................. 12

1.2.6 P. falciparum Atypical Protein Kinase 7 .................................................................................. 13

1.2.7 Current Technologies used to Identify Protein Kinase Substrates ......................................... 14

1.2.8 Quantitative Mass Spectrometry in Proteomics .................................................................... 16

1.2.9 Phosphoproteomic Analysis .................................................................................................... 20

1.2.10 The P. falciparum Proteome and Phosphoproteome ............................................................... 24

CHAPTER 2: GLOBAL ANALYSIS OF PROTEIN EXPRESSION AND PHOSPHORYLATION

OF THREE STAGES OF PLASMODIUM FALCIPARUM INTRAERYTHROCYTIC

DEVELOPMENT .................................................................................................................................... 27

2.1 Summary ....................................................................................................................................... 27

2.2 Materials and Methods ................................................................................................................. 28

vii
2.2.1 Plasmodium falciparum Culture ................................................................. 28
2.2.2 Sample Preparation and Mass Spectrometry .............................................. 29
2.2.3 Phosphopeptide Enrichment and Identification ........................................... 30
2.2.4 Western Blotting with Phospho-Specific Antibodies .................................. 31
2.2.5 Immunoprecipitation with PTMScan Direct ............................................... 32
2.2.6 Subdomain Location of Phosphorylation Sites in Protein Kinases .................. 34
2.2.7 Matching Phosphorylation Sites with Kinases Using NetPhorest .................. 34
2.3 Results and Discussion .................................................................................... 35
  2.3.1 Identification of 2,767 Proteins and 6,293 Phosphorylation Sites from the
  Intraerythrocytic Stages of Plasmodium falciparum ........................................... 35
  2.3.2 Developmental Stage-Specific Distribution of Proteins ................................. 37
  2.3.3 Stage-Specific Analysis of Functional Profiles .......................................... 38
  2.3.4 Developmental Stage-Specific Distribution of Phosphorylation Sites ............ 40
  2.3.5 Distribution of Phosphorylation Site Classes and Phosphorylation Motifs Across P.
  falciparum Stages and Phosphoproteins ............................................................ 43
  2.3.6 Phosphorylation Profiles of Putative MAPK/CDK Substrates and Tyrosine
  Phosphorylated Proteins .................................................................................. 44
  2.3.7 Sub-domain Distribution of Phosphorylation Sites within Protein Kinases .......... 47
  2.3.8 Preliminary Identification of Kinases Responsible for Kinase Domain Phosphorylation 49
  2.3.9 Peak Protein Expression and Phosphorylation are Unlinked ........................ 50
2.4 Conclusions ..................................................................................................... 52

CHAPTER 3: PHOSPHOPROTEOMIC ANALYSIS OF PLASMODIUM FALCIPARUM
REVEALS SUBSTRATES OF THE ATYPICAL KINASE PFPK7 ................................. 75
3.1 Summary .......................................................................................................................... 75
3.2 Materials and Methods................................................................................................. 76
  3.2.1 Plasmodium falciparum Culture .............................................................................. 76
  3.2.2 PfPK7 Verification through Southern Blot Analysis and PCR Analysis .............. 77
  3.2.3 Phosphopeptide Enrichment and Identification .................................................... 78
  3.2.4 Identification of PfPK7 Substrate Phosphorylation Motifs .................................. 79
  3.2.5 Assignment as a Potential Substrate and Generation of Expression Constructs .... 79
  3.2.6 Bacterial Expression and Purification of Recombinant Proteins ......................... 80
  3.2.7 In vitro Kinase Assays ............................................................................................. 81
3.3 Results and Discussion ............................................................................................... 81
  3.3.1 Verification of the Disruption of the PfPK7 Gene .................................................. 81
  3.3.2 Identification of 1,047 Phosphoproteins and 3,875 Phosphorylation Sites from the
      Schizont and Segmenter Intraerythrocytic Stages of Plasmodium falciparum .......... 82
  3.3.3 Analysis of Decreased Phosphorylation Events and the Identification of Unique
      Phosphorylation Motifs ............................................................................................... 84
  3.3.4 Identification and Characterization of Putative PfPK7 Substrates ....................... 86
  3.3.5 Revealing the Role of PfPK7 in Plasmodium falciparum Signaling Pathways .......... 88
3.4 Conclusions .................................................................................................................. 89

CHAPTER 4: ELUCIDATING THE FUNCTIONS OF PLASMODIUM FALCIPARUM PROTEIN
KINASES THROUGH SUBSTRATE/INTERACTOR CHARACTERIZATION .......... 105

4.1 Summary ....................................................................................................................... 105
4.2 Materials and Methods ............................................................................................... 106
  4.2.1 Plasmodium falciparum Culture ............................................................................ 106
4.2.2 Previous Global Phosphoproteomic Analysis Identifies Putative MAPK/CDK Substrates ................................................................. 107

4.2.3 Potential Substrate Expression Constructs, Recombinant Protein Expression, and Purification .......................................................................................................................... 108

4.2.4 In vitro Kinase Assays ........................................................................................................................................................................... 109

4.2.5 In vitro Kinase Assays to Determine Key Residues for PfPK6 Activity ................................................................. 109

4.2.6 Western Blot Analysis ........................................................................................................................................................................... 110

4.2.7 Immunofluorescence Assays .................................................................................................................................................................... 111

4.2.8 Co-Immunoprecipitation to Identify Interacting Partners ............................................................................................................. 112

4.2.9 Over Expression of Recombinant Interactors Using TnT SP6 Wheat Germ Expression System ........................................................................................................................................... 117

4.2.10 Validation of the PfPK6 Interactome through Pull-Down Analysis ............................................................................................ 118

4.3 Results and Discussion ............................................................................................................................................................................. 119

4.3.1 Our Global Phosphoproteomic Analysis Identifies Putative MAPK/CDK Substrates .. 119

4.3.2 Characterization of PfPK6 and Identification of Novel PfPK6 Substrates ............ 120

4.3.3 Key T-loop Residues Greatly Affect PfPK6 Auto-phosphorylation and Activity Towards Substrates ........................................................................................................................................... 122

4.3.4 Co-Immunoprecipitation Identifies the Interactome of PfPK6 and Pfmrk .......... 123

4.4 Conclusions ................................................................................................................................................................................................. 125

CHAPTER 5: GENERAL DISCUSSIONS AND CONCLUSIONS ................................................. 140

5.1 The Plasmodium falciparum Proteome and Phosphoproteome ................................................................. 141

5.2 The Use of Kinase Negative Phosphoproteomics to Delineate the Role of A Protein Kinase in Plasmodium ........................................................................................................................................... 145
5.3 Elucidating the Function of \textit{Plasmodium falciparum} Protein Kinases Through Substrate/Interactor Characterization

LIST OF REFERENCES

146

148
LIST OF FIGURES

Figure 1: The Malaria Parasite Life Cycle ............................................................... 5
Figure 2: Quantitative Mass Spectrometry Data Processing and Analysis Workflow .......... 19
Figure 3: Typical Quantitative Phosphoproteomics Workflow ...................................... 23
Figure 4: Identification of the Phosphoproteome and Proteome from the Intraerythrocytic Stages of Plasmodium falciparum ................................................................. 54
Figure 5: Distribution of Identified Human vs. Malaria Proteins and Phosphorylation Sites .... 56
Figure 6: Comparative Analysis of Stage-Specific Protein Expression ............................. 60
Figure 7: Stage-Specific Analysis of Functional Profiles .............................................. 61
Figure 8: Functional Profile of Hypothetical Proteins ................................................... 62
Figure 9: Gene Ontology Term Analysis of Intraerythrocytic Stage .................................. 63
Figure 10: Global Analysis of Stage-Specific Phosphorylation Patterns ............................ 64
Figure 11: Stage-Specific Distribution of Phosphorylation Site Classes and Motifs .......... 66
Figure 12: Phosphorylation Profiles of Putative MAPK/CDK Substrates and Tyrosine Phosphorylated Proteins .................................................................................. 68
Figure 13: Subdomain Location of Phosphorylation Sites in Protein Kinases ...................... 69
Figure 14: Associating Kinase Subdomain Phosphorylation Events with Probable Kinases .... 71
Figure 15: Comparison of Protein and Phosphoprotein Expression Across P. falciparum Intraerythrocytic Stages .................................................................................. 73
Figure 16: K-Medoids Cluster Analysis of Differential Proteins ...................................... 74
Figure 17: Verification of the Presence of PfPK7 in the Wild Type 3D7 Cell Line and the Absence of PfPK7 in the PfPK7 Cell Line ........................................................................ 91
Figure 18: Representative Images of 3D7 Wild Type and PfPK7- Time Points Harvested for Phosphoproteomic Analysis ................................................................. 92

Figure 19: Overview of the Procedure for Phosphopeptide Preparation and Analysis ......................... 93

Figure 20: Identification of the Phosphoproteome from the Schizont and Segmenter Intraerythrocytic Stages in the Presence and Absence of PfPK7 ........................................... 95

Figure 21: Examination of Decreased Phosphorylation Events in the PfPK7- Cell Line ...................... 96

Figure 22: Validation of Putative PfPK7 Substrates ........................................................................ 102

Figure 23: PfPK7's Potential Regulatory Role in Plasmodium Signaling Networks ......................... 104

Figure 24: PfPK6 Localization and Expression Pattern ..................................................................... 128

Figure 25: Identification of Novel Substrates of PfPK6 ................................................................. 130

Figure 26: Key T-loop Residues Greatly Affect PfPK6 Auto-Phosphorylation and Activity Towards Substrates ......................................................................................... 132

Figure 27: Identification of the PfPK6 Interactome ....................................................................... 136

Figure 28: Verification of the PfPK6 Interactome .......................................................................... 138

Figure 29: Preliminary Interaction Network of PfPK6 .................................................................... 139
LIST OF TABLES

Table 1: Quantification of Stage-Specific (A) Proteins and (B) Phosphorylation Sites ................. 57
Table 2: Quantification of the Top Five Decreased Phosphorylation Events in the PfPK7' Cell Line ................................................................. 98
Table 3: Identification of Putative PfPK7 Substrates ..................................................................... 100
Table 4: Identification of Putative Plasmodium CDK Substrates .............................................. 129
Table 5: Identification of the (A) Pfmrk and (B) PfPK6 Interactomes ......................................... 134
LIST OF ACRONYMS

- **A**: Alanine
- **ACN**: Acetonitrile
- **ACT**: Artemisinin based combination treatments
- **ANOVA**: Analysis of variance
- **Arg**: Arginine
- **Asn**: Asparagine
- **ATP**: Adenosine-triphosphate
- **CAK**: CDK activating kinase
- **CaMK**: Ca\(^{2+}\)- calmodulin kinase
- **CDC2**: Cell division control protein 2
- **CDK**: Cyclin dependent kinase
- **CK1**: Casein kinase 1
- **CK2**: Casein kinase 2
- **CLK**: CDC-like kinase
- **CLK2**: CDC-like kinase 2
- **Crk-1**: CDC-related protein kinase 1
- **Crk-3**: CDC-related protein kinase 3
- **Crk-4**: CDC-related protein kinase 4
- **D**: Aspartic Acid
- **DAPI**: 4’,6-diamidino-2-phenylindole
- **E**: Glutamic Acid
- **ePKs**: Eukaryotic protein kinases
- **FA**: Formic acid
- **FASP**: Filter assisted sample preparation
- **FDR**: False discovery rate
- **GPS**: Group-based Prediction System
- **GSK-3**: Glycogen-synthase kinase 3
- **HA**: Hemagglutinin
- **HPI**: Hours post infection
- **HPLC**: High performance liquid chromatography
- **ICAT**: Isotope-coded affinity tag
- **ICPL**: Isotope-coded protein label
- **IDA**: Iminodiacetic acid
- **IFA**: Immunofluorescence Assay
- **IMAC**: Immobilized metal ion affinity
- **IP**: Immunoprecipitation
- **IPI**: International Protein Index
- **iTRAQ**: Isobaric tag for relative and absolute quantification
- **MAP2K3**: Mitogen activated protein kinase 3
- **MAPK**: Mitogen-activated protein kinase
- **MBP**: Myelin basic protein
- **MIPS**: Munich Information Center for Protein Sequences
- **MS**: Mass spectrometry
- **MudPIT**: Multidimensional protein identification
- **NCBI**: National Center for Biotechnology Information
- **ORC**: Origin recognition complex
- **ORF**: Open reading frame
- **P**: Proline
- **P. falciparum**: *Plasmodium falciparum*
- **P. ovale**: *Plasmodium ovale*
- **P. vivax**: *Plasmodium vivax*
- **PBS**: Phosphate buffered saline
- **PCR**: Polymerase chain reaction
- **PfCK1**: *Plasmodium falciparum* casein kinase 1
- **Pfccyc1**: *Plasmodium falciparum* cyclin 1
- **PfMAT1**: *Plasmodium falciparum* MAT1 (CDK-activating kinase assembly factor)
- **PfMrk**: *Plasmodium falciparum* MO15-related protein kinase
- **PfPK6**: *Plasmodium falciparum* Protein Kinase 6
- **PfPK2**: *Plasmodium falciparum* Protein kinase 2
- **PfPKB**: *Plasmodium falciparum* Protein kinase B
- **PKs**: Protein kinases
- **Pro**: Proline
- **PTMs**: Post-translational modifications
- **R**: Arginine
- **RBC**: Red blood cell
- **S**: Serine
- **Ser**: Serine

xvii
- **SCX**: Strong cation exchange
- **SDS-PAGE**: Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
- **SILAC**: Stable isotope labeling by amino acids in cell culture
- **SR**: Serine/arginine proteins
- **T**: Threonine
- **TBS**: Tris buffered saline
- **TBS-T**: Tris buffered saline with Tween 20
- **Thr**: Threonine
- **TMT**: Tandem mass tag
- **Tyr**: Tyrosine
- **WT**: Wild Type
- **W/V**: Weight per Volume
CHAPTER 1: INTRODUCTION

1.1 Hypothesis and Specific Objectives

Regulation of the cell cycle is critical for maintaining cellular homeostasis. This is primarily carried out by protein kinases, which act as regulatory switches in the cell cycle by phosphorylating key proteins. Cyclin-dependent kinases (CDKs) and other protein kinases (PKs) have been shown to be master regulators of the eukaryotic cell cycle [2] and are proven therapeutic targets for many human diseases. Many of the Plasmodium kinases are structurally and functionally distinct from host kinases, which would avoid potential toxicity issues [3]. Although a number of CDK-like kinases and other PKs have been identified in Plasmodium falciparum, the molecular mechanisms underlying the regulation of the parasite’s intraerythrocytic cell division, causing the parasite cell cycle to deviate from the canonical eukaryotic cell cycle remain largely unknown. To decipher the exact role of PKs in regulating malaria parasite cell growth and differentiation, it is important to identify their specific cellular targets and their regulators. We hypothesize that global evaluation of protein phosphorylation events of wild-type and kinase deficient mutant parasites, and characterization of interactors and substrates of PKs will provide an understanding of the signaling pathways that orchestrate the P. falciparum intraerythrocytic cell cycle.

To prove this hypothesis, we pursued the following specific objectives:

1) **Global Analysis of Protein Expression and Phosphorylation of P. falciparum Intraerythrocytic Developmental Stages.** This specific aim investigated the dynamics of protein expression and phosphorylation during the entire intraerythrocytic developmental cycle of Plasmodium to reveal the constitutive versus regulated
expression of the proteome. Furthermore, distinct phosphorylation motifs utilized by *Plasmodium* kinases and potential substrates of MAPKs/CDKs were identified.

2) **Phosphoproteomic Analysis of *P. falciparum* to Reveal Substrates of the Atypical Kinase PfPK7.** The studies in this aim are directed to better define the function of the orphan kinase PfPK7 using a PfPK7 null cell line. *In vitro* kinase assays revealed the ability of PfPK7 to directly phosphorylate three previously unknown substrates of the kinase and aided in elucidating the role PfPK7 plays in signaling cascades.

3) **Elucidating the substrates/interactors of *Plasmodium* CDK-like Kinases.** Two different approaches, bioinformatics and proteomic analysis of the macromolecular complex from cell lines expressing epitope-tagged kinases, were used to define kinase-substrate/interactor relationships of CDK-like kinases PfPK6 and Pfmrk.

**1.2 Literature Review**

**1.2.1 Plasmodium falciparum: The Most Virulent Form of Malaria**

Despite intensive efforts to eradicate or control malaria, it remains a major threat to public health and economic development in both the tropics and subtropics. Approximately 40% of the world’s population, over three billion people, are at risk of contracting malaria. Malaria is caused by protozoan parasites of the genus *Plasmodium* in the Apicomplexan family. More than 200 species of *Plasmodium* have been identified that infect mammals, reptiles, or birds [4]. Four species of *Plasmodium* are traditionally recognized as the causative agents of human malaria: *Plasmodium falciparum, Plasmodium vivax, Plasmodium ovale*, and *Plasmodium malariae* [5]. Of the *Plasmodium* species that infect humans, *Plasmodium falciparum* causes the most lethal form of malaria, resulting in 300-600 million clinical cases and 1 to 3 million deaths annually [6].
Malaria infection presents a major global health challenge, currently afflicting over 500 million people worldwide and significantly increasing childhood mortality in the poorest nations, while causing an estimated 1.3% reduction in economic productivity in affected countries [7]. The majority of all malaria cases occur in the countries of sub-Saharan Africa. Approximately 25% of all deaths in children under the age of five in Africa are due to malaria [8]. Outside of Africa, there are an estimated 85 million malaria cases linked to *P. falciparum* infection [9]. Malaria infections once occurred in areas such as western Europe and the United States; however, the disease has achieved near or complete elimination due to economic development and the implantation of improved public health policies [8]. Cases of malaria that do present in the United States or Europe are primarily due to international travel from Africa, the Middle East, Asia, or the Amazon. Approximately 1,500–2,000 cases of malaria are reported every year in the United States. In 2011, 1,925 cases of malaria were reported in the United States of which 58% were due to *P. falciparum* infection [10]. Clinical complications resulting from *P. falciparum* infection include fever, anemia, lactic acidosis, cerebral malaria, acute respiratory distress, epilepsy, hypoglycemia, and multi-organ failures. The effect of infection in pregnant women is even more profound. When *P. falciparum* infection occurs in the first or second pregnancies, the result may be spontaneous abortion, still birth, low birth weight, and anemia or death for the mother [11].

Traditionally, there have been two avenues for the control of malaria: prevention of contact with humans (through the removal of mosquito breeding sites, insecticides, screens, and insecticide treated bed nets) and effective disease management (through antimalarial treatments such as chloroquine and artemisinin combination therapies). Although antimalarials have proven useful in the past, the extensive deployment of antimalarials has lead to selection pressure on *Plasmodium* parasites, leading to the emergence of widespread resistance. While artemisinin-based combination
treatments (ACTs) have played an effective role in controlling the disease in many malaria endemic areas, the appearance of parasites resistant to artemisinin derivatives in wide areas of Southeast Asia (encompassing south Vietnam to central Myanmar) underscores the fragility of the available malaria treatment measures [12, 13]. Therefore, there is an increased need for the identification of novel cellular targets for therapeutic intervention. The identification of novel cellular targets requires a nuanced understanding of the idiosyncratic molecular mechanisms that underlie progression of the *P. falciparum* life cycle. Despite long-standing interest due to its medical and biological significance, the atypical life cycle of *Plasmodium* remains ill defined at a molecular level; more generally, the signaling networks that govern all aspects of its biology, from recognition of environmental cues to cell growth and differentiation, are poorly understood [1].

### 1.2.2 Unique Life Cycles of Malaria Parasites

Most protozoan parasites, including many with large impacts on global public health such as apicomplexans, have complex life cycles, which are difficult to study with classical genetic approaches. The life cycle of the *Plasmodium* parasite consists of invasion from the mosquito midgut to the human liver, and subsequent progression to human erythrocytes. The malaria parasite life cycle, among the most complicated of any organism, is characterized by three invasive forms (sporozoites, merozoites, and ookinetes) and several intermediary forms. The *Plasmodium* life cycle consists of two asexual stages, tissue schizogony and erythrocytic schizogony, and one sexual stage referred to as mosquito sporogony (Fig. 1) [14].
The lone sexual replication stage in the malaria life cycle occurs in the midgut lumen of mosquitoes. The ookinete develops from a zygote as a direct product of fertilization between a male and female gametocyte. Once fertilized, the ookinete traverses the midgut epithelial layer to reach the basal lamina, becoming immobilized and maturing into oocysts [15]. The oocysts continue to grow and mature into sporozoites, which are released in the mosquito body and travel to the salivary glands. When the infected mosquito bites a human host, it releases the sporozoites from the salivary glands into the human blood stream where they are directed to the liver. In the liver, the
parasite invades hepatocytes and undergoes asexual replication within a parasitophorous vacuole in a process known as exo-erythrocytic schizogony. At this stage, *P. vivax* and *P. ovale* undergo a dormant period in which they form hypnozoites and can remain inactive up to 30 years, while *P. falciparum* hepatocyte infection only last approximately 6 days. The result of this asexual replication in the liver is a 10,000-fold amplification of parasite numbers and the release of merozoites into the human host blood stream [16]. Free-moving merozoites attach and invade enucleated erythrocytes where the parasite undergoes a 48-hour cycle of asexual replication and division. During the intraerythrocytic cell cycle the parasite undergoes three morphologically distinct stages referred to as ring (0-24 hours post infection), trophozoite (24-36 hours post infection), and schizonts or segmenters (36-48 hours post infection). The ring stage is the earliest intraerythrocytic stage and at this point in the cell cycle the parasite begins feeding on the newly invaded erythrocyte. As the parasite grows, morphological changes can be seen as the ring becomes more rounded and irregular in nature, marking the transition to the trophozoite stage. During the trophozoite stage the parasite’s metabolism is highly active, performing glycolysis, ingesting host cytoplasm, and performing proteolysis of hemoglobin to break it down into constituent amino acids [8]. The transition from trophozoite to schizont stage occurs as the parasite undergoes multiple rounds of nuclear division without cytokinesis in a process referred to as erythrocytic schizogony. The result of schizogony is the appearance of a multinucleated schizont parasite that matures into approximately 20 merozoites, which then rupture from the RBC and continue on to invade uninfected RBCs [17]. In *P. falciparum*, *P. ovale*, and *P. vivax* the intraerythrocytic cell cycle repeats approximately every 48 hours leading to the exponential growth of parasite numbers [8]. The intraerythrocytic stages of infection are responsible for the pathogenesis associated with malaria. One hallmark symptom of a malaria infection is acute or severe anemia due to the
destruction of host erythrocytes. Other clinical complications include fever, shivering, cough, respiratory distress, joint pain, headaches, diarrhea, vomiting and convulsions [18]. Cerebral and placental malaria may also arise from the sequestration of mature stage parasites in the brain and placenta. During the process of schizogony, for not well understood reasons, some of the merozoites differentiate into sexual forms of the parasite: macrogameotocytes (female) and microgametocytes (male), which can then be ingested by a female anopheline mosquito and continue the sporogonic cell cycle [5].

Clearly, the intraerythrocytic cell cycle of *Plasmodium* is unique and complex. Accordingly, a correlation between archetypal G1, S, G2, and M phases of eukaryotic cell cycle with those of the *Plasmodium* developmental cycle is difficult to establish, and a clear understanding of the molecular switches that coordinate DNA replication, mitosis, and cytokinesis to produce 8 to 32 merozoites from a single cell will be of great interest. The research done here focuses on using innovative approaches to delineate *P. falciparum* cell cycle regulatory protein networks involving PKs, particularly CDK-like kinases because the CDKs are master regulators that orchestrate the activities of key proteins to precisely execute various cell cycle events [19].

### 1.2.3 Identification of the *P. falciparum* Kinome

There is a pressing need for the development of new antimalarials with novel modes of action, such as regulators of the non-canonical cell cycle of *Plasmodium*. Protein kinases are an important family of proteins that regulate diverse cellular activities. Next to G-protein-coupled receptors, PKs are targeted heavily by the pharmaceutical industry in their screening programs. Throughout eukaryotes, protein kinases universally function as regulatory switches to drive cell cycle progression by phosphorylating key effector proteins, and malaria parasites are not likely to
be an exception. Genome-wide sequencing of *P. falciparum* has identified the repertoire of parasite protein kinases (PKs) based on homology searches [20]. Surprisingly, this revealed only 65 PKs related to eukaryotic PKs (ePKs), along with a novel ePK-related family of 20 members, named FIKKs after a shared Phe-Ile-Lys-Lys motif in their catalytic domain [21]. This number is rather low considering that (i) *Saccharomyces cerevisiae* encodes approximately twice as many ePKs within its comparably sized genome and (ii) the parasite’s developmental cycle spans multiple hosts, requiring specialized regulation in each. The protein sequences of the 65 identified ePKs were used to construct a phylogenetic tree, revealing that the parasite possesses enzymes belonging to most of the major serine/threonine kinase groups, while several of the kinases do not cluster with any of the traditional kinase groups [21].

Based on the amino acid sequence of the catalytic domain, *P. falciparum* protein kinases were classified into the major ePK serine/threonine kinase groups. There are 10 traditional ePK kinase groups: CK1, AGC, CaMK, CMGC, TKL, STE, TyrK, RGC, PKL and Atypical. The casein kinase 1 (CK1) family is a vast group composed of protein kinases with an acidic recognition sequence that partake in hierarchical protein phosphorylation reactions, and phosphorylate a wide range of substrates [22]. PfCK1 is the only malaria kinase that falls within this group [23].

The AGC group is composed of protein kinases that are directed by basic amino acids [24]. These include the cyclic nucleotide-dependent families PKA and PKG, the PKC family, the β-adrenergic receptor kinase family, the ribosomal S6 kinase family, and other relatives [23]. Five malarial kinases cluster within this group [21]. The Ca\(^{2+}\)-calmodulin kinase (CaMK) group are also a group of basic amino acid-directed kinases with a preference for phosphorylating serine or threonine residues close to arginine and lysine residues [24]. The CaMK group includes protein kinases that are regulated by Ca\(^{2+}\)-calmodulin, calcium-dependent protein kinases (CDPKs), the
SNF1/AMP-activated kinase family, and other relatives [23]. To date, 13 *P. falciparum* kinases have been identified in this group, which demonstrates the significance of calcium signaling in the malaria parasite [25].

The tyrosine-like kinases (TKL) are a group of serine/threonine protein kinases with close sequence similarity to tyrosine kinases. Five malarial kinases show homology to the TKL group; however, their ability to function as a TKL kinase has yet to be shown [21]. No malarial protein kinases have been identified that cluster with the STE, TyrK, RGC, or PKL groups [23]. The STE group of kinases are homologs of the yeast STE7, STE11, and STE20 genes, which form the MAPK cascade and act in a signal transduction pathway. Unlike traditional protein kinases that phosphorylate serine or threonine residues, members of the tyrosine kinase (TyrK) group almost exclusively phosphorylate tyrosine residues. The RGC is a small group consisting of receptor guanyl cyclases, while the PKL group is a larger group sharing a PKL fold and catalytic mechanism. Although no malaria protein kinases cluster with the STE, TyrK, RGC, or PKL groups, several do not clearly cluster with any defined group and are classified as “orphan” kinases. This group does not strongly associate with any established ePK group [26].

The most prominent group in the *Plasmodium* kinome is the CMGC group. Eighteen malarial kinases cluster in this group. The CMGC group are frequently proline-directed protein kinases and include the family of cyclin-dependent protein kinases (CDKs), mitogen-activated protein kinases (MAPKs), glycogen-synthase kinase 3 family, CDK-like kinases (CLK), and other close relatives [23]. MAPK pathways play crucial roles in the control of eukaryotic cell proliferation in response to various extra- or intracellular stimuli [27]. In eukaryotic systems, the majority of the CMGC kinases are involved in the control of cell proliferation and development, highlighting the significance of kinases belonging to this cluster. Kinases belonging to either ePKs
or “orphan” kinase groups are fitting potential drug targets due to their predicted roles in regulating the unique *P. falciparum* life cycle and due to the distinct structural and functional characteristics they possess in comparison to their human host which may provide *P. falciparum*-specific targets [28, 29].

**1.2.3.1 Cyclin-dependent Protein Kinases**

Cyclin-dependent kinases (CDKs) are a large group of proteins that have a critical role in cellular processes such as cell cycle progression, transcription, DNA replication and repair, apoptosis, and differentiation. Abnormalities in the regulation of CDKs lead to numerous cell cycle defects and a wide variety of diseases. Currently, CDKs are being investigated as potential drug targets in cancer, neurological disorders, and cardiovascular disease [30]. Due to their highly conserved nature and their role in cell growth, CDKs are attractive targets for therapeutic intervention. In its active form, CDKs are heterodimers consisting of a catalytic serine/threonine kinase subunit and a regulatory subunit. When an inactive CDK binds its cognate cyclin it becomes a partially active enzyme and there is a displacement of a sterically hindering “T-loop” exposing critical residues in the catalytic cleft [31]. CDKs will then become fully active when a CDK activating complex (CAK) phosphorylates a conserved threonine residue within the “T-loop” of the CDK [32]. The CAK complex traditionally is composed of CDK7, cyclin H, and the stabilizing partner MAT1 [2]. Several CDKs and cyclins have been identified in *P. falciparum* by using homology-based PCR and database mining [33]. Although CDKs and cyclins have been identified in *P. falciparum*, the regulatory mechanisms responsible for activating and inhibiting CDK activity in *Plasmodium* remain unknown. Furthermore, the substrates of CDKs and CDK regulation mechanisms of the non-canonical malarial cell cycle have yet to be determined.
In *P. falciparum*, 6 CDK-related protein kinases have been identified, PfPK5, PfPK6, Pfmrk, Pfcrk-1, Pfcrk-3, Pfcrk-4 [19]. PfPK5, a putative homologue of p34\(^{\text{cdc2}}\) cyclin-dependent kinase required for both entry into S-phase and mitosis in fission yeast, was the first CDK-like kinase characterized in *P. falciparum* with 60% identity to human cdc2 [34]. PfPK5 co-localizes with the nuclear stain at the onset of schizogony [35]. It possesses a fairly conserved cyclin binding PSTAIRE motif (PSTTIRE) and potential sites of regulatory phosphorylation (equivalent to Thr14, Tyr15, and Thr160 of CDK2). The structure of PfPK5 has been determined recently to a resolution of 1.9 Å [36]. PfPK5 has structural homology to human CDK2, the only other monomeric CDK structure solved.

### 1.2.4 *P. falciparum* Protein Kinase 6

*Plasmodium falciparum* protein kinase 6 (PfPK6) is a novel protein kinase that has sequence homology to mitogen activated protein kinases (MAPKs) and also to cyclin dependent kinases (CDKs). Both MAPKs and CDKs have important functions in the regulation of growth and cell cycle in eukaryotes. PfPK6 shows some structural and functional resemblances to eukaryotic CDKs. The similarity of the catalytic domain of PfPK6 to CDK2 and MAPK is 57.3% and 49.6%, respectively [37]. Another feature suggesting PfPK6 functions as a CDK-like kinase is the presence of conserved regulatory phosphorylation sites found in other CDKs: Thr14, Thr15, and Thr160 in CDK2 [19].

Although PfPK6 shows sequence homology to CDKs, there are several lines of evidence suggesting that PfPK6 is a cyclin-independent kinase. First, the recombinant PfPK6 shows significant auto-phosphorylation and phosphorylation of the control substrate histone in the absence of a cyclin. Secondly, PfPK6 activity is not stimulated by incubation with cyclins [19]. Thirdly,
several amino acids in the N-terminal lobe of CDKs that are involved in cyclin binding are absent from PfPK6. Furthermore, closer examination of the canonical cyclin binding PSTAIRE motif in CDKs reveals that the motif is replaced by a SKCILRE sequence in PfPK6 [37]. These observations suggest that PfPK6 is not a traditional CDK and instead is a unique “cyclin-independent” CDK that may be regulated through its own synthesis or degradation and not the binding of a cyclin partner [19].

PfPK6 transcript levels were found to peak at the trophozoite stage, the cell cycle stage thought to correspond to the onset of the classical S phase during which DNA replication occurs [37]. Protein expression levels of PfPK6 were detected at both the trophozoite and schizont stages, with peak expression occurring at the trophozoite stage. Previous in vitro kinase assays detected the ability of PfPK6 to auto-phosphorylate, phosphorylate histone, and phosphorylate the small subunit of the malarial ribonucleotide reductase. It was also shown that PfPK6 activity is sensitive to common CDK inhibitors such as roscovitine [37]. Identifying and characterizing substrates and interactors of PfPK6 will generate a preliminary network of interactions for PfPK6 and provide insight into its function(s) in Plasmodium signaling networks.

1.2.5 P. falciparum MO15-related Protein Kinase Pfmrk

Pfmrk was first isolated in 1997 through a PCR approach. Homology searches of Pfmrk revealed that it has the highest sequence identity (46%) to human CDK7 of all of the mammalian CDKs [38]. CDK7 has a dual function as the TFIIH associated kinase responsible for regulation of transcription and DNA repair and activates down-stream CDKs through “T-loop” phosphorylation [39]. As part of the CAK complex, CDK7 traditionally binds cyclin H and the stabilizing partner MAT1. Previous studies have shown the ability of Pfmrk to phosphorylate both histone H1 and the
carboxyl-terminal domain of the large subunit of RNA polymerase II [40, 41]. Furthermore, Pfmrk activity was stimulated in vitro when it was associated with human cyclin H (the cyclin partner of human CDK7) or Pfcyc-1 (a cyclin homolog in P. falciparum) [40, 42]. Although in vitro studies support the notion that Pfmrk is a Plasmodium CDK homologue, more recent studies suggest that Pfmrk may be a member of a unique protein kinase family [43]. To date, Pfmrk has not been shown to possess CAK activity against PfPK5 or other CDK-related kinases; however, Pfmrk has been shown to interact and co-localize with the C-terminal domain of PfMAT1 [31, 42]. Previous studies also showed the interaction of Pfmrk with two DNA replication proteins, PfRFC-5 and PfMCM6 and the ability of Pfmrk to utilize these proteins as substrates. Taken together, the data suggest a role for Pfmrk in the nucleus of the parasite presumably in regulation of the DNA replication machinery, rendering it a highly attractive target for therapeutic inhibitors [31].

1.2.6 P. falciparum Atypical Protein Kinase 7

Plasmodium falciparum protein kinase 7 (PfPK7) is one possible P. falciparum-specific target with no human homolog [21]. PfPK7 is an “orphan” kinase displaying regions of homology to more than one protein kinase family. The C-terminal lobe of PfPK7 displays maximal homology to the MEK family, while the N-terminal lobe is related to the fungal cyclin AMP-dependent kinases [27]. Sequence analysis revealed the two closest human homologs as dual specificity mitogen activated protein kinase 3 (MAP2K3) and serine/threonine-protein kinase 6 (Aurora-A) with 33% and 26% sequence homology, respectively [44]. Previous studies have demonstrated the in vitro kinase activity of PfPK7 to autophosphorylate and phosphorylate a number of substrates including myelin basic protein, histone H2A, and β-casein [45]. Despite possessing homology to the MEK family of kinases, PfPK7 was not able to phosphorylate MAPK homologs and was not
inhibited by MEK inhibitors; therefore, PfPK7 is not likely to function as a MEK homolog in *P. falciparum* [45].

PfPK7 has been shown to be expressed at both the asexual and sexual stages of the human host, as well as within the mosquito. Through the use of reverse genetics, Dorin-Semblat et al. disrupted the PfPK7 gene, which resulted in the slowing down of parasite growth through a reduction in the number of merozoites produced by each schizont and through a reduction in the ability to produce oocysts in the mosquito [27]. This observed phenotype suggests that PfPK7 is involved in parasite proliferation and development, marking PfPK7 as a promising *Plasmodium-* specific drug target.

1.2.7 Current Technologies used to Identify Protein Kinase Substrates

Protein kinases are the largest group of enzymes capable of post-translational modifications in eukaryotes [46]. Protein kinases are responsible for phosphorylating protein substrates by transferring a phosphate group from adenosine-triphosphate (ATP) to a serine, threonine, or tyrosine residue on the substrate protein [47]. This essential post-translational modification regulates nearly every aspect of biological pathways, from protein-protein interactions to signal transduction [48]. When protein phosphorylation becomes deregulated, it often leads to the development of numerous diseases such as cancer. Therefore, phosphorylation analysis to identify phosphorylated proteins and quantitative changes in phosphorylation is crucial to understanding how signaling networks function and how the disruption of these networks lead to disease states.

There are five traditional approaches used when investigating phosphorylation events: genetic screening, *in vitro* kinase assays, protein interaction-based substrate screening, mass spectrometry-based high throughput screening, and bioinformatics of kinase-substrate predictions.
Genetic screening is one of the most historically applied approaches used to discover potential substrates of protein kinases. Once the phenotype of a particular kinase knockout has been determined, high throughput genetic manipulation is performed, such as siRNA, on a genome-wide scale in order to screen all the genes that can mimic the kinase knockout phenotype. The genes capable of producing a similar phenotype are then tested to determine if they are true substrates of the kinase being investigated [46]. This approach has been successful in model organisms such as yeast, worms, and flies where genetic manipulation can be conducted easily; however, this approach is not as feasible for nontraditional organisms such as *Plasmodium*.

Another approach to identifying protein kinase substrates is *in vitro* kinase assays. This is the most commonly used method when determining kinase activity toward substrates. This technique is performed by incubating a purified kinase with a purified putative substrate in the presence of ATP. The ability of the kinase to utilize any given substrate is then assessed through colorimetric, radioactive, chemiluminescent, or fluorometric techniques [49]. Perhaps the biggest limitation to this approach is the fact that *in vitro* phosphorylation may differ from true physiological events. Another limitation is the fact that a researcher must have an idea of what proteins are being phosphorylated by a particular kinase, which is why *in vitro* kinase assays are typically used as a confirmatory technique to verify previously identified substrates.

More recently, focus has been placed on the use of protein interaction-based substrate screening in order to identify kinase-interacting proteins. A number of approaches have been used to study protein-protein interactions in order to identify potential kinase substrates; however, phosphorylation is usually a transient protein-protein interaction, making the detection of substrates through protein-protein interactions difficult [46]. The use of mass spectrometry-based high throughput screening and bioinformatics of kinase-substrate predictions have recently become
widely used due to the advancements in technology in the past decade. These techniques are discussed in detail within the following two sections.

1.2.8 Quantitative Mass Spectrometry in Proteomics

Traditional proteomic quantification approaches utilize dyes, fluorophores, or radioisotopes to determine the relative abundance of a protein in a particular sample. The advantages associated with the classical methods include good sensitivity and range, but they require high-resolution protein separation and do not lead to the identification of the protein, demonstrating the need for modified techniques [50]. In recent years, the field of proteomics has grown vastly with an emphasis placed on the development of techniques that allow for the comparison of protein abundance among biological samples. This quantitative comparison allows for the characterization of changes in protein abundance in various organisms under many different biological conditions. The most successful strategies currently in use rely on the introduction of an isotope-coded tag, which is used to differentiate among samples. In 1999 3 different laboratories reported the first usage of stable isotope labeling which has become a staple of the proteomics field since then [51-53]. Currently, there are a variety of approaches used to introduce the isotope tags including metabolically, chemically, or enzymatically; however, once the tags have been incorporated, the fundamental strategy for each technique is the same. More specifically, after tag incorporation, the samples are combined and analyzed simultaneously in order to reduce experimental error between samples [54]. Next the intensities of the different isotope tags can be compared in order to determine the relative abundance of each peptide, which can then be used to infer the relative abundance of each parent protein [54].
Metabolic labeling is the first point possible for introducing a stable isotope tag into proteins, which occurs during cell growth and division. The first use of metabolic labeling was performed using \(^{15}\text{N}\)-enriched cell culture medium for total labeling of bacteria [52]. The technique has since expanded and been used in a wide range of organisms. One of the most widely used approaches for metabolic labeling is the stable isotope labeling by amino acids in cell culture (SILAC) method, which was first introduced in 2002 [55]. This method utilizes a culture medium containing \(^{13}\text{C}_6\)-arginine and \(^{13}\text{C}_6\)-lysine in order to ensure that all tryptic peptide products carry at least one labeled amino acid. Proteins are identified on the basis of fragmentation spectra from one “heavy” and one “light” peptide and subsequently the relative quantification was determined based on the comparison of the intensities of isotopes [50]. The utility of metabolic labeling comes from the fact that samples treated differently can be pooled for analysis, which reduces the amount of introduced experimental error during protein extraction, protein digestion and fractionation. One of the biggest limitations of metabolic labeling is the small varieties of labels are available for use. For example, SILAC labeling can only use two labels, which restricts the number of conditions that can be compared in one experiment to three (two labeled and one unlabeled). Overall, metabolic labeling is one of the most accurate quantitative MS methods when comparing a small number of conditions due to the early combination of samples [50].

Chemical or enzymatic incorporation of isotope tags is performed \textit{in vitro} as a post-biosynthetic mechanism. Enzymatic labeling is usually performed after proteolytic digestion and was traditionally labeled by the integration of \(^{18}\text{O}\) into the COOH-termini of peptides [56]. When using this method, it is rare that complete labeling could be achieved; therefore, the rate of incorporation of labels into individual peptides vary enormously, making accurate analysis of data more complicated [57]. In 1999 a novel approach was created utilizing an affinity-tag that is
isotope-coded (ICAT) that specifically labels cysteine residues with zero or eight deuterium atoms as well as a biotin group. A disadvantage to this approach is the fact that there are a significant number of proteins that contain few to no cysteine residues and therefore will not be labeled [51].

Another method of protein and peptide labeling utilizes labeling reagents that target the peptide N-terminus and the epsilon-amino group of lysine residues [50]. The isotope-coded protein label (ICPL), isotope tags for relative and absolute quantification (iTRAQ), and tandem mass tags (TMT) are among the tags capable of this type of labeling. When using a protein or peptide tag, relative quantification is accomplished using the integration of mass spectrometric signals originating from “heavy” and “light” labeled peptides [50]. However, the concept of isobaric mass tagging is divergent from the above concept by attaching tags that generates isobaric labeled peptides which co-migrate in chromatographic separations. Once fragmentation of peptides has been performed, the different tags can then be identified from the MS data. Thus this approach enables us to determine simultaneously the identity as well as the relative abundance of pair of peptides resulting from tandem mass spectra, which is a great advantage. Another advantage of this approach is the commercial availability of iTRAQ reagents allowing for multiplexed quantification of up to eight samples and the comparison of up to eight time points or treatments in the same experiment [58].

In order to determine the absolute quantification of proteins, an isotope-labeled synthetic internal standard must be used. This is attained by the introduction of a known amounts of a standard peptides that are stable isotope-labeled to the digested protein followed by the comparison of the signals from MS to the experimental sample peptides [59]. More recently, synthetic genes have been constructed that, upon digestion with trypsin yield peptides of the same protein to be uses as a standard for quantification. This approach increases confidence in the quantification process.
and removes experimental bias that could occur during the digest process [60]. Although the introduction of an internal standard greatly enhances sample quantification, sample quantification is limited by the co-elution of peptides with a similar mass. This limitation is overcome by decreasing sample complexity by introducing biochemical fractionation prior to LC-MS analysis [50]. The general procedure for obtaining information from MS and MS/MS spectra that is quantitative is similar among different quantification approaches. Figure 3 depicts in yellow the quantification steps that are common between approaches with and without utilizing stable isotopes, while the blue boxes illustrate the additional steps that are required when the MS signal intensity values are used for quantification [50].

![Figure 2: Quantitative Mass Spectrometry Data Processing and Analysis Workflow](image-url)

Figure 2: Quantitative Mass Spectrometry Data Processing and Analysis Workflow

Figure adapted from [50]
Spectra counting is most frequently used when determining relative protein quantification between multiple samples. For this analysis, the number of spectra observed for a protein in each condition is counted and compared among the different conditions in order to determine if there is a significant difference. This method can lead to variation among experiments due to varying levels of peptide detection. In order to diminish this effect, most quantitative proteomic studies average the ratio of spectra between experiments; however, this approach can also vary based on signal intensity [61]. Protein quantification can be improved by either calculating a weighted average of intensities, addition up all measured quantities, and by calculating the ratio of proteins, or by computing a linear regression to determine the protein ratio [61-63].

Quantitative mass spectrometry is most frequently used to correlate a protein’s abundance in an experimental condition or to determine the protein interactions. The most common use of quantitative mass spectrometry is to detect protein abundance changes among experimental conditions. This method has recently been applied as a comprehensive investigative tool of biological phenomena, as was performed in this study to define the constitutive and regulated expression of the proteome during the intraerythrocytic developmental cycle.

1.2.9 Phosphoproteomic Analysis

Protein phosphorylation is one of the post-translational modifications (PTMs) in living cells that is of major significance in protein phosphorylation. This is because the reversible protein phosphorylation has a central regulatory role in signal transduction mechanisms across species. Protein phosphorylation is involved in a diverse regulatory processes such as regulation of cell cycle, transmission of signals from cell surface to the nucleus, cellular differentiation, normal and aberrant cell proliferation, and metabolic processes [64]. Protein kinases and phosphatases account
for 2-4% of eukaryotic proteomes and are responsible for controlling substrate modifications through reversible phosphorylation and dephosphorylation [65, 66]. Protein phosphorylation plays a crucial role in the cell and it is estimated that about 30% of all eukaryotic proteins are reversibly phosphorylated. The level of phosphorylation in an individual protein ranges from less than 1% to greater than 90% [67, 68]. Protein phosphorylation acts as a regulatory mechanism in signal transduction pathways by activating or deactivating key proteins. A disruption in the balanced control of protein phosphorylation has been linked to many diseases, including cancer [48, 69].

Protein phosphorylation can be discovered and visualized in SDS PAGE gels of proteins that are $^{32}$P labeled proteins or by western blot analysis with phospho-specific antibodies. Although these approaches are limited in their ability to identify novel phosphoproteins or to localize phosphorylation sites within a protein [64]. Mass spectrometry has evolved drastically in recent years and can now identify, localize, and quantify thousands of in vivo phosphorylation sites. It can also be applied in studies examining almost any question in basic cell biology in a wide variety of biological contexts. Mass spectrometry based phosphoproteomics is frequently used to delineate cell signaling pathways. More specifically, the relative quantification of signaling molecules involved in various signaling cascades as well as their phosphorylation sites can be determined [64]. Phosphoproteome analysis has been performed in many organisms ranging from yeast to mouse and has proven to be an important tool for defining key proteins within cellular signaling networks.

Phosphoproteomic analysis is frequently used to study whole-tissue physiology, cell differentiation status, signal transduction cascades, and kinase/substrate specificity. The preferred method in quantitative mass spectrometric based proteomic analysis utilizes the enzymatic digestion of complex samples into peptide mixtures that are resolved by liquid chromatography
followed by MS analysis [64]. Sample preparation for phosphoproteomic analysis is similar to quantitative proteomic sample preparation; except that phosphopeptides, representing only a small percentage of total peptides present in a total lysate sample, needs to be enriched for efficient detection by mass spectrometry. There are many methods currently being used to enrich total lysate samples for phosphopeptides, the most common are affinity and antibody-based methods. The typical workflow used for quantitative phosphoproteomics is depicted in Fig. 4 [64].

One of the most common phosphopeptide enrichment strategies being used today is immobilized metal affinity chromatography (IMAC). IMAC is based on the concept that covalent bonds form between specific amino acids, such as histidine and cysteine, and metal ions. Proteins with an affinity for metal ions will be retained in a column containing immobilized metal ions, such as cobalt, nickel, copper, iron, zinc, aluminum, or gallium. Iron, aluminum, and gallium are the preferred ions for phosphopeptide enrichment and are typically immobilized to iminodiacetic acid (IDA) derived resin [70]. The IMAC concept was first introduced by Anderson et al in 1986 [71] and later improved upon by Tempst et al who compared the selectivity of a variety of metals in the IMAC procedure and their ability to bind phosphopeptides [72]. During the optimization process, Tempst et al observed the best selectivity with IDA columns complexed with gallium in the pH range of 2.0-3.5 [72].

Another approach used for the enrichment of phosphorylated peptides is strong cation exchange chromatography (SCX). This method is based on the concept that phosphorylated and non-phosphorylated peptides are going to have a difference in the solution charge states. More specifically, at a pH of 2.7, the negatively charged phosphate group on a phosphorylated peptide causes a decrease in net charge, which allows for their enrichment over non-phosphorylated peptides [64]. Although this technique is useful, it does not allow for multi-phosphorylated peptides
to be bound to the SCX column due to their net zero charge, which is why SCX is typically a first step in the phosphopeptide enrichment workflow and is usually followed by IMAC. The combination of SCX and IMAC is a strong analytical tool used in many large-scale phosphoproteomic studies [73-75].

Figure 3: Typical Quantitative Phosphoproteomics Workflow

Figure publically available at https://www.biochem.mpg.de/225898/Phosphorylation
Global phosphoproteomic studies based on mass spectrometry have generated qualitative and quantitative data describing protein phosphorylation events in various biological systems. Reversible phosphorylation of proteins regulates many cellular processes. Activation of most signaling pathways in eukaryotes is because of coordinated phosphorylation events across many proteins over time. The malaria parasite life cycle is complex and the regulation of proliferation and differentiation events by protein phosphorylation is likely a crucial aspect. The divergence between human and malarial protein kinases suggest that specific inhibition of malarial protein kinases is feasible.

1.2.10 The P. falciparum Proteome and Phosphoproteome

Despite long-standing interest due to its medical and biological significance, the atypical life cycle of Plasmodium remains ill defined at a molecular level. The signaling networks that govern all aspects of the parasite’s biology, from recognition of environmental cues to cell growth and differentiation, are poorly understood. To date, our knowledge of this organism has been restricted due to challenges in applying traditional biochemical, cell biological, and genetic screens in this system. However, in recent years, a variety of technological advances have enabled a paradigm shift in biology, broadening the focus from studies restricted to individual genes or proteins to a system-wide, global analysis of cellular networks. Importantly, many of these technologies, including mass spectrometry-based methods for global quantitative analysis of proteomes and phosphoproteomes, have proven to be applicable to Plasmodium, thus offering unprecedented insights into the molecular foundations of this organism’s unique and medically important life cycle [76-78].

A variety of previous proteomic studies utilizing a wide range of technologies for fractionation and quantification have explored various stages of the Plasmodium life cycle [79-82].
These studies have investigated sporozoite, merozoite, trophozoite, schizont, and gametocyte stages independently of one another and typically identified several hundred to approximately 2,500 proteins [81-85]. Multidimensional protein identification (MudPIT) technology was used for a system-wide comparison of the transcriptome and proteome; however, this study detected proteins from only 50% of the transcripts in the asexual stages [79]. Significantly, this study revealed a delay in the appearance of proteins compared to that of the transcripts, indicative of a major role of post-transcriptional mechanisms in the regulation of gene expression. Profiling of schizont stage protein expression by two-dimensional differential gel electrophoresis followed by mass spectrometric analysis provides evidence for post-transcriptional regulation of gene expression [80].

Recently, merozoite proteome analysis using isobaric tagging and strong cation exchange (SCX) fractionation found 677 proteins of which 92 are post-transcriptionally regulated [81]. Another study using MudPIT identified 802 proteins in the *P. falciparum* nuclear proteome [82].

Phosphoproteome analysis has been performed in many organisms ranging from yeast to mouse in order to profile the phosphorylation status of the proteome. By providing a global overview of phosphorylation, these studies have proven to be important tools for defining key proteins within cellular signaling networks. An initial attempt to profile schizont stage protein expression was performed with two-dimensional differential gel electrophoresis followed by mass spectrometric analysis and identified 170 proteins [86]. More recently, reports describing the characterization of the *P. falciparum* phosphoproteome focused on the schizont stage of the intraerythrocytic growth of the parasite. The first study used immobilized metal ion affinity (IMAC) and TiO$_2$ chromatography followed by LC-MS/MS analysis to identify 650 proteins with 1,177 phosphorylation sites [77]. Another study used SCX chromatography, IMAC, and LC-MS/MS to detect 1,673 proteins that are phosphorylated at 8,463 sites [76]. A third study used two sample
preparation procedures (i) SDS-PAGE protein gel fractionation and phosphopeptide enrichment with TiO$_2$ followed by LC-MS/MS and (ii) filter assisted sample preparation (FASP) with strong anion exchange chromatography followed by TiO$_2$ phosphopeptide enrichment and LC-MS/MS to identify 929 proteins with 2,541 phosphorylation sites [78]. Nonetheless, these earlier studies have examined only mature schizont stage parasites. Notably, these previous studies did not investigate the dynamics of protein expression and phosphorylation during the entire intraerythrocytic developmental cycle of Plasmodium. Here we present a comprehensive analysis of protein expression and phosphorylation across the three major developmental stages in the *P. falciparum* intraerythrocytic asexual cycle using isobaric labeling [58, 87]. This comparative analysis revealed stage-specific profiles of protein expression and phosphorylation.
CHAPTER 2: GLOBAL ANALYSIS OF PROTEIN EXPRESSION AND PHOSPHORYLATION OF THREE STAGES OF *PLASMODIUM FALCIPARUM* INTRAERYTHROCYTIC DEVELOPMENT


2.1 Summary

During asexual intraerythrocytic development, *Plasmodium falciparum* diverges from the paradigm of eukaryotic cell cycles by undergoing multiple rounds of DNA replication and nuclear division without cytokinesis. A better understanding of the molecular switches that coordinate the numerous events governing the progression of the parasite through the intraerythrocytic developmental stages will be of fundamental importance for rational design of intervention strategies. To achieve this goal, we performed isobaric tag-based quantitative proteomics and phosphoproteomics analyses of three developmental stages in the *Plasmodium* asexual cycle and identified 2,767 proteins, 1,337 phosphoproteins, and 6,293 phosphorylation sites. Approximately 34% of identified proteins and 75% of phosphorylation sites exhibit changes in abundance as the intraerythrocytic cycle progresses. Our study identified 43 distinct phosphorylation motifs and a range of potential MAPK/CDK substrates. Further analysis of phosphorylated kinases identified 30 protein kinases with 126 phosphorylation sites within the kinase domain or in N- or C-terminal tails. Many of these phosphorylation events are likely CK2-mediated. We defined the constitutive and regulated expression of the *Plasmodium* proteome during the intraerythrocytic developmental cycle, offering an insight into the dynamics of phosphorylation during asexual cycle progression.
Our system-wide comprehensive analysis is a major step toward defining kinase-substrate pairs operative in various signaling networks in the parasite.

2.2 Materials and Methods

2.2.1 Plasmodium falciparum Culture

Parasites of Plasmodium falciparum strain 3D7 were initially grown at a 4-10% parasitemia and 4% hematocrit in RPMI 1640 culture medium supplemented with A+ erythrocytes and 5% Albumax as previously described [88]. Schizont stage parasites were magnetically synchronized as described by Kim et al. using a MACS LD (Miltenyi Biotec Auburn, CA) column with a Midi-MACS Separator [89, 90]. Parasite-infected erythrocytes were added to a fresh culture plate and supplemented with A+ blood and fresh medium to adjust to a 4% hematocrit.

Approximately 7-8 hours after MACS synchronization, the parasites had re-infected and entered the ring stage. At this point, parasites were re-synchronized by treatment with 5% sorbitol (w/v). Parasite growth and parasitemia were monitored by evaluating Giemsa-stained blood smears under the microscope. Tightly synchronized ring (approximately 16 hours ± 4 hours post-invasion), trophozoites (approximately 26 hours ± 4 hours post-invasion), and schizonts (approximately 36 hours ± 4 hours post-invasion) were harvested following established protocols [91]. Parasites were isolated by lysing the red blood cells in 0.1% saponin followed by thorough washing in PBS. The subsequent pellets were weighed and proteins were immediately extracted by lysis in an 8M urea lysis buffer supplemented with protease and phosphatase inhibitors (8M urea, 75mM Tris, pH 8.2, 1X HALT protease inhibitor, 1X HALT phosphatase inhibitor) [92]. Lysates were cleared by
centrifugation for 10 minutes at 4°C at 14,000 rpm. Protein concentration was determined for each sample by BioRd Bradford Assay. Duplicate samples were prepared for each developmental stage.

### 2.2.2 Sample Preparation and Mass Spectrometry

Equal quantities of duplicate protein extracts (1mg) from each of the three intraerythrocytic *P. falciparum* stages were reduced, alkylated, and digested with Lys-C prior to labeling with one of 6 TMT isobaric labeling reagents [87]. All 6 samples were then mixed together and separated via strong cation exchange chromatography and a total of 20 fractions were collected. Each of the 20 fractions was then analyzed via LC-MS/MS on an LTQ-Orbitrap-Velos mass spectrometer essentially as described previously [93] except that high-resolution MS/MS spectra were acquired for all precursors and used for both identification and quantification. Individual peptides were identified using Sequest [94] and proteins were filtered to a false discovery rate (FDR) of 1% as described previously [95]. Database searches employed a composite database containing *Plasmodium falciparum* sequences obtained from NCBI [96] as well as human protein sequences derived from the International Protein Index [97] and common contaminants such as Lys-C; all sequences were included in forward and reverse orientations. For quantification, peptides matching each protein were grouped and their MS2 TMT reporter ion intensities were scaled to a sum of 1.0. An ANOVA model was used to identify differentially expressed proteins based on the quantification of individual peptides. The Benjamini-Hochberg approach was used to correct for multiple hypothesis testing.
2.2.3 Phosphopeptide Enrichment and Identification

Following protein extraction, samples containing 1mg of protein per stage were reduced, alkylated and digested with Lys-C and subsequently labeled with one of 6 TMT isobaric labeling reagents. The resulting peptides were fractionated via SCX (strong cation exchange chromatography). The resulting 20 fractions were subjected to immobilized metal affinity chromatography (IMAC) to specifically pull out phosphorylated peptides [92]. Phosphopeptides were analyzed via LC-MS/MS on an LTQ-Orbitrap mass spectrometer essentially as described previously [93], except that all MS/MS spectra were acquired at high resolution in the Orbitrap and used for both phosphopeptide identification and quantification. Peptides were identified by matching individual MS/MS spectra with peptide sequences using Sequest [94]. As before, searches employed a database containing *Plasmodium* sequences from NCBI [96] as well as human protein sequences from IPI [97]. The peptides were filtered using a multivariate approach to remove questionable identifications based on a target-decoy strategy [95, 98]. Peptides were assembled into proteins and proteins were further filtered to ensure a protein-level false discovery rate of 1%. Individual phosphorylation sites on each peptide were scored to assess how confidently the modification could be localized to a single residue using AScore [99]. Finally, the peptides and proteins were filtered to account for protein redundancy.

For quantification, TMT reporter ion signals were extracted from MS/MS spectra and normalized assuming equal protein loading in each channel. Each phosphorylation site was quantified by gathering together all matching peptides, filtering out peptides that didn’t meet the quantification standards, and re-scaling reporter ion intensities so that they summed to 1.0.
2.2.4 Western Blotting with Phospho-Specific Antibodies

*Plasmodium falciparum* 3D7 strain parasites were grown as described above. Ring stage parasites were synchronized by treatment with 5% sorbitol (w/v). Parasite growth and parasitemia were monitored for a 48-hour recovery period. After the parasites progressed through an entire intraerythrocytic cell cycle, parasites were harvested every 8 hours for 48 hours following established protocols [91]. Parasites were isolated by lysing the red blood cells in 0.1% saponin (w/v), followed by thorough washing in PBS. The subsequent pellets were weighed and recorded. Proteins were immediately extracted by lysis in an 8M urea lysis buffer as described above and centrifuged for 10 minutes at 4°C at 14,000 rpm. A BioRad Bradford Assay was used to determine the protein concentrations of the cleared lysates. The protein lysates were then stored at -80°C until use.

For Western blotting, 100μg of protein extract from each time point were loaded on a NuPAGE 4-12% Bis-Tris gel with NuPAGE MOPS SDS Running Buffer and run at 100 volts for approximately 3 hours following the manufacturer’s instructions (Invitrogen). After electrophoresis, the gels were rinsed in transfer buffer (NuPAGE 1X Transfer Buffer) and transferred onto nitrocellulose using the NuPAGE wet transfer module at 30 volts for 1 hour at 4°C according to the manufacturer’s instructions (Invitrogen). The membranes were blocked in 5% BSA in tris-buffered saline containing 0.1% Tween 20 for 1 hour at room temperature. The membranes were then incubated with the relevant antibody from Cell Signaling Technology (Phospho-Tyrosine Mouse mAb (P-Tyr-100) #9411 1:2000 and Phospho-MAPK/CDK Substrates (PXSP or SPXR/K) (34B2) Rabbit mAb #2325 1:1000) overnight in 5% BSA tris-buffered saline containing 0.1% Tween 20 at 4°C following the suggested dilutions. The membranes were washed, incubated with horseradish peroxidase-conjugated secondary antibody at a 1:2000 dilution for 1.5 hours at room temperature,
washed, and developed using enhanced chemiluminescence detection according to the manufacturer’s instructions (Cell BioSciences).

2.2.5 Immunoprecipitation with PTMScan Direct

The PTMScan Direct protocol is adapted from the PhosphoScan® method developed at Cell Signaling Technology with licensed use [100].

Cell Lysate Preparation- Parasite extracts were prepared as described above. The protein pellets were lysed in urea lysis buffer (20mM HEPES, pH 8.0, 9M urea, 1X HALT protease inhibitor, and 1X HALT phosphatase inhibitor). A total of 45mg of protein was used for each experiment. The pellets were re-suspended in the lysis buffer and incubated on ice for 10 minutes with intermittent vortexing. The cells were sonicated two times for 30 seconds each at 15-W output with 1 minute cooling on ice between bursts. Lysates were centrifuged at 20,000 rpm for 15 minutes at 15°C. The supernatant was collected in a fresh 15mL conical tube and the concentration was determined via Bradford Assay. The cleared supernatant was diluted with lysis buffer to a concentration of 4mg/mL to avoid protein aggregation problems. Supernatants were then reduced with 1/10 volume of 45mM DTT for 20 minutes at 60°C and cooled on ice for 10 minutes. Reduced lysates were alkylated with the same volume of 110mM iodacetamide as the DTT and incubated in the dark at room temperature for 15 minutes. The samples were diluted 1:4 with 20mM HEPES, pH 8.0 and methanol-chloroform precipitation of proteins was performed. Briefly, 4X the volume of 100% methanol was added to the sample, mixed well, and centrifuged at 9,000 rpm for 10 seconds at room temperature. Then 1X the volume of chloroform was added, mixed well, and centrifuged at 3,500 rpm for 1 minute followed by 3X the volume of HPLC grade water, mixed well, and centrifuged at 4,000 rpm for 14 minutes. The upper layer was then removed and discarded. Next,
3X the volume of methanol was added, mixed, and centrifuged at 9,000 rpm for 2 minutes. The supernatant was removed and discarded, while the pellet was allowed to air dry. The pellet was re-suspended in 4M urea, 50mM Tris-HCl, pH 8.8 buffer at a concentration of 3mg/mL. The sample was then diluted to 1M urea with 50mM Tris-HCl, pH 8.8 and digested overnight at room temperature with LysC (Wako 129-02451 10AU) in a 1/100 enzyme/protein ratio. Digested peptide lysates were desalted over Sep-Pak® C18 Columns, volume 0.7mL (Waters WAT05910). Peptides were eluted with 40% acetonitrile, 1% FA, lyophilized for two days and then stored at -80°C.

Immunoprecipitation- Lyopholized peptides were re-suspended in 1.4mL of 1X IAP buffer and centrifuged at 1,800 rpm for 5 minutes. The resulting supernatant was pipetted directly into a slurry containing PTMScan® Immunoaffinity beads and mixed overnight at 4°C. The beads were then pelleted by centrifugation for 30 seconds at 1,500 rpm and washed twice with 1mL of 1X IAP buffer as well as three times with 1mL HPLC grade water. Peptides were eluted from the beads with 55μL of 100mM FA for 10 minutes at room temperature, followed by a second elution of 45μL of the same solution. The sequential elutions were combined and desalted as previously described [100]. Finally, samples were re-suspended in 5% ACN/5% FA prior to MS analysis.

Mass Spectrometry Analysis- All samples were analyzed via LC-MS on an LTQ-Velos-Orbitrap mass spectrometer configured as described previously [93] and operated in data-dependent mode with low-resolution ion trap mass spectra acquired for the top ten precursors identified in each MS survey scan. Data were then searched using Sequest [94] and resulting peptide and protein identifications were filtered to a final protein-level FDR of 1% using the target-decoy approach as described previously [95, 98]. Site localization was evaluated via AScore [101] and finally phosphopeptides were grouped according to the phosphorylation sites they contained.
2.2.6 Subdomain Location of Phosphorylation Sites in Protein Kinases

Putative kinases were selected from the table of all phosphorylated proteins based on the appearance of the string “kinase” in the sequence description. From these, we identified and aligned 30 eukaryotic protein kinase domains using the program MAPGAPS [102] and sequence profiles designed for apicomplexan protein kinases [103], while additional sequences representing atypical kinases or other kinase-associated proteins were removed from this sequence set. The alignment regions corresponding to each sub-domain were identified according to the definition by Hanks and Hunter [24]. An in-house Python script was used to map each observed phosphorylation site to a subdomain or tail region, count the number of phosphorylation sites in each region, and visualize the totals.

2.2.7 Matching Phosphorylation Sites with Kinases Using NetPhorest

Sequences of likely eukaryotic protein kinases in the NCBI Plasmodium proteome were uploaded to the public web server for NetPhorest [104] along with lists of phosphorylated residues. The probabilities associating each phosphorylation site with particular kinases and kinase families were then downloaded and sorted in order of decreasing probability. The top 2.5% of putative kinase-substrate relationships were retained and plotted as a heat map reflecting the probability that the indicated kinase or kinase family would be expected to phosphorylate the indicated residue.
2.3 Results and Discussion

2.3.1 Identification of 2,767 Proteins and 6,293 Phosphorylation Sites from the Intraerythrocytic Stages of Plasmodium falciparum

To perform a comprehensive analysis of protein expression and phosphorylation across the intraerythrocytic stages of *Plasmodium falciparum*, we used isobaric labeling for quantitative analysis of Lys-C digested parasite extracts prepared from tightly synchronized ring (16 hours ± 4 hours), trophozoite (26 hours ± 4 hours), and schizont (36 hours ± 4 hours) stages. The phosphoproteomic work flow is outlined in Fig. 5A. This quantitative proteomic analysis identified a total of 3,365 proteins from both the proteomic and phosphoproteomic datasets. Of these proteins, 451 were human-derived, likely from contaminating RBC material, while 2,914 were of malaria origin, representing 55% of the total predicted *P. falciparum* proteome (see below) (Fig. 6) [20]. After human proteins were excluded, the final data set contained 2,767 proteins obtained in the absence of phosphopeptide enrichment (“Proteome” table in Fig. 5B) and 1,337 phosphoproteins (out of which 1,190 were also found in the proteome analysis; see discussion below) containing a total of 6,293 phosphorylation sites, obtained from the phosphoproteome analysis (Fig. 5B) at a protein-level false discovery rate (FDR) of <1%. Fig. 5C displays a histogram showing how the quantified peptide count varied across quantified proteins; at least two peptides were used for quantification of 90% of the identified proteins, while the average number of peptides per identified protein was 15. The observation that the vast majority of proteins were both identified and quantified on the basis of several peptides enhances the reliability of this analysis. We identified a total of 11,881 phosphopeptides bearing 6,293 phosphorylation sites from 1,337 proteins. In contrast, another recent report [76] identified 7,835 phosphopeptides matching 8,463
phosphorylation sites on 1,673 proteins. Importantly, in addition to identification, in this study we provide quantitative analysis of the phosphorylation changes for the first time for 4,731 sites, across the intraerythrocytic cell cycle (Fig. 5B).

Since many proteins of the *Plasmodium* proteome are exclusively expressed in the mosquito stages [84, 85], and many others are restricted to liver and gametocyte stages during infection of the human host, we propose that our present dataset likely encompasses the majority of intraerythrocytic stage-specific proteins. Our data significantly enhances the number of *Plasmodium* identified proteins observed. Several reasons underpin the greater depth of our analysis compared to previous reports [79, 83, 91]. Firstly, by simultaneously surveying three developmental stages, each with its own protein complement, our study encompasses a larger fraction of the proteomic landscape. Secondly, the continuous improvements in methodology and instrument sensitivity naturally result in an increased number of protein and phosphorylation identifications [92].

An intriguing aspect of intraerythrocytic development of *Plasmodium* is the mechanism by which parasites regulate gene expression. Based on transcriptome analysis, it was initially believed that the majority of genes are globally controlled in a stage-dependent manner at the transcriptional level [105, 106], but it is now generally accepted that post-transcriptional and post-translational regulatory mechanisms have dominant roles in regulation of intraerythrocytic gene expression [79, 80]. To further understand the importance of post-transcriptional regulation in *P. falciparum* we compared the dynamics across the *Plasmodium* proteomic landscape to those of the published transcriptome [105]. Only the sorbitol-synchronized ring, trophozoite, and schizont transcripts were compared to our proteomic data, as this transcriptome profile is closest to our dataset in terms of sample preparation. Of the 2,767 proteins identified in our study, 429 were not present in the...
transcriptomics [105] data set. For a large percentage of genes (57.1%), timing of peak mRNA expression does not coincide with maximal protein abundance (Fig. 5D). In some instances, such as Early Transcribed Membrane Protein 10.1 (PF10_0019) and the putative protein kinase PfTKL4 (PFF1145c), the delay in protein expression is significant: transcription peaks in segmenters or in rings, while the protein expression peaks in schizonts, suggesting that these mRNAs have a long half-life and are not translated until the late trophozoite stage. Analysis of mRNA decay in *Plasmodium* supports a long half-life of the mRNAs for these proteins [107]. Overall, our analysis is in line with previous findings [79, 80] that post-transcriptional regulation of gene expression plays a major role in *Plasmodium*.

### 2.3.2 Developmental Stage-Specific Distribution of Proteins

To determine the dynamics of protein expression during the intraerythrocytic cell cycle, we first compared expression levels by calculating three ratios across the three stages (trophozoite: ring, schizont: ring, and schizont: trophozoite). Proteins exhibiting greater than 1.5-fold difference in abundance between any two stages, with an adjusted p value ≤0.05, were considered enriched in that particular stage, while proteins showing less than 1.5-fold changes for all pair-wise comparisons, or an adjusted p value >0.05, were considered globally expressed [108]. Several abundant proteins enriched in a specific stage of intraerythrocytic development are shown in Table 1A. For example, these proteins include merozoite surface protein 1 (enriched in the ring/schizont sample), Pfmc2-2TM Maurer’s Cleft Two Transmembrane protein (enriched in the trophozoite sample), and ring-exported protein 1 (enriched in the trophozoite/schizont sample). Pfmc2-2TM Maurer’s Cleft Two Transmembrane protein is a Maurer’s Cleft associated protein that has been implicated in transporting proteins secreted from the parasite to the erythrocyte surface [109].
Previously, its transcription has been observed during the trophozoite stage [105] and our study shows peak protein expression also at the trophozoite stage. This pattern suggests that Pfmc2-2TM may have a role in the translocation of secreted proteins during trophozoite maturation.

A total of 1,832 proteins were globally expressed, while 935 showed a significant change in expression among the intraerythrocytic stages (Fig. 7A). Of these, 353 (12.8%) proteins were enriched in the ring stage, 59 (2.1%) were enriched in the trophozoite stage, and 86 (3.1%) were enriched in the schizont stage (Fig. 7B). The large proportion of proteins exhibiting variable expression across stages demonstrates that the *P. falciparum* proteome undergoes significant changes during the intraerythrocytic cell cycle. Next, we analyzed the expression pattern of the 33.8% of the proteome that fluctuates across the intraerythrocytic cell cycle (Fig. 7C and 7D). Of the constitutively expressed proteins, approximately 65% are hypothetical or unclassified proteins, which is not significantly different from their prevalence in the predicted proteome. Expectedly, a large portion of the recognizable constitutively expressed proteins has general housekeeping and structural functions such as energy, metabolism, protein synthesis, and transcription.

2.3.3 Stage-Specific Analysis of Functional Profiles

In an effort to understand the functional classes of identified proteins, we sorted them using the Munich Information Centre for Protein Sequences (MIPS) catalog, with some adaptations for classes specific to the parasite, such as proteins pertaining to malaria pathogenesis (Fig. 8) [110]. The distribution of the functional profiles is summarized in Fig. 8. Forty percent of the identified proteins were *P. falciparum* hypothetical proteins. We made an effort to categorize the hypothetical proteins into functional classes where possible, based on sequence homology to orthologs in other organisms (Fig. 9). Proteins belonging to important functional categories such as cell cycle and
DNA processing (2%), cellular transport (5%), interaction with the environment (3%), metabolism (4%), protein fate (5%), protein with binding function (9%), and transcription (4%) account for 32% of the total proteins identified. Many of these classes of proteins are necessary for invasion, proliferation, defense, and cell communication, which are critical for pathogenesis of the parasite. Functional classification revealed some marked differences between intraerythrocytic stages. More specifically, the “pathogenesis” functional category was not represented in the ring and trophozoite stage proteomes, but reaches approximately 7% in the schizont stage. Furthermore, <1% of globally expressed proteins are classified as functioning in pathogenesis. This increased enrichment of proteins involved in the pathogenesis functional category at the schizont stage suggests the parasite is preparing for rupture from the red blood cells and invasion of new red blood cells to continue the pathogenic process. Another fascinating observation involves the “transcription” functional category, which encompasses 2% of the ring stage proteome and 2.3% of the schizont stage proteomes; however, this number significantly increases to 8.5% in the trophozoite stage (p<0.05, Chi-Squared Test). Interestingly, “protein expression” protein levels is at its lowest during the trophozoite stage, while proteins involved in transcription reach their highest level at the trophozoite stage. Also, there are no proteins involved in protein synthesis with peak expression at the trophozoite stage. This implies the parasite’s emphasis on transcription and not protein synthesis during the trophozoite stage. This also suggests of an effort to conserve energy for subsequent schizogony and an emphasis on post-transcriptional regulation.

Proteins were further characterized by assigning GO SLIM terms using “Generic GO” [85] (http://go.princeton.edu/cgi-bin/GOTermMapper). A Plasmodium-specific GO analysis for “Molecular Function” and “Cellular Component” was performed using collected PlasmoDB identifiers. All proteins with PlasmoDB identifiers were compared to P. falciparum proteins within
each molecular function category and cellular component category. The high percentage of proteins identified across multiple “Molecular Function” categories and “Cellular Components” demonstrates the depth and comprehensiveness of this analysis (Fig. 10A and Fig. 10B).

2.3.4 Developmental Stage-Specific Distribution of Phosphorylation Sites

Because reversible protein phosphorylation is an important regulator of protein activity, we analyzed how protein phosphorylation fluctuates with the progression of the *Plasmodium* asexual life cycle. Table 1B includes examples of abundant stage-specific phosphorylation events with the most significant p values. We examined numbers of phosphorylated peptides, numbers of unique sites, and the numbers of phosphoproteins identified in the three intraerythrocytic stages. Of the identified phosphoproteins, 452 were differentially expressed and 758 were globally expressed (Fig. 5B). The number of phosphorylated peptides, unique phosphorylation sites, and total phosphoproteins varied among the intraerythrocytic stages (Fig. 11A), reflecting the likely role of protein phosphorylation in regulating developmental progression. The highest number of phosphopeptides and phosphorylation sites was found at the schizont stage, while maximum expression of phosphopeptides was detected at the ring stage. This finding underscores the importance of protein phosphorylation in controlling specialized signaling networks that parasites employ to progress through the intraerythrocytic cell cycle. A similar trend has been found in the yeast centrosome phosphoproteome: only 54 phosphorylation sites were detected during the G1 stage, which is analogous to the ring stage in *Plasmodium*, but almost twice as many were detected during mitosis, which is comparable to the *Plasmodium* schizont stage [111].

Seventy-five percent of phosphoproteins identified in this study contained multiple phosphorylation sites (Fig. 11B); the majority are phosphorylated on three or more residues (Fig.
11B). Phosphorylation on multiple residues suggests that a protein’s activity and interactions are regulated via phosphorylation at distinct sites. Phosphorylation sites were scored using the AScore algorithm to assess the confidence of site localization [101]. Any site with an AScore above 13 was considered correctly localized. Ninety percent of the sites were localized to a single amino acid, and 87%-93% of sites (depending on stage) were localized (Fig. 11C).

Phosphorylation was most abundant on Ser (82%), followed by Thr (13.5%), and Tyr (4.5%). The enriched phosphorylation of Ser residues surpasses its relative abundance among all residues that are subject to phosphorylation (Fig. 11D), indicating a strong bias towards Ser phosphorylation. The observed 4.5% of tyrosine phosphorylation in this study is similar to the typical ~1%-4% tyrosine phosphorylation seen in other organisms [95, 112] which is somewhat surprising given the absence of true tyrosine kinases in *Plasmodium* [21, 76, 113-115]. The observed 4.5% tyrosine phosphorylation is higher than that reported earlier [76, 77, 116], at least in part because our dataset includes the ring and trophozoite stages. Stage-specific tyrosine phosphorylation levels are moderate at the ring stage (14.8%), decrease at the trophozoite stage (2%), and increase to peak levels during the schizont stage (27.6%) (Fig. 11F). Although many proteins were phosphorylated on multiple residues, only a small portion of Ser, Thr, and Tyr residues within each phosphoprotein were modified (Fig. 11E).

Although tyrosine kinase activity in *Plasmodium* has been documented through biochemical, pharmacological, and immunological approaches [117-120], the parasite’s kinome is devoid of members of the tyrosine kinase group [21]. Two dual-specificity phosphatases, YVH1 and PRL, have been identified in *Plasmodium* [115, 120, 121], which probably account for at least some of the observed Tyr phosphorylation. These reports also allude to the existence of reversible Tyr phosphorylation in malaria parasites.
We detected 226 Tyr phosphorylated proteins. Tyrosine phosphorylation of such a large number of *Plasmodium* proteins in the absence of true tyrosine kinase homologues underscores the importance of dual-specificity kinases in the parasite. Recent reports have identified tyrosine auto-phosphorylation in the activation loops of PfGSK2 and PfCLK3, dual-specificity CMGC kinases [76, 77]. Tyr auto-phosphorylation activities by mammalian “dual-specificity tyrosine phosphorylation-regulated kinases” (DYRK) have been observed [122]. Recently, the *Plasmodium* NIMA-like kinase Pfnek3 has also been shown to have tyrosine auto-phosphorylation activity [123]. Characterization of *Plasmodium* kinases with trans-Tyr phosphorylation ability will be an area of considerable interest.

We examined each phosphorylation site to define those with a >1.5-fold increase or decrease between stages. Fig. 11F illustrates the stage-specific distribution of serine, threonine, tyrosine and overall phosphorylation. It is interesting to note that although the overall phosphorylation of serine, threonine, and tyrosine residues varies greatly between the stages, the distribution of phosphorylation between these residues is relatively stable as the parasite progresses through the developmental cycle. Phosphorylation of 25% of the sites did not significantly fluctuate, while 75% of the sites were stage-dependent (Fig. 11G and 11H). Our data indicate that each stage of the *P. falciparum* intraerythrocytic cell cycle displays a distinct phosphoproteome (Fig. 11G and 11H). Peak protein phosphorylation was observed in the schizont stage, suggesting that reversible protein phosphorylation is presumably a regulatory feature of nuclear division and merozoite ontogeny. The importance of phosphorylation in merozoite invasion has been demonstrated previously [21, 124]. More recently, the phosphorylation of glideosome motor components and other proteins with known involvement in parasite egress and invasion have been
documented, supporting the involvement of phosphorylation in egress and invasion regulation [116].

2.3.5 Distribution of Phosphorylation Site Classes and Phosphorylation Motifs Across P. falciparum Stages and Phosphoproteins

The phosphorylation sites were analyzed to determine the kinase classes responsible for the modification. We used a decision tree approach to classify each site as acidic, basic, proline-directed, tyrosine, or other based on the identities of surrounding amino acids [92, 95]. Acidic sites were the most common (36%) followed by basic (29%), while tyrosine and proline-directed sites were the least common (4% each) (Fig. 12A). Approximately 60% of phosphoproteins contained phosphorylation sites from multiple kinase classes, whereas 2% of phosphoproteins were predicted to be targets for all kinase classes (Fig. 12B). Variations were detected in the frequencies of these classes across the developmental stages, suggesting that distinct kinases operate preferentially at various stages. Overall, site classes were represented to a much lesser extent during the early stages (ring and trophozoite), where phosphorylation levels were the lowest, and increased drastically as the parasite progressed into the schizont stage (Fig. 12C). Proline-directed sites that are utilized by CDKs and MAP kinases had the most significant increase at the schizont stage (Fig. 12C).

The P. falciparum genome is extremely A:T rich, which results in an unusual amino acid bias in the parasite’s proteome. Our analysis suggests that P. falciparum codon bias has led to unique phosphorylation motifs that direct its kinase specificity. The phosphoproteomic data enabled us to identify parasite-specific phosphorylation motifs using the MotifX algorithm [125]. To this end we inputted all the identified phosphorylation sites along with six amino acids N- and C-terminal of the phosphorylated residue, generating a list of motifs that were enriched in the
phosphorylation sequences compared to randomly selected Ser, Thr, and Tyr residues. A total of 43 distinct phosphorylation motifs were identified. Of these, 33 motifs related to phosphorylated Ser residues and 10 motifs to phosphorylated Thr residues. Interestingly, some of the identified motifs are unusual. For example, there are 14 motifs with prominent Asn residues, in line with an earlier report for the schizont phosphoproteome [76]. The overall trends of phosphorylation for the identified Ser, Thr, Arg, Asn, Gly, and Pro-residues containing motifs resemble the general trend of phosphorylation. One notable difference is the substantial increase in phosphorylation at the trophozoite and schizont stages for proline-containing phosphorylation motifs (Fig. 12D).

Our data provides an opportunity for the discovery of substrate motifs that may point to specific kinases, which will help to delineate the parasite signaling networks. We predict possible kinase-substrate relationships by comparing phosphorylated peptide sequences to protein Ser/Thr kinase consensus phosphorylation sequences. Fig. 12E lists five putative substrates with consensus phosphorylation sequences for known human Ser/Thr kinases. In addition to proline-containing CDK-like motifs, we found motifs similar to those for mammalian GSK3, TGF-βR1, and numerous other candidates. We also identified some previously undescribed motifs that are likely to be specific for as of yet unknown kinases.

2.3.6 Phosphorylation Profiles of Putative MAPK/CDK Substrates and Tyrosine Phosphorylated Proteins

The intraerythrocytic development of the malaria parasite diverges from the eukaryotic cell cycle paradigm that has emerged largely from studies in yeast cells. The six identified CDK-related protein kinases [21, 112] are likely key regulators of cell cycle progression. To validate this prediction and gain insight into physiological functions of Plasmodium CDK-related kinases, it is
necessary to identify their cellular substrates. CDKs are proline-directed serine/threonine kinases having a strong requirement for basic amino acids, Arg and Lys in particular, at the +3 site [126]. Filtering 206 proline-directed Ser/Thr phosphorylated proteins for those with the strong CDK consensus pS/T-P-X-K/R motif identified 45 proteins that are putative *Plasmodium* CDK substrates. Among the proteins that may be phosphorylated by *Plasmodium* CDKs are homologs of origin recognition complex subunit 2 (Orc2) (MAL7P1.21), Orc4 (PF13_0189), and Orc1/CDC6 (PFE0155w), components of the pre-replication complex that are known CDK substrates in other eukaryotes. Additional pre-replicative complex proteins that are phosphorylated at the weak CDK consensus motif (pS/T-P) are MCM4 and MCM5. Mammalian ORC and MCM subunits are phosphorylated by CDKs to prevent helicase loading beyond G1 phase of the cell cycle, thus preventing re-replication [127, 128]. Other proteins of significance in potentially regulating the parasite’s DNA replication, transcription, or mitosis that are phosphorylated at the strong CDK motif are homologs of cell division cycle protein 48 (PFF0940c) and regulator of chromosome condensation (MAL7P1.38).

Developmental stage-specific protein phosphorylation profile analysis using phospho-MAPK/CDK substrates antibody demonstrated minimal phosphorylation at the ring stage and maximal phosphorylation at the schizont stage (Fig. 13A). This correlates well with our protein phosphorylation analysis of the 45 putative MAPK/CDK substrates, concurring that CDK activity is predominant in late stages of the asexual cycle (Fig. 13B). Using an antibody-based pull-down of phospho-motif peptides (PTMScan® Direct Cell Signaling Technology) in duplicate experiments, we identified a total of 48 unique phosphorylation events. Significantly, the PTMScan® Immunoprecipitation identified two phosphorylation events, homologs of cell division cycle protein 48 (PFF0940c) (S512) and regulator of chromosome condensation (MAL7P1.38) (S602) that were
predicted MAPK/CDK substrates from the global phosphoproteomic analysis. Furthermore, 7 of the identified phosphopeptides identified in the PTMScan® Immunoprecipitation were also identified in the global phosphoproteomic analysis.

We used a phospho-tyrosine specific monoclonal antibody (P-Tyr-100) for western blot analysis of stage-specific extracts (Fig. 13C) to demonstrate that tyrosine phosphorylation events also fluctuate with cell cycle progression (Fig. 13D). It is evident from these data that phosphorylation of tyrosine residues is quite common in Plasmodium. Therefore, we examined the sensitivity of the parasite to the common tyrosine kinase inhibitors Genistein, Herbimycin, PP1, and PP2. The observed IC$_{50}$ values were as follows: Genistein 83μM, Herbimycin 0.3μM, PP1 2.4μM, and PP2 0.9μM (unpublished data), which reveals that P. falciparum is quite sensitive to tyrosine kinase inhibitors. Herbimycin and PP2, inhibitors of Src-family kinases, exhibited submicromolar potency. Recent data implicating host erythrocyte signaling pathways in parasite survival [129] suggest that some of these effects may be mediated by host tyrosine kinases. Our data suggest that tyrosine kinases may prove to be promising targets for future malaria therapeutics developments.

A characteristic of Plasmodium intraerythrocytic cell cycle is the formation of multinucleated schizonts through multiple nuclear divisions without cytokinesis [17]. In addition to CDKs, Aurora kinases play important roles in the regulation of mitosis in eukaryotes; thus we identified putative substrates of Plasmodium Aurora kinase A homologues using an Aurora kinase A phosphorylation motif, (K/R)-(K/R)-X-(pS/pT)-L. Our data filtering identified proteins with a MORN-repeat, Kid domain, and MIF4G domain among 28 putative Plasmodium Aurora kinase A substrates. A MORN repeat-containing protein has been shown in another apicomplexan parasite, Toxoplasma gondii, to be a component of cell division machinery [130].
2.3.7 Sub-domain Distribution of Phosphorylation Sites within Protein Kinases

Because a common regulatory mechanism of eukaryotic phosphosignaling pathways is the auto-phosphorylation of key regions of the kinase domain itself, we investigated the subdomain distribution of phosphorylation in *Plasmodium* kinases. Thirty of the phosphorylated proteins from our phosphoproteomic data set were identified as eukaryotic protein kinases (ePKs). In these proteins we identified the location of each of the 126 observed phosphorylation sites relative to the kinase domain, whether in the kinase domain, the N-terminal tail, or the C-terminal tail. Those in the kinase domain were further assigned to a sub-domain location (Fig. 14A). The majority of phosphorylation sites, 71 in total, appear in the N-terminal extension of the kinase domain, while 25 appear in the C-terminal tail and 30 appear in the kinase domain itself. This apparent enrichment in the N-terminal extension region is mostly explained by the presence of extended N-segments in many of the *P. falciparum* protein kinases.

When examining the stage-specific expression profile of the 30 identified ePKs, we observed peak expression at the ring stage (Fig. 14B). This is in contrast to the stage-specific profile of phosphorylation of these kinases, which exhibits peak phosphorylation at the schizont stage (Fig. 14B). Furthermore, the percentage of ePKs with global expression is large (30%) in comparison to the 11% of global phosphorylation events. Fig. 14C depicts the stage-specific distribution of the number of phosphorylation sites at a given time during the intraerythrocytic cell cycle for each of the kinase sub-domains, as well as the N-terminal and C-terminal tails, where overall phosphorylation levels are the highest at the trophozoite and schizont stages.

The protein kinase sub-domains VII and VIII contain the activation loop, a structural region whose phosphorylation is a key step in the activation of many kinases [131]. We observed activation loop phosphorylation in three kinases: PfCDPK1 (PFB0815w), PfCRK3 (PFD0740w),
and an orphan kinase (PF11_0227). Of these, the single phosphorylation sites on PfCDPK1 and PfCRK3 are consistent with a typical mechanism of kinase activation by phosphorylation of the activation loop, while the orphan kinase PF11_0227 contains a greatly extended activation loop of 135 amino acids, of which three were phosphorylation sites, suggesting the possibility of a novel functional elaboration.

Phosphorylation in sequence regions between the defined sub-domains was observed in two kinases, FIKK4.1 (PFD1165w) [132] and PfPK1 (PF08_0044), which is related to glycogen synthase kinases [103, 133] (Fig. 14A). The N-terminal region of subdomain VIb, which includes the catalytic loop and beta-7 strand, was phosphorylated in two other kinases, PfSRPK1 (PFC0105w) at 6 sites and the orphan kinase PF14_0392 at 3 sites. In PfSRPK1 this is the site of a large insert known as the spacer domain, which influences localization of the protein in yeast and human SRPKs [134]. Given that the spacer region itself if not required for PfSRPK1 activity [135], the extensive phosphorylation of the spacer region of PfSRPK1 may instead be a mechanism controlling localization or binding of PfSRPK1.

While tyrosine phosphorylation is rare overall in the P. falciparum phosphoproteome [76-78], those occurrences are noteworthy as they may point to novel signaling pathways or mechanisms in the parasite. We observed only one phosphorylated tyrosine site within the kinase domain, in subdomain IV of PfCRK4 (PFC0755c). Structurally, this is the beta-4 sheet in the ATP-binding N-lobe of the kinase domain; PfCRK4 was also phosphorylated on serine or threonine residues at 6 other sites in subdomain IV. Four more instances of tyrosine phosphorylation were observed in the N-terminal tails of PfCRK4, PfCRK1 (PFD0865c), PfPKG (PF14_0346), and the orphan kinase PF14_0392. In previous studies tyrosine phosphorylation has been observed in the
activation loop of dual-specificity kinases such as PfGSK3 and PfCLK3 [77]. However, we did not find any instances of tyrosine activation loop phosphorylation in this study.

2.3.8 Preliminary Identification of Kinases Responsible for Kinase Domain Phosphorylation

The activity of each kinase in signaling cascades is determined in part by specific reversible phosphorylation events. Since these phosphorylation events govern the activities of these kinases, identifying the kinases responsible can reveal the phosphorylation networks that regulate \textit{Plasmodium} biology. We used a computational approach to predict those kinases that are likely responsible for observed kinase phosphorylation events. Though such predictions on their own do not prove particular kinase-substrate relationships, they nevertheless can provide working hypotheses that can be tested experimentally. We used the NetPhorest algorithm [104] to identify candidate kinases that may be responsible for the 126 phosphorylation events we observed upon 30 \textit{Plasmodium} kinases. The 2.5% most likely kinase-substrate pairings are depicted in Fig. 15 for all phosphorylation sites observed on \textit{Plasmodium} kinases in conjunction with structural and quantitative data.

Although 22 kinases and kinase families are associated with the 126 kinase phosphorylation sites, the most confident predictions implicated kinases of the CK2 family (similar to Casein Kinase 2). Sites on nearly a dozen kinases were located in highly acidic regions that were judged to be favorable for CK2 activity; essentially all of these sites reached maximum abundance during the schizont stage, which is consistent with shared regulation. In contrast, two sites on separate kinases bear a signature SQ motif that are predicted to be substrates for ATM/ATR kinases; however, while one of these sites reaches maximum expression during the schizont stage, the other peaks during the ring stage and is nearly completely absent during the trophozoite stage. Finally, we observed a
single phosphorylation site on NEK1, which was scored as a candidate substrate for DMPK, PAKB, or a NEK-family kinase: thus, this site could potentially be a case of auto-phosphorylation.

### 2.3.9 Peak Protein Expression and Phosphorylation are Unlinked

Differential levels of phosphopeptide detection can arise from changes in protein level, or from genuine changes in the phosphorylation occupancy of the target site. In order to differentiate between these two possibilities, we compared the proteomic data of the intraerythrocytic stages to our phosphoproteomic analysis. Overall, we identified 2,914 proteins, 40% of which were identified both with and without enrichment of phosphopeptides (Fig. 16A). A similar overlap was observed across developmental stages (Fig. 16B). The proteins that are identified solely on the basis of phosphorylation data are those of low abundance. Very low abundance proteins are often difficult to detect using the shotgun approach that we used, but when they are phosphorylated they can often be identified following phosphopeptide enrichment.

Phosphorylated and non-phosphorylated protein profiles varied throughout the asexual cell cycle (Fig. 16C). Seventy-five percent of proteins with stage-specific phosphorylation sites were expressed across multiple intraerythrocytic stages in the non-phosphorylated form (Fig. 16D). In general, protein expression and phosphorylation display obvious differences throughout the *P. falciparum* intraerythrocytic cell stages. Malaria parasite protein expression can be classified as a global event (66.2%) more frequently than is observed for protein phosphorylation (25%). Among those proteins that are variably expressed across developmental stages, a significant number attained peak expression during ring stage in contrast to the phosphorylation, which peaks at the schizont stage, demonstrating that peak phosphorylation events are generally not due to stage-specific protein expression.
Proteins and phosphorylation sites that showed statistically significant differences in protein abundance, phosphorylation levels, or both across the three growth stages were classified via K-medoids clustering into groups with similar patterns of both protein expression and phosphorylation. The silhouette method was then used to determine the optimal number of clusters for the data, which showed 7 clusters in our case. Fig. 17A summarizes numbers of proteins and phosphorylation sites that were assigned to each of the 7 clusters, accompanied by the trends observed for proteins and sites in each group. Also, heat maps were generated for each cluster to illustrate changes in protein expression and phosphorylation. The results suggest that protein abundance and phosphorylation levels change independently of one another. There are relatively few cases where both change in tandem, while there are many cases where phosphorylation increases while protein abundance stays the same or decreases. For example, in cluster PEP3, there are 1,526 phosphorylation sites with associated protein measurements. Members of this cluster generally showed increasing levels of phosphorylation while protein levels decreased (Fig. 17A).

In an effort to further characterize the proteins in each of the 7 clusters and demonstrate that membership in a cluster is nonrandom; we examined the functional classes of all of the proteins in each cluster using the MIPS catalogue. Interestingly, some of the 7 clusters show a significant enrichment in functional classes. When applying a Chi-Squared test and determining a two-tailed p value for each cluster, the distribution of functional classes proved to be highly nonrandom. This signifies that proteins following similar expression and phosphorylation trends also possess similar functional roles. The percentage of proteins in each functional category for each cluster can be seen in Fig. 17B. A remarkable trend can be seen for Cluster PE2 functional analysis. Cluster PE2 proteins show a slight increase in protein expression at the schizont stage, while the parasite prepares for egress from the erythrocyte and invasion into a new host cell. Supporting this is the
fact that 100% of proteins in the Host Invasion, Entry into Host-Cell, and Antigenic Variation functional categories are members of cluster PE2. Another interesting trend can be seen for cluster PEP3, which comprises a significant number of cell cycle regulatory proteins. Cluster PEP3 shows a dramatic increase in protein phosphorylation at the schizont stage. This is the period in the *Plasmodium* developmental cycle where multiple rounds of DNA replication are ongoing. The fact that approximately 29% of proteins in the Cell Cycle and DNA Processing functional category are members of Cluster PEP3 demonstrates the involvement of phosphorylation in the control of *P. falciparum* cell cycle progression.

### 2.4 Conclusions

This is the only quantitative study that compares dynamics of phosphorylation events as the parasite progresses through its asexual developmental stages in the host erythrocyte. A major challenge in the post-genome sequencing era is to understand the cellular roles of the vast number of proteins whose function is not deducible from sequence analysis. Achieving a comprehensive functional understanding of the *Plasmodium falciparum* proteome is particularly challenging, given that 60% of putative proteins have no known orthologs in other systems. Many *Plasmodium* proteins, including kinases, are atypical and possess hybrid features [136]; thus, this work brings a significant contribution to the ongoing global and multi-pronged efforts at implementing functional genomics studies of *Plasmodium falciparum*, the end goal of which is to facilitate the discovery of novel targets for intervention. Our phosphoproteomic analysis also represents a step toward elucidation of the kinase-substrate pairs that will eventually form the basis of a comprehensive *Plasmodium* signaling network. In view of the phylogenetic divergence of malaria parasites relative
to model eukaryotes such as yeast and mammalian cells, this is likely to shed light from a different angle on a domain that is of central importance in biology.

Data from this study are available at http://plasmodb.org/plasmo/
Figure 4: Identification of the Phosphoproteome and Proteome from the Intraerythrocytic Stages of Plasmodium falciparum

(A) Overview of the procedure for phosphopeptide preparation and enrichment from stage-specific P. falciparum lysates. Duplicate samples from each stage were processed to generate the lists of proteins and phosphorylation sites. Protein samples were digested with LysC, followed by labeling with TMT and fractionation on a SCX column. Enrichment of phosphopeptides was conducted using IMAC. The resulting phosphopeptide samples were analyzed on an LTQ-Velos Orbitrap mass spectrometer. Spectra were identified using SEQUEST, and the resulting data were filtered to a protein-level false discovery rate of 1%. AScore was used to assess phosphorylation site
localization. Phosphorylation sites with AScores above 13 were considered to be localized. (B) Tables representing the total numbers of stage-specific distributions of identified peptides and phosphopeptides, phosphorylation sites (before AScore filtering), and proteins and phosphoproteins in the six stage-specific samples indicated to the left. (C) Histogram depicting the numbers of peptides used for identification and quantification of the proteins reported in this study. (D) Correlation of the transcriptome (as stated on PlasmoDB) and the proteome of identified malaria proteins.
Figure 5: Distribution of Identified Human vs. Malaria Proteins and Phosphorylation Sites

(A) Comparison of the number of human and malaria proteins identified in the proteomic analysis.

(B) Percentage of human and malaria phosphorylation sites identified in the phosphoproteomic study.
### Table 1: Quantification of Stage-Specific (A) Proteins and (B) Phosphorylation Sites

#### Table 1A: Quantification of Stage-Specific Proteins

<table>
<thead>
<tr>
<th>Peak Expression</th>
<th>Accession Number</th>
<th>Annotation</th>
<th>Adj. pValue</th>
<th>Peptide Count</th>
<th>Stage</th>
<th>Schizont Average</th>
<th>Trophozoite Average</th>
<th>Ring Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizont/Ring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A5A7B0</td>
<td>Q8IDR3</td>
<td>Merozoite Surface Protein 1</td>
<td>1.81472E-69</td>
<td>426</td>
<td>0.19822</td>
<td>0.12845</td>
<td>0.17333</td>
<td></td>
</tr>
<tr>
<td>Q4AE85</td>
<td>A5.28286E-26</td>
<td>RhopH1 Clag3.1</td>
<td>1.98323E-26</td>
<td>173</td>
<td>0.20033</td>
<td>0.12982</td>
<td>0.16985</td>
<td></td>
</tr>
<tr>
<td>Q8I2J4</td>
<td>1.94174E-23</td>
<td>Profilin</td>
<td>31</td>
<td>0.20973</td>
<td>0.12672</td>
<td>0.16355</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q6LFH8</td>
<td>Q8I2J4</td>
<td>Ornithine Aminotransferase</td>
<td>4.9966E-181</td>
<td>123</td>
<td>0.26665</td>
<td>0.14896</td>
<td>0.08438</td>
<td></td>
</tr>
<tr>
<td>Q8IDQ9</td>
<td>Q8IJA9</td>
<td>Phosphoethanolamine N-Methyltransferase</td>
<td>2.9399E-151</td>
<td>148</td>
<td>0.25202</td>
<td>0.14008</td>
<td>0.10789</td>
<td></td>
</tr>
<tr>
<td>Q8IJA9</td>
<td>Adenosine Deaminase</td>
<td></td>
<td>2.7421E-138</td>
<td>108</td>
<td>0.25039</td>
<td>0.13255</td>
<td>0.11706</td>
<td></td>
</tr>
<tr>
<td>Q8IJA9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ring/Trophozoite</td>
<td>Q8IC05</td>
<td>Heat Shock Protein 86</td>
<td>5.2475E-178</td>
<td>403</td>
<td>0.21866</td>
<td>0.16784</td>
<td>0.11350</td>
<td></td>
</tr>
<tr>
<td>Q8LUJ7</td>
<td>Q8IJA9</td>
<td>Enolase</td>
<td>2.1139E-152</td>
<td>249</td>
<td>0.22732</td>
<td>0.16188</td>
<td>0.11079</td>
<td></td>
</tr>
<tr>
<td>Q71T02</td>
<td>Q8IJA9</td>
<td>L-Lactate Dehydrogenase</td>
<td>9.3727E-118</td>
<td>171</td>
<td>0.23005</td>
<td>0.15436</td>
<td>0.11559</td>
<td></td>
</tr>
<tr>
<td>Trophozoite</td>
<td>Q8I1L6</td>
<td>Heat Shock Protein 90</td>
<td>5.62519E-73</td>
<td>62</td>
<td>0.14727</td>
<td>0.26304</td>
<td>0.08969</td>
<td></td>
</tr>
<tr>
<td>Q8I492</td>
<td>Q8IJA9</td>
<td>Mature Parasite-Infected Erythrocyte Surface Antigen (MESA) or PIEMP2</td>
<td>2.44953E-66</td>
<td>132</td>
<td>0.13203</td>
<td>0.25505</td>
<td>0.11292</td>
<td></td>
</tr>
<tr>
<td>Q8U1J11</td>
<td>Q8IJA9</td>
<td>Pmc-2TM Maurer’s Cleft Two Transmembrane Protein</td>
<td>2.86193E-51</td>
<td>84</td>
<td>0.13169</td>
<td>0.22994</td>
<td>0.13838</td>
<td></td>
</tr>
<tr>
<td>Trophozoite/Schizont</td>
<td>Q8I2G1</td>
<td>Ring-Exported Protein 1</td>
<td>1.08717E-44</td>
<td>89</td>
<td>0.07421</td>
<td>0.21638</td>
<td>0.20940</td>
<td></td>
</tr>
<tr>
<td>Q8ILA1</td>
<td>Q8I2G1</td>
<td>Conserved Plasmodium Protein</td>
<td>2.85479E-44</td>
<td>190</td>
<td>0.12001</td>
<td>0.18267</td>
<td>0.19732</td>
<td></td>
</tr>
<tr>
<td>P13830</td>
<td>Q8ILA1</td>
<td>RecName: Full- Ring-Infected Erythrocyte Surface Antigen</td>
<td>4.38793E-30</td>
<td>73</td>
<td>0.08755</td>
<td>0.18778</td>
<td>0.22466</td>
<td></td>
</tr>
<tr>
<td>Schizont</td>
<td>Q8S5M3</td>
<td>Conserved Plasmodium Protein</td>
<td>2.59184E-42</td>
<td>126</td>
<td>0.13449</td>
<td>0.15767</td>
<td>0.20784</td>
<td></td>
</tr>
<tr>
<td>Q8UO4</td>
<td>Q8S5M3</td>
<td>Conserved Plasmodium Protein</td>
<td>5.28286E-26</td>
<td>49</td>
<td>0.11808</td>
<td>0.10549</td>
<td>0.27644</td>
<td></td>
</tr>
<tr>
<td>Q8IDR3</td>
<td>Q8S5M3</td>
<td>Myosin A</td>
<td>8.48E-26</td>
<td>54</td>
<td>0.15048</td>
<td>0.08967</td>
<td>0.25985</td>
<td></td>
</tr>
</tbody>
</table>
Table 1B: Quantification of Stage-Specific Phosphorylation Sites

<table>
<thead>
<tr>
<th>Peak Phosphorylation Number</th>
<th>Accession Number</th>
<th>Annotation</th>
<th>Adj. pValue</th>
<th>Peptide Count</th>
<th>Class</th>
<th>Stage</th>
<th>Ring Average</th>
<th>Trophozoite Average</th>
<th>Schizont Average</th>
<th>Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizont/ Ring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O97239</td>
<td></td>
<td>Conserved Plasmodium Protein</td>
<td>5.67584E-05</td>
<td>8</td>
<td>A</td>
<td>0.23169</td>
<td>0.14715</td>
<td>0.16912</td>
<td>S406</td>
<td></td>
</tr>
<tr>
<td>Q8ILC8</td>
<td></td>
<td>DNA Topoisomerase II</td>
<td>0.000263595</td>
<td>3</td>
<td>O</td>
<td>0.24746</td>
<td>0.11807</td>
<td>0.18188</td>
<td>S1296</td>
<td></td>
</tr>
<tr>
<td>C6KT13</td>
<td></td>
<td>Conserved Plasmodium Protein</td>
<td>0.006320528</td>
<td>2</td>
<td>O</td>
<td>0.22058</td>
<td>0.13932</td>
<td>0.14850</td>
<td>S25</td>
<td></td>
</tr>
<tr>
<td>Ring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q8I296</td>
<td></td>
<td>Ubiquitin Carboxyl-Terminal Hydrolase</td>
<td>8.91046E-11</td>
<td>3</td>
<td>A</td>
<td>0.36738</td>
<td>0.09242</td>
<td>0.07079</td>
<td>S2603</td>
<td></td>
</tr>
<tr>
<td>C5HEV4</td>
<td></td>
<td>Ringgly-Associated Protein 1</td>
<td>1.436E-09</td>
<td>8</td>
<td>B</td>
<td>0.35644</td>
<td>0.08576</td>
<td>0.06399</td>
<td>S99</td>
<td></td>
</tr>
<tr>
<td>O7K6A5</td>
<td></td>
<td>Multi-Drug Resistance Protein</td>
<td>4.05683E-09</td>
<td>19</td>
<td>A</td>
<td>0.26736</td>
<td>0.14217</td>
<td>0.11477</td>
<td>S513</td>
<td></td>
</tr>
<tr>
<td>Ring/ Trophozoite</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q8IDG2</td>
<td></td>
<td>Kech Protein</td>
<td>2.50453E-09</td>
<td>14</td>
<td>B</td>
<td>0.25089</td>
<td>0.17969</td>
<td>0.13337</td>
<td>T192</td>
<td></td>
</tr>
<tr>
<td>P39898</td>
<td></td>
<td>RecName: Full-Plasmeprin 1</td>
<td>1.201E-08</td>
<td>4</td>
<td>O</td>
<td>0.30932</td>
<td>0.25867</td>
<td>0.05553</td>
<td>T23</td>
<td></td>
</tr>
<tr>
<td>O97238</td>
<td></td>
<td>Conserved Plasmodium Protein</td>
<td>1.68106E-06</td>
<td>9</td>
<td>B</td>
<td>0.24925</td>
<td>0.17173</td>
<td>0.14156</td>
<td>S275</td>
<td></td>
</tr>
<tr>
<td>Trophozoite</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q8IDG8</td>
<td></td>
<td>Membrane-Associated Histidine Rich Protein 2</td>
<td>8.66673E-11</td>
<td>7</td>
<td>O</td>
<td>0.12973</td>
<td>0.41704</td>
<td>0.13912</td>
<td>T136</td>
<td></td>
</tr>
<tr>
<td>Q8IDM6</td>
<td></td>
<td>Nucleoside Transporter 1</td>
<td>3.16966E-07</td>
<td>6</td>
<td>A</td>
<td>0.16576</td>
<td>0.32821</td>
<td>0.11905</td>
<td>S16</td>
<td></td>
</tr>
<tr>
<td>Q8I492</td>
<td></td>
<td>Mature Parasite-Infected Erythrocyte Surface Antigen (MESA) or PIEMP2</td>
<td>1.52967E-05</td>
<td>7</td>
<td>A</td>
<td>0.12416</td>
<td>0.21326</td>
<td>0.17549</td>
<td>S177</td>
<td></td>
</tr>
<tr>
<td>Trophozoite/ Schizont</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q6JW6</td>
<td></td>
<td>Transcription Factor with AP2 Domain</td>
<td>2.3356E-35</td>
<td>25</td>
<td>A</td>
<td>0.06787</td>
<td>0.23552</td>
<td>0.27157</td>
<td>S913</td>
<td></td>
</tr>
<tr>
<td>Q8I3A3</td>
<td></td>
<td>Ubiquitin Specific Protease</td>
<td>9.15377E-22</td>
<td>16</td>
<td>A</td>
<td>0.08560</td>
<td>0.21405</td>
<td>0.29959</td>
<td>S1457</td>
<td></td>
</tr>
<tr>
<td>Q8IB78</td>
<td></td>
<td>Nucleoside Transporter</td>
<td>1.64292E-20</td>
<td>10</td>
<td>O</td>
<td>0.02631</td>
<td>0.35909</td>
<td>0.28085</td>
<td>T336</td>
<td></td>
</tr>
<tr>
<td>Schizont</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q8ILA9</td>
<td></td>
<td>Conserved Plasmodium Protein</td>
<td>1.86363E-22</td>
<td>18</td>
<td>P</td>
<td>0.07735</td>
<td>0.16488</td>
<td>0.29758</td>
<td>S121</td>
<td></td>
</tr>
<tr>
<td>Q8IC17</td>
<td></td>
<td>Recognition Complex Subunit 2</td>
<td>7.10986E-19</td>
<td>11</td>
<td>T</td>
<td>0.13324</td>
<td>0.22689</td>
<td>0.37323</td>
<td>S347</td>
<td></td>
</tr>
<tr>
<td>Q8IHR4</td>
<td></td>
<td>Dynamin-Like Protein</td>
<td>7.40701E-19</td>
<td>16</td>
<td>O</td>
<td>0.12588</td>
<td>0.14555</td>
<td>0.24440</td>
<td>S782</td>
<td></td>
</tr>
</tbody>
</table>
Table 1: Quantification of Stage-Specific (A) Proteins and (B) Phosphorylation Sites

(A) Stage-specific abundant proteins exhibiting a >1.5-fold difference in expression among intraerythrocytic stages. The representative proteins for each stage displayed the most significant $p$ values. The columns corresponding to the ring average, trophozoite average, and schizont averages are the normalized intensities for peptides associated with each particular stage. (B) Stage-specific abundant phosphorylation sites exhibiting a >1.5 fold difference in phosphorylation levels among intraerythrocytic stages. The proteins for each stage displayed the most significant $p$ values. Site classes were assigned using a decision tree algorithm: A: acidic; B: basic; P: proline-directed; T: tyrosine; O: other. The columns listing the ring average, trophozoite average, and schizont averages are the normalized intensities for phosphorylated peptides associated with each stage.
Figure 6: Comparative Analysis of Stage-Specific Protein Expression

(A) Venn diagram depicting the distribution of identified malaria proteins over the intraerythrocytic stages. (B) Percentage of globally expressed proteins and proteins quantified at each of the intraerythrocytic stages. (C) Expression level changes as determined by a >1.5-fold difference. (D) Number of proteins up- and down-regulated in each stage in comparison with ring stage expression levels.
Figure 7: Stage-Specific Analysis of Functional Profiles

Functional profiles of expressed proteins using GO annotations downloaded from PlasmoDB (www.plasmodb.org) or UniProt (www.uniprot.org) as defined by the MIPS catalogue. To avoid redundancy, we assigned only one class per protein.
Figure 8: Functional Profile of Hypothetical Proteins

The functional profile of hypothetical proteins was generated using GO annotations downloaded from PlasmoDB (www.plasmodb.org) or UniProt (www.uniprot.org) as defined by the MIPS catalogue. To avoid redundancy, only one functional class was assigned per protein.
Figure 9: Gene Ontology Term Analysis of Intraerythrocytic Stage

(A) Enrichment for GO “Molecular Function” terms of proteins detected in *P. falciparum* intraerythrocytic stages. The figure shows the percentage of GO terms identified in our proteomic data compared to terms of all predicted *P. falciparum* proteins. (B) Enrichment for GO “Cellular Component” terms of proteins detected in the intraerythrocytic stages. The number of each “Cellular Component” identified in our proteomic data is compared to all predicted *P. falciparum* proteins. The percentage identified in our data for each component is shown.
Figure 10: Global Analysis of Stage-Specific Phosphorylation Patterns

(A) Relative distributions of stage-specific phosphopeptides, phosphorylated sites, and phosphoproteins. Raw values are provided in Figure 5B. (B) Histogram depicting the numbers of
phosphorylation sites identified for each phosphorylated protein. (C) Distribution of the number of localized and non-localized phosphorylation sites throughout the intraerythrocytic stages. (D) Relative abundance of serine, threonine, and tyrosine residues within all phosphoproteins compared with their frequency of phosphorylation. (E) Extent of phosphorylation of serine, threonine, and tyrosine residues within all phosphoproteins. (G) Venn diagram depicting the distribution of identified malaria phosphorylation sites over the intraerythrocytic stages. (H) Phosphorylation level changes as determined by a >1.5-fold difference in comparison with observed ring stage levels.
Figure 11: Stage-Specific Distribution of Phosphorylation Site Classes and Motifs

Phosphorylation sites were categorized as acidic, basic, proline-directed, tyrosine, or other as previously described [95]. (A) Overall percentage of each site class and percentage of each throughout the intraerythrocytic cell cycle of *P. falciparum*. (B) Proportion of phosphoproteins
containing phosphorylation sites from only one site class, from two, three, or four different site classes, or from all five different site classes. (C) Proportion of residues phosphorylated in each site class through the intraerythrocytic stages. (D) Comparison of phosphorylation across the *P. falciparum* intraerythrocytic stages for motifs containing serine, threonine, arginine, asparagine, glycine, and proline residues. (E) Table identifying potential Ser/Thr kinases responsible for observed phosphorylation on representative putative substrates.
Figure 12: Phosphorylation Profiles of Putative MAPK/CDK Substrates and Tyrosine Phosphorylated Proteins

(A) Immunoblot analysis using a MAPK/CDK phospho-specific antibody on 8 hour time point lysates. (B) Analysis of 45 putative MAPK/CDK substrates revealing the number of proteins with peak expression in each stage and the number of phosphorylation events peaking at each stage. (C) Immunoblot analysis using a phospho-tyrosine antibody on 8 hour time point lysates. (D) Stage-specific distribution of proteins containing tyrosine phosphorylation events.
Figure 13: Subdomain Location of Phosphorylation Sites in Protein Kinases
(A) Subdomain location of 126 phosphorylation sites observed on 30 eukaryotic protein kinases. Bar height indicates the number of observed phosphorylation sites relative to the kinase domain, whether in the domain, the N-terminal tail, or the C-terminal tail. Those in the kinase domain were further assigned to a subdomain location. (B) Stage-specific comparison of eukaryotic protein kinase expression levels and phosphorylation levels. (C) Histogram depicting the stage-specific distribution of the number of phosphorylation sites at a given time during the intraerythrocytic cell cycle for each of the kinase subdomains as well as the N-terminal tail and C-terminal tail.
Those localized phosphorylation events that mapped to known kinase subdomains were analyzed using NetPhorest to identify the kinase or kinase groups most likely to phosphorylate each site.
based on neighboring sequences [104]. A total of 22 kinases and kinase groups accounted for the most confident 2.5% of all predicted kinase-site relationships and are mapped below. The red heat map displays the predicted confidence that each kinase or kinase family would phosphorylate the sequence in question; the yellow heat map displays relative levels of each phosphorylation site throughout three stages of the *Plasmodium* life cycle.
Figure 15: Comparison of Protein and Phosphoprotein Expression Across *P. falciparum* Intraerythrocytic Stages

(A) Comparison of proteins identified with and without phosphopeptide enrichment. (B) Numbers of proteins and phosphoproteins with peak expression at each of the intraerythrocytic stages. (C) Stage-specific comparison of protein expression and phosphorylation levels. (D) Bar chart reflecting proportions of proteins with global, moderate, and stage-specific expression in an unmodified form, modified form, and those containing stage-specific phosphorylation sites.
Figure 16: K-Medoids Cluster Analysis of Differential Proteins

(A) Proteins and phosphorylation sites that showed statistically significant differences across the three growth stages in protein abundance, phosphorylation levels, or both were classified via K-Medoids clustering. The numbers of proteins and phosphorylation sites that were assigned to each cluster as well as the trends observed for protein expression and phosphorylation are depicted. (B) Functional profiles of proteins in each cluster using GO annotations downloaded from PlasmoDB (www.plasmodb.org) or UniProt (www.uniprot.org) as defined by the MIPS catalogue.
CHAPTER 3: PHOSPHOPROTEOMIC ANALYSIS OF *PLASMODIUM FALCIPARUM* REVEALS SUBSTRATES OF THE ATYPICAL KINASE PfPK7

3.1 Summary

Options for malaria therapy is increasingly becoming limited because of widespread drug resistance. Therefore, there is a need to identify new pharmacophores targeted against novel targets. Atypical *Plasmodium* protein kinases are considered to be ideal targets due to the distinct structural and functional characteristics in comparison to their host. *Plasmodium falciparum* protein kinase 7 (PfPK7) is an “orphan” protein kinase displaying regions of homology to more than one protein kinase family. Furthermore, PfPK7 has been shown to be involved in regulating parasite proliferation and development. To better understand the involvement of PfPK7 in *Plasmodium falciparum*, we performed isobaric tag-based quantitative phosphoproteomics analysis of the schizont and segmenter asexual stages in both the presence and absence of PfPK7. A total of 1,047 proteins and 3,875 phosphorylation sites were identified in the wild type and PfPK7− samples. In order to determine which proteins PfPK7 may be phosphorylating, we examined the phosphoproteomic data for proteins displaying decreased phosphorylation in the absence of PfPK7 during both the schizont and segmenter stages. We identified 146 proteins with a total of 239 phosphorylation sites exhibiting decreased phosphorylation in the absence of PfPK7 during both intraerythrocytic stages. Further analysis of the decreased phosphorylation events revealed three distinct motifs that are enriched among phosphorylated sites in *Plasmodium* proteins that are down regulated. Due to this decreased phosphorylation, it is possible that these proteins may act as substrates of PfPK7 and function in parasite proliferation and/or development. The ability of recombinant PfPK7 to utilize these putative substrates was examined and we have demonstrated
that PfPK7 is active towards various substrates that were predicted from the phosphoproteomics analysis. More specifically, the atypical kinase PfPK7 was able to directly phosphorylate two conserved *Plasmodium* proteins with unknown functions (PF10_0257 and PFB0490c) and 60s ribosomal protein L7Ae/L30e putative (PFD0960c). We also determined that PfPK7 was not able to directly phosphorylate many of the phosphorylation events that displayed decreased phosphorylation in the absence of PfPK7, suggesting the role of PfPK7 as an upstream regulatory kinase. Our phosphoproteomics analysis has revealed insight into the dynamics of PfPK7 phosphorylation during both the schizont and segmenter stages, while making a major step toward defining PfPK7-substrate pairs.

### 3.2 Materials and Methods

#### 3.2.1 *Plasmodium falciparum* Culture

The PfPK7- cell line was kindly provided by C. Doerig (Monash University, Clayton Australia). Parasites of *Plasmodium falciparum* strain 3D7 and PfPK7- were grown at a 4-10% parasitemia and 4% hematocrit in RPMI 1640 culture medium supplemented with A+ erythrocytes and 0.5% albumax as previously described [88]. Blasticidin was added to a final concentration of 2.5μg/ml in the PfPK7- cultures. Parasites were doubly synchronized as described by Pease et al. [1]. Briefly, schizont stage parasites were magnetically synchronized using a MACS LD (Miltenyi Biotec, Auburn, CA) column with a Midi-MACs Separator. Approximately 8 hours after MACS synchronization, the parasites had re-infected fresh red blood cells and entered the ring stage. Parasites were then re-synchronized by treatment with 5% sorbitol (w/v).

Parasite growth and parasitemia were periodically monitored by Giemsa-stained blood smears. Tightly synchronized schizonts (approximately 36 hours ± 4 hours post-invasion) and
segmenters (46 hours were ± 4 hours post-invasion) were harvested following established protocols [91]. Parasites were isolated by lysing the red blood cells in 0.1% saponin followed by thorough washing in PBS. The subsequent parasite pellets were weighed and proteins were immediately extracted by lysis in an 8M urea lysis buffer supplemented with protease and phosphatase inhibitors (8M urea, 75mM Tris, pH 8.2, 1X HALT protease inhibitor, 1X HALT phosphatase inhibitor) [16]. Lysates were cleared by centrifugation for 10 minutes at 4°C at 14,000 rpm. Protein concentration was determined for each sample by BioRad Bradford Assay. Duplicate samples were prepared for each developmental stage and pooled.

3.2.2 PfPK7 Verification through Southern Blot Analysis and PCR Analysis

Genomic DNA was isolated from wild type 3D7 and from PfPK7− parasites. A washed pellet of parasites obtained by saponin lysis was treated with 250mM EDTA, 2% sodium dodecyl sulfate, and 150μg/ul RNAse at 37°C for 1 hour. Next, the pellet was treated with 150μg/ml proteinase K at 60°C for 2 hours. The DNA was extracted with chloroform: phenol: isoamyl alcohol (25:24:1) and chloroform: isoamyl alcohol (24:1) followed by precipitation with ethanol and 0.3M sodium acetate. One microgram of each wild type 3D7 and PfPK7− DNA was digested with EcoRV and NcoI. The digested DNA was then fractionated on a 0.8% agarose gel and transferred to a Schleicher and Schuell Nytran SPC membrane. Full length PfPK7 was used as the probe. The probe was generated using the following primers: primer 1: ATGAAGGATATTTTATCTAATTATTC and primer 2: TTATAATTTTTTCTCTTTTATAAAG. The full length PfPK7 gene was labeled with (32P)dCTP using the Prime-it II random primer labeling kit following the manufacturer’s instructions (Stratagene). After the probe was labeled, it was purified using the QIAquick Nucleotide Removal kit as instructed (QIAgen). Hybridization was carried out at 60°C overnight as
described by Sambrook and Russell [137]. After thorough washing, autoradiograms of hybridized blots were prepared using Kodak Biomax MS film exposed at -80°C with the use of an intensifying screen.

To further verify the presence of PfPK7 in the wild type 3D7 cell line and the absence of PfPK7 in the knock out cell line, genotype characterization through PCR analysis was performed. Genomic DNA extracted from the wild type 3D7 cell line and PfPK7⁻ cell line was used for detection of a partial PfPK7 gene following the genotype characterization used when creating the PfPK7⁻ cell line [27]. The sequences for the primers had minor modifications, compared to the original protocol, to remove the cut sites and amplify only the PfPK7 partial gene. The primers were designed as follows: primer 1: ATGAAGGATATTTTATCTAATTATTCAAAC and primer 2: ACCCAAA ACTCCATATCCACCTTTGC. The resulting PCR product would be 702 base pairs in size.

3.2.3 Phosphopeptide Enrichment and Identification

Following protein extraction, samples containing 1mg of protein per stage for the wild type 3D7 *P. falciparum* and PfPK7⁻ samples were reduced, alkylated, and digested with Lys-C prior to labeling with TMT isobaric labeling reagents [87]. The resulting peptides were fractionated via strong cation exchange chromatography (SCX). The resulting 20 fractions were subjected to immobilized metal affinity chromatography (IMAC) to specifically pull out phosphorylated peptides [92]. Phosphopeptides were analyzed via LC-MS/MS on an LTQ-Velos-Orbitrap mass spectrometer as described previously [1]. Peptides were identified by matching individual MS/MS spectra with peptide sequences using Sequest [94]. Searches employed a database containing *Plasmodium* sequences from NCBI [96] as well as human protein sequences from the International
Protein Index [97]. The peptides were filtered using a multivariate approach to remove questionable identifications based on a target-decoy strategy [95, 98]. Peptides were assembled into proteins and proteins were further filtered to ensure a protein-level false discovery rate of 1%. Individual phosphorylation sites on each peptide were scored to assess how confidently the modification could be localized to a single residue using AScore [99]. Finally, the peptides and proteins were filtered to account for protein redundancy.

For quantification, TMT reporter ion signals were extracted from MS/MS spectra and normalized assuming equal protein loading in each channel. Each phosphorylation site was quantified by gathering together all matching peptides, filtering out peptides that didn’t meet our quantification standards, and re-scaling reporter ion intensities so that they summed to 1.0.

3.2.4 Identification of PfPK7 Substrate Phosphorylation Motifs

To identify primary sequence motifs directly or indirectly associated with PfPK7 activity; those phosphorylation sites whose abundance decreased in the absence of PfPK7 were extracted. Sites were separated according to the residue phosphorylated (Ser, Thr, or Tyr) and enriched motifs were identified using the algorithm Motif-X [125] to compare down-regulated sites with all Ser, Thr, and Tyr residues in the Plasmodium proteome. Motifs were required to occur at least 10 times among down-regulated sites and the minimum significance was set to $1 \times 10^{-5}$. While no significant motifs were observed for Thr and Tyr sites due to their infrequent detection, three motifs were detected for Ser phosphorylation sites.

3.2.5 Assignment as a Potential Substrate and Generation of Expression Constructs

When analyzing the wild type 3D7 and PfPK7− phosphoproteomic data sets, we compared phosphorylation levels by calculating two ratios among the two cell lines and developmental stages
(PfPK7\(^{-}\) schizont: 3D7 schizont and PfPK7\(^{-}\) segmenter: 3D7 segmenter). Phosphorylation events exhibiting greater than 1.5-fold differences in abundance between the two cell lines were considered enriched in that particular cell line, while phosphorylation events showing less than 1.5-fold changes for both comparisons were considered phosphorylated equally in both cell lines. Of the 3,875 phosphorylation sites identified in the phosphoproteomic analysis, 239 phosphorylation events on 146 phosphoproteins displayed a greater than 1.5-fold decrease in phosphorylation in the PfPK7\(^{-}\) cell line at both the schizont and segmenter stages. In order to investigate if any of these 146 phosphoproteins are in fact substrates of PfPK7, we generated His-tagged constructs of the potential substrates with selected open reading frames exhibiting the largest decrease in phosphorylation in the PfPK7\(^{-}\) cell line for both the schizont and segmenter stages and three potential substrates containing the identified RxxS* motif whose lengths were under 2,000 base pairs in size. The following proteins were cloned into the pET30 EK/LIC vector following the manufacturer’s instructions (EMD Millipore): PFB0100c, PF10\_0257, PF14\_0068, PF10\_0068, PFD0960c, PFI1555w, PF13\_0102, PFB0490c, and PF14\_0190.

### 3.2.6 Bacterial Expression and Purification of Recombinant Proteins

The potential PfPK7 substrates were optimally expressed in BL21-CodonPlus (RIPL) cells (Stratagene) and purified from the soluble fraction by affinity chromatography using HisLink resin and following the manufacturer’s instructions (Promega). The plasmid encoding GST-PfPK7 were kindly provided by C. Doerig (Monash University, Clayton Australia). The expression and purification of GST-PfPK7 was performed following published procedures [138].
3.2.7 *In vitro Kinase Assays*

Kinase reactions to measure PfPK7 auto-phosphorylation and activity toward the potential substrates were performed using 1μg of recombinant PfPK7 and 5μg of recombinant substrate in a total volume of 30μl as described previously [45, 139]. A standard kinase buffer containing 20mM Tris/HCl (pH 7.5), 20mM MgCl₂, 2mM MnCl₂, 10μM ATP, and 5μCi (γ-³²P) ATP was used. Kinase reactions were incubated for 30 minutes at 30°C and stopped by the addition of 5X gel loading buffer. The entire kinase reaction was loaded on a 12% SDS/polyacrylamide gel. The gels were then stained with Coomassie blue stain, de-stained using a 30% methanol and 10% glacial acetic acid solution, dried, and exposed for autoradiography. Signal intensity values were obtained and normalized by using the kinase buffer control as the 0% value and PfPK7 auto-phosphorylation as the 100% value. The ability of PfPK7 to phosphorylate potential substrates was determined by the normalized percent signal intensity compared to the 100% value of PfPK7 auto-phosphorylation.

3.3 Results and Discussion

3.3.1 Verification of the Disruption of the PfPK7 Gene

The PfPK7 cell line was kindly provided by C. Doerig (Monash University, Clayton Australia). This cell line was generated through the use of reverse genetics to disrupt the PfPK7 gene [27]. Briefly, PfPK7 was knocked out by utilizing a single crossover homologous recombination to insert a corresponding central region of the PfPK7 catalytic domain next to a domain conferring blasticidin resistance in the pCAM-BSD vector and transferred into 3D7 parasites through electroporation [45, 140]. The homologous recombination at the PfPK7 locus
resulted in a pseudo-diploid configuration with two truncated copies of PfPK7 lacking essential regions of the catalytic domain and disrupting the ability of PfPK7 to function properly [27].

To verify that the PfPK7- cell line had lost the wild type gene, we used both PCR and Southern blot analysis. Either 50ng or 25ng of genomic DNA was used to detect a partial PfPK7 product (702 base pairs) similar to the genotype characterization carried out in the creation of the PfPK7- cell line [27]. PCR analysis showed that the 3D7 wild type cell line contained the expected 702 base pair product (lanes 1 & 3) and that the PfPK7- cell line had lost the wild type PfPK7 gene (lanes 2 & 4) (Fig. 18A).

As can be seen from Fig. 18C, a hybridization signal corresponding to pfpk7 was obtained only in the genomic DNA isolated from the 3D7 wild type cell line when using the PfPK7 probe. Fig. 18B demonstrates equal loading of digested genomic DNA for both wild-type 3D7 and PfPK7-.

Taken together, the PCR and Southern blot analysis confirms the lack of full-length PfPK7 in the PfPK7- cell line; however, PfPK7 is still present and functional in the wild-type 3D7 cell line.

3.3.2 Identification of 1,047 Phosphoproteins and 3,875 Phosphorylation Sites from the Schizont and Segmenter Intraerythrocytic Stages of Plasmodium falciparum

We used isobaric labeling for quantitative analysis of LysC-digested parasite extracts prepared from tightly synchronized schizont (approximately 36 hours ± 4 hours post-invasion) and segmenter (approximately 46 hours ± 4 hours post-invasion) stages. Fig. 19 depicts representative images of the wild type 3D7 cultures (Fig. 19A) and PfPK7- cultures (Fig. 19B) harvested at each of the cell cycle stages. A workflow depicting the phosphoproteomic analysis is outlined in Fig. 20. This quantitative phosphoproteomic analysis identified a total of 1,047 proteins and 6,795 phosphorylation sites across both cell lines and both intraerythrocytic cell cycle stages. Of these
phosphorylation events, only 3,875 met our strict quantification standards at a protein-level FDR of <1%.

In addition to identification, we also provided quantitative analysis of phosphorylation changes comparing the two cell lines at the schizont and segmenter stages. To determine which phosphorylation events were occurring predominately in the 3D7 wild type cell line, the PfPK7 cell line, or equally in both cell lines, we examined each phosphorylation site to define those with a >1.5-fold increase or decrease between the cell lines. Phosphorylation events exhibiting greater than a 1.5-fold difference in abundance between the two were considered to be enriched in that particular cell line, while phosphorylation events showing less than a 1.5-fold change between the two cell lines were considered to be present equally in both the wild type 3D7 and PfPK7. Fig. 21A demonstrates the distribution of phosphorylation events for each of the cell lines. At the schizont stage, a total of 1,490 phosphorylation events were classified as occurring in the 3D7 wild type cell line, which means that these phosphorylation events exhibited a greater than 1.5-fold decrease in phosphorylation when PfPK7 was no longer functional. Furthermore, 609 phosphorylation events were classified as 3D7 at the segmenter stage. Interestingly, 794 phosphorylation events at the schizont stage and 1,027 phosphorylation events at the segmenter stage were up regulated in the absence of PfPK7 (Fig. 21A). The increase in phosphorylation seen on some proteins that occurs in the absence of PfPK7 could possibly be a compensatory mechanism the parasite is utilizing to ensure survival in the absence of PfPK7. Further analysis across the cell lines revealed that approximately 41% of the schizont stage phosphorylation events and 58% of the segmenter phosphorylation events were detected equally in both cell lines, while 20.5% of the phosphorylation events at the schizont stage and 26.5% of the phosphorylation events at the segmenter stage have decreased phosphorylation in the absence of a functional PfPK7 (Fig. 21B).
3.3.3 Analysis of Decreased Phosphorylation Events and the Identification of Unique Phosphorylation Motifs

To analyze phosphorylation events that may be occurring due to PfPK7 activity, we examined the phosphoproteomic data for proteins with a greater than 1.5-fold decrease in phosphorylation in both the schizont and segmenter stages in the absence of PfPK7. A total of 239 phosphorylation events on 146 phosphoproteins exhibited decreased phosphorylation at both intraerythrocytic stages (Fig. 22A). The five phosphoproteins with the biggest decrease in phosphorylation at the schizont and segmenter stages are shown in Table 2A and Table 2B, respectively.

We sorted proteins into functional classes using the Munich Information Centre for Protein Sequences (MIPS) catalog with some adaptations for classes specific to the Plasmodium parasite, such as proteins pertaining to malaria pathogenesis [1, 110]. Functional classification was carried out on the entire phosphoproteomic data set, which was then compared to the functional classification of the 146 phosphoproteins with decreased phosphorylation in the absence of PfPK7 at both the schizont and segmenter stages. The distribution of the functional profiles is summarized in Fig. 22B. The comparison of functional profiles did not result in any marked differences (Fig. 22B). For example, 3% of the phosphoproteins from the overall data set and 3% of the phosphoproteins with decreased phosphorylation at both intraerythrocytic stages in the absence of PfPK7 are involved in the cell cycle and DNA processing category. This demonstrates that in the absence of PfPK7, no single functional category is dramatically affected. More specifically, this indicates that PfPK7 does not specifically phosphorylate proteins within a given functional category; if so, we would have detected a significant decrease in the percentage of proteins in that particular functional category when comparing the functional profiles.
To further investigate the involvement of PfPK7 in phosphorylating proteins at the schizont and segmenter stages, we compared the phosphorylated peptides in the overall data set to those with decreased phosphorylation in the absence of PfPK7 at both of these stages. MotifX analysis was performed to identify primary sequence motifs directly or indirectly associated with PfPK7 activity. Three motifs were identified that are enriched among pS sites in *Plasmodium falciparum* that were down regulated (Fig. 22C). When compared to the entire data set, the RxxS* motif had the greatest fold increase (4.12-fold) in the proteins with decreased phosphorylation when PfPK7 was no longer functional. The RxxS* motif is a well-known phosphorylation motif matching the AGC-family of kinases, which includes members such as cAMP-dependent protein kinase (PKA), cGMP-dependent protein kinase (PKG), protein kinase C, AKT, and RSK. This class of kinases recognizes an Arg at position -3 relative to the phosphorylated residue [141, 142]. Akt has been shown to play a central role in the mediation of cellular processes such as cell growth and survival and transcriptional regulation [143]. The fact that PfPK7 functions similarly to AKT by regulating cell proliferation and also recognizes the RxxS* motif supports the idea that PfPK7 may have a role in regulating cell growth and proliferation of *Plasmodium*.

The MotifX analysis also identified a greater than 3-fold enrichment of the S*xD and Rxxxx*S* motifs among the proteins with decreased phosphorylation at both the schizont and segmenter stages in the absence of PfPK7 (Fig. 22C). The S*xD motif is traditionally utilized by casein kinase 1 (CK1) and casein kinase 2 (CK2), while the Rxxxx*S* motif is commonly recognized by protein kinase C (PKC) or CDC2-like kinase (CLK2). The CK1 and CK2 protein kinase families are serine/threonine protein kinases, which have been evolutionarily conserved through many species [144]. Members of the CK1 family function as regulators of signal transduction pathways in many eukaryotic cells. These kinases are responsible for phosphorylating...
substrates involved as key regulatory proteins in the control of cell differentiation, proliferation, and chromosome segregation [144]. Members of the CK2 family of protein kinases have functions similar to CK1 members and have been shown to have involvement in cell cycle control, DNA repair, and regulation of circadian rhythm of cellular processes [145]. The CLKs are an evolutionarily conserved group of dual specificity kinases belonging to the CMGC group of kinases. The CLK family of protein kinases interact and/or phosphorylate serine/arginine-rich (SR) proteins of the spliceosomal complex. CLK2 is involved in the regulation of several cellular processes including cell cycle progression [146, 147]. PfPK7’s ability to phosphorylate proteins with motifs traditionally recognized by AKT, CK1, CK2, and CLK2 supports the role of PfPK7 in cell cycle control and cell proliferation.

3.3.4 Identification and Characterization of Putative PfPK7 Substrates

In an effort to identify putative substrates of PfPK7, we examined all of the phosphorylation sites displaying decreased phosphorylation at both the schizont and segmenter stages in the absence of PfPK7. A total of 146 phosphoproteins fit this criteria and these proteins may represent putative substrates of PfPK7. The phosphoproteins with the largest decrease in phosphorylation at both intraerythrocytic stages are more likely to be true substrates of PfPK7 than those with only small decreases in phosphorylation. After examining the phosphoproteins with decreased phosphorylation in the absence of PfPK7, we generated His-tagged recombinant forms of the six proteins with the largest decrease in phosphorylation at both asexual stages. Table 3A displays the cloned proteins as well as the phosphorylated peptides that exhibited a significant decrease in phosphorylation, the biological process the protein is involved in, and it’s molecular function. Next, we identified the three phosphorylation events occurring at an RxxS* motif with the largest decrease in
phosphorylation and a recombinant His-tagged form of these proteins were generated. Table 3B shows the recombinant proteins as well as the phosphorylated peptide, the biological process the protein is involved in, and the protein’s molecular function. The phosphorylated peptides were then subjected to motif analysis in order to predict the protein kinase that may be responsible for the phosphorylation event. Significantly, the majority of phosphorylated peptides with large decreases in phosphorylation in the absence of PfPK7 are predicted to be recognized by CK1, CK2, and AKT (Table 3A and 3B). This prediction and the fact that motifs recognized by CK1, CK2, and AKT kinases are enriched among phosphopeptides with decreased phosphorylation when PfPK7 is absent, suggests the functional similarities of PfPK7 to CK1, CK2, and AKT protein kinases. Alternatively, PfPK7 may act upstream of PfCK1, PfCK2, and PfAKT, thus regulating their downstream activities.

In order to better define the role of PfPK7 in *Plasmodium falciparum* we performed *in vitro* kinase assays to test that identified substrates are utilized by the kinase. Fig. 23A and 23C show 1 μg of recombinant PfPK7 and 5 μg of recombinant substrate were used for each assay condition. After five independent kinase assays, it was clear that PfPK7 was able to robustly phosphorylate a conserved *Plasmodium* protein (PF10_0257) compared to the generic substrate myelin basic protein (MBP) (Fig. 23B). PfPK7 was also able to directly phosphorylate 60s ribosomal protein L7Ae/L30e, putative (PFD0960c) and a conserved *Plasmodium* protein (PFB0490c) to a lesser extent (Fig. 23D). The ability of PfPK7 to phosphorylate these substrates was determined by normalizing signal intensity values to the master mix alone (0% value) and PfPK7 auto-phosphorylation (100% value). PfPK7 phosphorylation of PF10_0257 was 59%, which is greater than PfPK7’s phosphorylation of the control substrate MBP at 38% (Fig. 23E). PfPK7’s phosphorylation of PFD0960c and PFB0490c was 7% and 7%, respectively (Fig. 23E). Although
PFD0960c and PFB0490c were not phosphorylated as much as was detected for PFB0490c, all three of these proteins were directly phosphorylated by PfPK7 to some extent.

### 3.3.5 Revealing the Role of PfPK7 in Plasmodium falciparum Signaling Pathways

It was interesting to note that PfPK7 was only able to directly phosphorylate three of the tested potential substrates and was not able to directly phosphorylate four of the tested potential substrates. Out of the three phosphorylated proteins, two contained the S*xD motif and one contained the RxxS* motif. The ability of PfPK7 to utilize the RxxS* motif was tested by determining the ability of PfPK7 to phosphorylate four different proteins containing this motif and only one was phosphorylated, suggesting that PfPK7 does not directly recognize the RxxS* phosphorylation motif and instead may be utilizing the S*xD motif. Further supporting PfPK7’s use of the S*xD phosphorylation motif is the fact that the substrate with the most robust phosphorylation (PF10_0257) contained this motif. Due to the inability of PfPK7 to directly phosphorylate all of the proteins identified with decreased phosphorylation in *Plasmodium falciparum* when PfPK7 was absent, it is likely that PfPK7 may act as an upstream regulatory kinase responsible for phosphorylating intermediary kinases. One avenue that needs to be investigated further is whether PfCK1, PfCK2, PfPKB (*Plasmodium falciparum* homologue of AKT), or PfPK2 (*Plasmodium falciparum* homologue of CLK2) are the intermediary kinases, activated when phosphorylated by PfPK7, which then phosphorylate the proteins detected in the phosphoproteomic analysis displaying decreased phosphorylation in the absence of PfPK7. PfCK1, PfCK2, PfPKB, and PfPK2 are likely candidates of intermediary kinases in the PfPK7 signaling cascade due to the fact that the RxxS* phosphorylation motif recognized by these kinases are significantly enriched among the 239 decreased phosphorylation events detected in the absence of
PfPK7; however, PfPK7 does not seem to directly phosphorylate this motif. Furthermore, PfCK1, PfCK2, and PfPKB are expected to have a role in the PfPK7 signaling pathway because the phosphorylated peptides with the largest decreases in phosphorylation when PfPK7 is no longer present are predicted to be phosphorylated by PfCK1, PfCK2, and PfPKB. The potential regulatory role of PfPK7 phosphorylation in a signaling cascade is depicted in Fig. 24.

3.4 Conclusions

A major challenge in understanding the noncanonical nature of the *Plasmodium falciparum* cell cycle is the fact that nothing is currently known about malaria signaling cascades. In depth analysis of the molecular switches governing intraerythrocytic cell cycle progression is needed; and characterizing the protein kinases and their substrates involved in this process is crucial. However, a significant obstacle in defining cellular signaling networks is to establish physiological kinase-substrate relationships. Many *Plasmodium* protein kinases are atypical, possessing hybrid features, marking them as unique *Plasmodium*-specific targets for therapeutic intervention. This kinase negative phosphoproteomic study identified 146 proteins with decreased phosphorylation at the schizont and segmenter stages in the absence of PfPK7, suggesting the role of PfPK7 in their phosphorylation. Although phosphoproteome profiling may provide clues to substrate identification, establishing direct kinase substrate relationships require activity-based assays. Our analysis also supports this statement because of the identified potential substrates, PfPK7 was only able to directly phosphorylate three of the tested proteins. Furthermore, of the three phosphorylation motifs that are enriched among phosphorylated sites in *Plasmodium* proteins that were down regulated, the S*xD* motif is present in two of the substrates. The inability of PfPK7 to directly phosphorylate all of the phosphoproteins that displayed decreased phosphorylation in the absence of
PfPK7, suggests a role of PfPK7 as an upstream regulatory kinase. Our phosphoproteomic analysis of the PfPK7 kinase deficient cell line highlights the limitations of this approach because it is difficult to ascertain if a change in protein phosphorylation is a direct effect due to the loss of PfPK7 or the orchestration of events of a signaling cascade. Although it is beyond the scope of this work, a recently developed approach to experimentally determine kinase-substrate pairs through the integration of data obtained from phosphorylation reactions on protein microarrays, bioinformatics, and phosphoproteomics [148] would possibly be a better approach to generate a high resolution signaling network. Nevertheless, our data defines the proteomics landscape in Plasmodium whose phosphorylation is affected due to the loss of PfPK7 function and is a step towards better understanding the mechanism of cellular signaling in the parasite.
Figure 17: Verification of the Presence of PfPK7 in the Wild Type 3D7 Cell Line and the Absence of PfPK7 in the PfPK7- Cell Line

(A) PCR analysis to verify the presence or absence of PfPK7 in the representative cell lines. The PCR analysis was performed with 25ng and 50ng of genomic DNA for both cell lines. (B) One microgram of digested 3D7 and PfPK7- genomic DNA was run on a 0.8% agarose gel. The image demonstrates equal loading of DNA for the two cell lines prior to transfer to the Nytran SPC membrane. (C) Autoradiogram of the Southern blot analysis performed using a full length PfPK7 probe against the wild type 3D7 cell line and the PfPK7- cell line.
Figure 18: Representative Images of 3D7 Wild Type and PfPK7- Time Points Harvested for Phosphoproteomic Analysis

Geimsa stained images of the cultures harvested for the two cell lines at the two stated intraerythrocytic stages. (A) Representative images of the schizont and segmenter stage parasites harvested for the wild type 3D7 cell line. (B) Representative images of the schizont and segmenter stage parasites harvested for the PfPK7- cell line.
Figure 19: Overview of the Procedure for Phosphopeptide Preparation and Analysis
Duplicate samples from each time point for each cell line were pooled and processed to generate the lists of phosphorylation sites. Protein samples were digested with LysC, labeled with TMT, fractionated on a SCX column, and enriched for phosphopeptides using IMAC. The resulting phosphopeptide samples were analyzed on an LTQ-Velos Oribtrap mass spectrometer. SEQUEST was used to identify spectra and the data was filtered to a protein-level false discovery rate of 1%. Phosphorylation site localization was analyzed through AScore and phosphorylation sites with an AScore above 13 were considered to be localized. Phosphorylation sites with a greater than 1.5-fold decrease at both stages were classified as potential PfPK7 substrates and MotifX analysis was performed. After potential PfPK7 substrates were cloned, in vitro kinase assays were performed to test the ability of PfPK7 to phosphorylate the substrates.
Figure 20: Identification of the Phosphoproteome from the Schizont and Segmenter Intraerythrocytic Stages in the Presence and Absence of PfPK7

Intraerythrocytic Stages in the Presence and Absence of PfPK7

(A) The table depicts the number of phosphorylation events that were classified as belonging to each cell line. The phosphorylation events that did not display greater than a 1.5-fold change between cell lines were classified as both lines. Three phosphorylation events that were detected in the schizont stage were not detected in the segmenter stage; therefore, there were only 3,872 phosphorylation events detected at this stage. (B) Number of phosphorylation sites classified at the schizont and segmenter stages for the wild type 3D7 cell line, PfPK7- cell line, and both cell lines.
Figure 21: Examination of Decreased Phosphorylation Events in the PfPK7- Cell Line
(A) A total of 1,490 phosphorylation events were decreased at the schizont stage, 609 were decreased at the segmenter stage, and 239 were decreased in the PfPK7 cell line at both stages when compared to the wild type. (B) Functional profiles of identified proteins using GO annotations downloaded from PlasmoDB (www.plasmodb.org) or UniProt (www.uniprot.org) as defined by the MIPS catalogue. To avoid redundancy, we assigned only one class per protein. A comparison of the percentage of proteins in each functional group in the overall phosphoproteomic data set compared to the proteins identified with decreased phosphorylation in the absence of PfPK7. The comparison shows that both data sets had the same relative distribution of proteins in each of the functional categories. (C) MotifX analysis was performed to compare the phosphorylation peptides in the overall data set to those with decreased phosphorylation in the absence of PfPK7. Three motifs were identified that are enriched among pS sites in Plasmodium falciparum that were down-regulated. When compared to the entire data set, the RxxS* motif had the best fold increase (4.12-fold) in the proteins with decreased phosphorylation when PfPK7 was knocked out.
Table 2: Quantification of the Top Five Decreased Phosphorylation Events in the PfPK7* Cell Line

Table 2A: Quantification of the Top Five Decreased Phosphorylation Events in the PfPK7* Cell Line at the Schizont Stage

<table>
<thead>
<tr>
<th>Accession Number</th>
<th>Annotation</th>
<th>Site</th>
<th>Motif Peptide</th>
<th>Schizont Score</th>
<th>Segmenter Score</th>
<th>Schizont Score</th>
<th>Segmenter Score</th>
<th>Schizont Comparison</th>
<th>Segmenter Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>O9KT20</td>
<td>Conserved Plasmodium Membrane Protein</td>
<td>S499</td>
<td>YLKDJKSLIEKK</td>
<td>244.08</td>
<td>1300.64</td>
<td>5.80</td>
<td>784.94</td>
<td>0.02377</td>
<td>0.60349</td>
</tr>
<tr>
<td>Q8BC01</td>
<td>Cg4 Protein</td>
<td>T863</td>
<td>EKEKSTKNENSA</td>
<td>91.53</td>
<td>394.09</td>
<td>2.37</td>
<td>90.28</td>
<td>0.02593</td>
<td>0.22909</td>
</tr>
<tr>
<td>Q9TV99</td>
<td>Knob-associated Histidine-Rich Protein</td>
<td>S558</td>
<td>EAKTEASTKEAT</td>
<td>6.95</td>
<td>1278.25</td>
<td>0.195</td>
<td>398.84</td>
<td>0.02799</td>
<td>0.31202</td>
</tr>
<tr>
<td>Q8IE2</td>
<td>Cyclic Related Protein</td>
<td>S688</td>
<td>WNDLESEKGRGR</td>
<td>114.56</td>
<td>580.33</td>
<td>4.39</td>
<td>126.73</td>
<td>0.03838</td>
<td>0.21837</td>
</tr>
<tr>
<td>Q8IDF2</td>
<td>Conserved Plasmodium Protein</td>
<td>S193</td>
<td>PKNLQHSDNEKNQ</td>
<td>199.88</td>
<td>618.91</td>
<td>7.72</td>
<td>129.03</td>
<td>0.03860</td>
<td>0.20847</td>
</tr>
</tbody>
</table>

Table 2B: Quantification of the Top Five Decreased Phosphorylation Events in the PfPK7* Cell Line at the Segmenter Stage

<table>
<thead>
<tr>
<th>Accession Number</th>
<th>Annotation</th>
<th>Site</th>
<th>Motif Peptide</th>
<th>Schizont Score</th>
<th>Segmenter Score</th>
<th>Schizont Score</th>
<th>Segmenter Score</th>
<th>Schizont Comparison</th>
<th>Segmenter Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>O9KT20</td>
<td>Smara Related Protein</td>
<td>S1327</td>
<td>DDDDDNSVDAKYN</td>
<td>127.64</td>
<td>11.46</td>
<td>15.87</td>
<td>0</td>
<td>0.12431</td>
<td>0</td>
</tr>
<tr>
<td>Q8BZ5</td>
<td>Cg7 Protein</td>
<td>S171</td>
<td>KNNLSYEEKOL</td>
<td>418.79</td>
<td>183.79</td>
<td>218.22</td>
<td>13.30</td>
<td>0.52108</td>
<td>0.07236</td>
</tr>
<tr>
<td>Q8FM2</td>
<td>Surface- Associated Interspersed Genes 4.2</td>
<td>S1395</td>
<td>ELFDGKSEEWEEK</td>
<td>872.13</td>
<td>73.64</td>
<td>236.55</td>
<td>10.13</td>
<td>0.27123</td>
<td>0.13761</td>
</tr>
<tr>
<td>Q8ID3</td>
<td>Myc-like DNA-Binding Domain</td>
<td>S2073</td>
<td>ISNNKSIISNNK</td>
<td>482.59</td>
<td>1251.63</td>
<td>41.47</td>
<td>202.91</td>
<td>0.06592</td>
<td>0.16212</td>
</tr>
<tr>
<td>Q7384</td>
<td>Conserved Plasmodium Protein</td>
<td>S1049</td>
<td>NDKNKQSEIODDG</td>
<td>67.78</td>
<td>287.39</td>
<td>11.13</td>
<td>49.79</td>
<td>0.16414</td>
<td>0.16742</td>
</tr>
</tbody>
</table>

(A) The top five phosphorylation events exhibiting a greater than 1.5-fold difference between the wild type 3D7 cell line and the PfPK7* cell line at the schizont stage. The representative phosphorylation events displayed the largest decrease in phosphorylation in the absence of PfPK7 at the schizont stage. The columns corresponding to the schizont and segmenter scores for the 3D7 wild type and PfPK7* cell lines are the normalized intensities associated with each particular stage.
(B) The top five phosphorylation events exhibiting a greater than 1.5-fold difference between the wild type 3D7 cell line and the PfPK7− cell line at the segmenter stage. The representative phosphorylation events displayed the largest decrease in phosphorylation in the absence of PfPK7 at the segmenter stage. The columns corresponding to the schizont and segmenter scores for the 3D7 wild type and PfPK7− cell lines are the normalized intensities associated with each particular stage.
### Table 3: Identification of Putative PfPK7 Substrates

#### Table 3A: Identification of Putative PfPK7 Substrates

<table>
<thead>
<tr>
<th>Annotation</th>
<th>PlasmoDB ID</th>
<th>Phosphorylated Peptide</th>
<th>Site Position</th>
<th>Biological Process</th>
<th>Molecular Function</th>
<th>Predicted Kinase Responsible</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knob-Associated Histidine-rich protein</td>
<td>PFB0100c</td>
<td>EATKEASTKEAT</td>
<td>558</td>
<td>Unclassified</td>
<td>Protein Binding</td>
<td>AKT, CK1, CK2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>KSTNAATPGAKDK</td>
<td>506</td>
<td></td>
<td></td>
<td>AK1, CK1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SADNKTNAATPG</td>
<td>502</td>
<td></td>
<td></td>
<td>CK1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TSADNKTNAATP</td>
<td>501</td>
<td></td>
<td></td>
<td>CK1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DAEVSKSKKHSH</td>
<td>406</td>
<td></td>
<td></td>
<td>CK1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATKEASTTEGATK</td>
<td>589</td>
<td></td>
<td></td>
<td>AKT, CK1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GATKEASTTEGAT</td>
<td>588</td>
<td></td>
<td></td>
<td>AKT, CK2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DNEDEAESVSKKH</td>
<td>403</td>
<td></td>
<td></td>
<td>CK1, AKT</td>
</tr>
<tr>
<td>Conserved Plasmodium protein</td>
<td>PF10_0257</td>
<td>QKVDQDGSADEKKD</td>
<td>250</td>
<td>Unclassified</td>
<td>Unknown</td>
<td>CK1, CK2, AKT</td>
</tr>
<tr>
<td>Fibrillarin, putative</td>
<td>PF14_0068</td>
<td>GNFKRNNSNNFGKG</td>
<td>16</td>
<td>rRNA Processing</td>
<td>RNA binding Methyltransferase Activity</td>
<td>PKA, CK1</td>
</tr>
<tr>
<td>RNA Binding Protein, putative</td>
<td>PF10_0068</td>
<td>LKNEKNSSKELIG</td>
<td>205</td>
<td>Unclassified</td>
<td>Nucleic acid binding Nucleotide binding</td>
<td>CK1, CK2, AKT</td>
</tr>
<tr>
<td>60S Ribosomal Protein L7Ae/L30e, putative</td>
<td>PFD0960c</td>
<td>ILKTNNSLDDEEKQ</td>
<td>107</td>
<td>Translation</td>
<td>Structural constituent of ribosome</td>
<td>CK1, CK2</td>
</tr>
<tr>
<td>Conserved Plasmodium protein, unknown function</td>
<td>PF11555w</td>
<td>HFRRVPSSSEKAD</td>
<td>16</td>
<td>Unclassified</td>
<td>Unknown</td>
<td>PKA, CK1, CK2, AKT</td>
</tr>
</tbody>
</table>

#### Table 3B: Identification of Putative PfPK7 Substrates with the RxxS* Motifs

<table>
<thead>
<tr>
<th>Annotation</th>
<th>PlasmoDB ID</th>
<th>Phosphorylated Peptide</th>
<th>Site Position</th>
<th>Biological Process</th>
<th>Molecular Function</th>
<th>Predicted Kinase Responsible</th>
</tr>
</thead>
<tbody>
<tr>
<td>DnaJ/Sec63 protein, putative</td>
<td>PF13_0102</td>
<td>NKKRVSSDDEDDDD</td>
<td>605</td>
<td>Post-translational protein targeting to membrane</td>
<td>Protein transporter activity</td>
<td>CK1, CK2, AKT</td>
</tr>
<tr>
<td>Conserved Plasmodium protein, unknown function</td>
<td>PFB0490c</td>
<td>FKGRFGSTSGSRVF</td>
<td>111</td>
<td>Unclassified</td>
<td>Unknown</td>
<td>PKA, AKT</td>
</tr>
<tr>
<td>Conserved Plasmodium protein</td>
<td>PF14_0190</td>
<td>NKKREISDDDDDN</td>
<td>405</td>
<td>Unclassified</td>
<td>ATP Binding</td>
<td>AKT, CK2, PKA</td>
</tr>
</tbody>
</table>
(A) The top six proteins exhibiting the largest decrease in phosphorylation in the PfPK7- cell line when compared to the wild type 3D7 cell line at the schizont and segmenter stages. The phosphorylated peptide was analyzed using GPS 3.0 (Group-based Prediction System) to predict the kinase/s that may be responsible for the phosphorylation event. The columns also describe the biological process and molecular function of the phosphorylated protein as stated using GO term analysis. (B) The top three proteins exhibiting the largest decrease in phosphorylation at a peptide containing the RxxS* motif in the PfPK7- cell line when compared to the wild type 3D7 cell line at the schizont and segmenter stages. The phosphorylated peptide was analyzed using GPS 3.0 to predict the kinase/s that may be responsible for the phosphorylation event.
Figure 22: Validation of Putative PfPK7 Substrates
(A) Coomassie blue stained gel of the \textit{in vitro} kinase assays. Lane 1: Master mix, Lane 2: MBP, Lane 3: PfPK7, Lane 4: PfPK7 + MBP, Lane 5: PF10_0257, Lane 6: PfPK7 + PF10_0257, Lane 7: PF14_0068, Lane 8: PfPK7 + PF14_0068, Lane 9: PF10_0068, Lane 10: PfPK7 + PF10_0068. (B) Phosphoimage of the \textit{in vitro} kinase assay gel depicted in (A). (C) Coomassie blue stained gel of the \textit{in vitro} kinase assays. Lane 1: PFD0960c, Lane 2: PfPK7 + PFD0960c, Lane 3: PFI1555w, Lane 4: PfPK7 + PFI1555w, Lane 5: PF13_0102, Lane 6: PfPK7 + PF13_0102, Lane 7: PFB0490c, Lane 8: PfPK7 + PFB0490c, Lane 9: PF14_0190, Lane 10: PfPK7 + PF14_0190. (D) Phosphoimage of the \textit{in vitro} kinase assay gel depicted (C). (E) Normalized signal intensity values. Master mix alone was set as the 0\% value and PfPK7 auto-phosphorylation was set as the 100\% value.
Figure 23: PfPK7’s Potential Regulatory Role in *Plasmodium* Signaling Networks

Schematic representation of PfPK7’s potential regulatory role in *Plasmodium*. Protein kinases are depicted in blue, proteins containing the S*xD* motif are purple, proteins containing the RxxS* motif are green, and proteins containing the RxxxxxS* motif are orange. PfPK7 directly phosphorylates PF10_0257, PFD0960c, and PFB0490c. PfPK7 may act as an upstream regulatory kinase, phosphorylating PfCK1, PfCK2, PfPKB, PfPKC, and/or PfPK2 activating these kinases through phosphorylation (shown with dashed lines). The activated kinase/s may then phosphorylate the proteins identified with decreased phosphorylation when PfPK7 was absent.
CHAPTER 4: ELUCIDATING THE FUNCTIONS OF *PLASMODIUM FALCIPARUM* PROTEIN KINASES THROUGH SUBSTRATE/INTERACTOR CHARACTERIZATION

4.1 Summary

The life cycle of the malaria parasite is complex and oscillates between stages of intense cell proliferation and stages where the cell withdraws from proliferation. It is expected that members of cyclin-dependent kinases and other protein kinases function as key regulators of the *Plasmodium* cell division stages, similarly to other eukaryotes. Although a number of *Plasmodium* protein kinases, including CDKs, have been identified, very little is known about their precise physiological functions. For a better understanding of the protein kinases that regulate the intraerythrocytic developmental cycles of *Plasmodium* we recently reported comprehensive isobaric-tag based quantitative proteomics and phosphoproteomics analyses. This study identified 43 distinct phosphorylation motifs and a range of proline-directed potential MAPK/CDK substrates. Additionally, to identifying substrates and interactors of *Plasmodium* CDK-like kinases, we have used HA-tagged PfPK6 and Pfmrk cell lines where endogenous protein kinase loci were replaced with epitope-tagged genes by homologous recombination. Using co-immunoprecipitation of the HA-tagged PfPK6 and Pfmrk baits with highly specific anti-HA antibody cross-linked to agarose beads, followed by mass spectrometric analyses of tryptic peptides, we have identified the components of the protein interaction complexes of these kinases. To eliminate non-specific interactions we have developed a novel filtering system based on Shannon entropy, using the number of spectral counts observed in a pair of replicates. The validity of our scoring system is underscored by the identification of Pfcyc-1 and PfMAT1 as specific interactors of Pfmrk, which have been previously reported. Our analyses of HA-PfPK6 and HA-Pfmrk immunoprecipitates have
identified 15 and 21 proteins in the interaction complex, respectively. Among the proteins identified in the PfPK6 interaction complex are many with nuclear roles e.g. Nucleolar GTP-binding protein 1 (PFF0625w) and an uncharacterized protein (MAL8P1.127) with a role in mRNA processing and RNA splicing. Components of the Pfmrk interaction complex appear to contain proteins with diverse functions such as a DNA excision-repair helicase homolog (PFI1650w) and a methyltransferase (PF07_0015). The ability of recombinant PfPK6 to interact and/or utilize any of the proteins identified in the interaction complex as substrates revealed the ability of PfPK6 to directly phosphorylate a regulator of chromosome condensation, putative (MAL7P1.38). Furthermore, site-directed mutagenesis of PfPK6 demonstrated the necessity of the residue T178 in the T-loop for activation of PfPK6 through auto-phosphorylation in order for PfPK6 to phosphorylate its substrates.

4.2 Materials and Methods

4.2.1 Plasmodium falciparum Culture

The HA-tagged PfPK6 and Pfmrk cell lines were kindly provided by C. Doerig (Monash University, Clayton Australia). Parasites of Plasmodium falciparum strain 3D7 with either HA-tagged PfPK6, HA-tagged Pfmrk, or wild type 3D7 were grown at a 4-10% parastemia and 4% hematocrit in RPMI 1640 culture medium supplemented with A+ erythrocytes and 0.5% albumax as previously described [88]. Blasticidin was added to a final concentration of 2.5μg/ml in the HA-tagged cultures.

Parasite growth and parastemia were periodically monitored by Giemsa-stained blood smears. Once parastemia reached 10%, parasites were harvested from 120 plates for each cell line following established protocols [91]. Parasites were isolated by lysing the red blood cells in 0.1%
saponin followed by thorough washing in PBS. The subsequent parasite pellets were weighed and proteins were immediately extracted by lysis in an 8M urea lysis buffer supplemented with protease inhibitors (8M urea, 75mM Tris, pH 8.2, 1X HALT protease inhibitor, 1X HALT phosphatase inhibitor) [16]. Lysates were cleared by centrifugation for 10 minutes at 4°C at 14,000 rpm. Protein concentration was determined for each sample by BioRad Bradford Assay.

4.2.2 Previous Global Phosphoproteomic Analysis Identifies Putative MAPK/CDK Substrates

Our previous global phosphoproteomics analysis was used to identify putative MAPK/CDK substrates. The analysis was performed as described previously [1]. Equal quantities of duplicate protein extracts (1mg) from each of the three intraerythrocytic P. falciparum stages were reduced, alkylated, and digested with Lys-C prior to labeling with one of 6 TMT isobaric labeling reagents [87]. All 6 samples were then mixed together and separated via strong cation exchange chromatography and a total of 20 fractions were collected. The resulting 20 fractions were subjected to IMAC to specifically isolate phosphorylated peptides [92]. Phosphopeptides were analyzed via LC-MS/MS on an LTQ-Orbitrap-Velos mass spectrometer essentially as described previously [93], except that all MS/MS spectra were acquired at high resolution in the Orbitrap and used for both phosphopeptide identification and quantification. Peptides were identified by matching individual MS/MS spectra with peptide sequences using Sequest [94]. As before, searches employed a database containing Plasmodium sequences from NCBI [96] as well as human protein sequences from IPI [97]. The peptides were filtered using a multivariate approach to remove questionable identifications based on a target-decoy strategy [95, 98]. Peptides were assembled into proteins and proteins were further filtered to ensure a protein-level false discovery rate of 1%. Individual phosphorylation sites on each peptide were scored to assess how confidently the modification could
be localized to a single residue using AScore [99]. Finally, the peptides and proteins were filtered to account for protein redundancy.

To identify putative MAPK/CDK substrates, all phosphorylation events identified in the global phosphoproteomic analyses were examined for the presence of the proline-directed Ser/Thr phosphorylated peptides. The 206 identified peptides were then filtered further to recognize only those containing the strong CDK consensus pS/T-P-X-K/R motif, which revealed 45 proteins that are putative Plasmodium CDK substrates. Another method used to identify putative MAPK/CDK substrates was immunoprecipitation with PTMScan® Direct. PTMScan® Direct allows for the isolation of proteins containing post-translational modifications; more specifically, in this study, the isolation of proteins that are phosphorylated at the strong CDK consensus sequence. The PTMScan® Direct protocol was adapted from the PhosphoScan® method developed at Cell Signaling Technology with licensed use [100]. The protocol used for the PTMScan® Immunoprecipitation procedure was followed as previously described [1].

4.2.3 Potential Substrate Expression Constructs, Recombinant Protein Expression, and Purification

After analyzing the global phosphoproteomic data and performing PTMScan® Direct we identified a number of putative MAPK/CDK substrates. In order to investigate if any of these proteins are in fact substrates of PfPK6, we generated His-tagged constructs of the potential substrates. Due to the large size of many of the potential substrates, truncated versions of the proteins containing putative phosphorylation site(s) were cloned. Each open reading frame (ORF) was truncated to ~1500 base pairs while ensuring the presence of the phosphorylated peptide containing the pS/T-P-X-K/R motif. The following proteins were cloned into the pET30 EK/LIC
vector following the manufacturer’s instructions (EMD Millipore): MAL7P1.38, PFE0155w, PF11_0332, PF10_0047, and PFC0805w. All ORFs were sequence confirmed prior to further use. The potential substrates were optimally expressed in BL21-CodonPlus (RIPL) cells (Stratagene) and purified from the soluble fraction by affinity chromatography using HisLink resin and following the manufacturer’s instructions (Promega).

4.2.4 In vitro Kinase Assays

Kinase reactions to measure PfPK6 auto-phosphorylation and activity toward the potential substrates were performed by using 1μg of recombinant PfPK6 and 5μg of recombinant substrate in a total volume of 30μl as described previously [45, 139]. A standard kinase buffer containing 20mM Tris/HCl (pH 7.5), 20mM MgCl₂, 2mM MnCl₂, 10μM ATP, and 5μCi (γ-32P) was used. Kinase reactions were incubated for 30 minutes at 30°C and stopped by the addition of 5X gel loading buffer. The entire kinase reaction was loaded on a 12% SDS/polyacrylamide gel. The gels were then stained with Coomassie blue stain, de-stained using a 30% methanol and 10% glacial acetic acid solution, dried, and exposed for autoradiography. Signal intensity values were obtained and normalized by using the kinase buffer control as the 0% value and PfPK6 phosphorylation of histone as the 100% value. The ability of PfPK6 to phosphorylate potential substrates was determined by the normalized percent signal intensity compared to the 100% value of PfPK6 phosphorylation of histone.

4.2.5 In vitro Kinase Assays to Determine Key Residues for PfPK6 Activity

Previous work in our lab used site-directed mutagenesis to investigate the role of key residues within the T-loop for PfPK6’s activity. The T-loop of eukaryotic protein kinases is also referred to as the activation loop. It is a flexible loop in close proximity to the catalytic loop.
contains a key residue that, once phosphorylated, activates the kinase. Previous work generated T173A, T175A, and T178A mutated PfPK6 constructs to investigate the importance of these T-loop residues in PfPK6’s activity. The T178D and T178E mutant constructs were generated to mimic phosphorylation at residue 178. In this study, the recombinant PfPK6 mutants were tested for their affect on PfPK6 auto-phosphorylation, phosphorylation of a control substrate histone, and phosphorylation of a novel PfPK6 substrate identified through substrate predictions from our global phosphoproteomics analysis described above. In vitro kinase assays were performed using wild type PfPK6, T173A PfPK6, T175A PfPK6, T178A PfPK6, T178D PfPK6, and T178E PfPK6. The kinase assays were performed as described above.

4.2.6 Western Blot Analysis

*Plasmodium falciparum* 3D7 strain parasites, HA-tagged PfPK6 parasites, and HA-tagged Pfmrk parasites were harvested following established protocols once parastemia reached 10% [91]. Parasites were isolated by lysing the red blood cells in 0.1% saponin (w/v), followed by thorough washing in PBS. The subsequent pellets were weighed and recorded. Proteins were immediately extracted by lysis in an 8M urea lysis buffer supplemented with 1X HALT protease inhibitor (8M urea, 75mM Tris, pH 8.2) and centrifuged for 10 minutes at 4°C at 14,000rpm. A BioRad Bradford Assay was used to determine the protein concentrations of the cleared lysates. The protein lysates were stored at -80°C until use.

For Western blotting, 100μg of protein extract from each cell line was loaded on a 12% SDS-PAGE gel with protein running buffer and run at 100 volts for approximately 3 hours. After electrophoresis, the gels were rinsed in transfer buffer (NuPAGE 1X Transfer Buffer) and transferred onto nitrocellulose using the NuPAGE wet transfer module at 30 volts for 1 hour at 4°C
according to the manufacturer’s instructions (Invitrogen). The membrane was blocked in 5% milk in tris-buffered saline containing 0.1% Tween 20 for 1 hour at room temperature. The membrane was then incubated with a 1:1000 dilution of goat anti-HA primary antibody overnight in 5% milk tris-buffered saline containing 0.1% Tween 20 at 4°C. The membrane was washed 3 times for 10 minutes each in 5% milk tris-buffered saline containing 0.1% Tween 20. Next, the membrane was incubated with horseradish peroxidase-conjugated secondary antibody at a 1:2000 dilution for 1.5 hours at room temperature, washed, and developed using enhanced chemiluminescence detection according to the manufacturer’s instructions (Cell BioSciences).

4.2.7 Immunofluorescence Assays

Immunofluorescence assays were performed in order to confirm the presence of the HA-tagged PfPK6 and Pfmrk as well as determine the proteins’ cellular localization. The following immunofluorescence assay (IFA) procedure was used for both HA-tagged cell lines. A total of 3ml of culture were harvested per IFA at 4% hematocrit and 7% parastemia by centrifuging at 1,500 rpm for 3 minutes at 4°C. Each sample was washed once in serum-free RPMI media and centrifuged as before. The pellet was re-suspended in 1ml/ml of original culture fixation solution (4% paraformaldehyde 0.0075% glutaraldehyde in 1X PBS, pH 7.4) and incubated at 37°C for 30 minutes with gentle mixing. The samples were washed with 1X PBS, pH 7.4 and centrifuged at 1,850rpm for 3 minutes at 4°C. The supernatant was removed and the pellet was re-suspended in 3ml of permeabilization solution (0.1% Triton X 100 in 1 X PBS, pH 7.4), incubated for 10 minutes at room temperature with gentle mixing, centrifuged, and washed as before. Next, the pellet was re-suspended in 3ml of blocking solution (3% BSA in 1X PBS, pH 7.4), incubated for 1 hour at room temperature with gentle mixing, centrifuged, and washed as before. After removal of the
supernatant, the pellet was re-suspended in 1ml of blocking solution containing the goat anti-HA primary antibody at a 1:100 dilution. The sample was incubated overnight at 4°C with end-over-end mixing. After centrifuging at 2,000 rpm for 2 minutes at 4°C, the pellet was washed 3 times for 10 minutes each in blocking solution. The pellet was then re-suspended in 1ml blocking solution containing donkey anti goat 488 flurophore secondary antibody. From this point forward, the samples were kept covered to prevent light exposure. Samples were incubated for 2 hours at room temperature with end-over-end mixing, followed by washing as described before. The final pellet was re-suspended in 500µl of 1X PBS, pH 7.4.

In order to prepare the coverslips and slides, the coverslips were dipped in 70% molecular grade ethanol. The coverslips were placed on paper and allowed to air dry. Next, 300µl of 1mg/ml poly-L-lysine was added to each coverslip and incubated for 15 minutes. The solution was removed and coverslips air dried. Then, 200µl of 1X PBS pH 7.4 was added to the coverslip, 8µl of the immunofluorescence assay was added to the PBS, and incubated for 15 minutes. The solution was removed and the coverslip was air dried. To prepare the slide, 10µl of Dapi fluoromount was added to the slide and the coverslip was placed cell side down onto the fluoromount. The slides were stored in the dark for 24 hours prior to imaging via confocal microscopy.

4.2.8 Co-Immunoprecipitation to Identify Interacting Partners

The Pierce HA Tag IP/Co-IP kit from Thermo Scientific was used to isolate proteins that are interacting with either PfPK6 or Pfmrk. The Co-immunoprecipitation procedure was carried out for the wild type 3D7 cell line as a negative control and was also carried out on the experimental HA-tagged PfPK6 and HA-tagged Pfmrk cell lines. The procedure provided by the manufacturer was
followed with some modifications. The co-immunoprecipitation experiment for each cell line was performed in duplicate and the resulting elutions were pooled.

The following procedure was completed in duplicate for each cell line. A positive control was also performed. A total of 8 reactions were used: 2 for the 3D7 wild type control, 2 for the positive control provided by the manufacturer, 2 for the HA-tagged PfPK6, and 2 for the HA-tagged Pfmrk. A total of 4.6mg of protein was used for each of the experimental conditions as well as the 3D7 control. For the positive control, 50μl of the HA-tagged positive control diluted in 150μl of TBS was used. The appropriate amount of lysate or positive control was added to a plugged spin column. The anti-HA agarose was thoroughly re-suspended prior to dispensing 20μl of resin into each labeled spin column using a wide-bore pipette tip. The cap was then screwed on to seal each column. The lysate plus anti-HA agarose was incubated with gentle end-over-end mixing overnight at 4°C. During this time, the TBS plus 0.5% Tween-20 (TBS-T) wash solution was prepared. After lysate incubation, the cap was loosened on the column and the bottom plug was removed. Each column was placed in a collection tube and centrifuged for 10 seconds at 4°C. The flow-through was saved for SDS-PAGE analysis.

A total of 500μl of TBS-T was added to each column. The cap and bottom plug were replaced and the column was inverted three times to gently wash the agarose. After removing the bottom plug and loosening the cap, the column was placed back in the collection tube and centrifuged for 10 seconds at 4°C. The collected wash was saved for future analysis. The washing steps were repeated two additional times for a total of three washes. To elute, the bottom plug and cap were removed, the spin column was placed in a new collection tube, 10μl of elution buffer was added to the HA-agarose, the cap was loosely screwed on and gently tapped to mix, and the spin column was centrifuged for 10 seconds at 4°C. The elution steps were repeated two additional times
and collected in the same collection tube for a total of 30μl of elution. The eluent was immediately neutralized by adding 1μl of 1M Tris, pH 9.5 per 20μl of elution buffer.

In order to perform mass spectrometry analysis on the co-immunoprecipitation eluent, we resolved 10% of the eluted sample on an SDS-PAGE gel to verify the presence of sufficient proteins for subsequent detection. The eluents from duplicate experiments were pooled for a total volume of 60μl per condition; therefore, 6μl of eluent was run on an SDS-PAGE gel for each condition. After the gel was run to completion, each gel was silver stained to detect the presence of protein. First, each gel was placed in a clean staining tray and rinsed for 5 minutes with ultrapure water. The gels were fixed overnight in 100ml fixative solution (40% ethanol and 10% acetic acid) with gentle rotation. After decanting the fixative solution, gels were washed in 30% ethanol for 10 minutes. The ethanol solution was decanted and gels were incubated in 100ml of sensitizing solution (0.02% sodium thiosulfate) for 10 minutes. The sensitizing solution was removed and the gels were washed again in 30% ethanol for 10 minutes, followed by washing for 10 minutes in ultrapure water. Next, the gels were incubated in 100ml of staining solution (0.1% silver nitrate) for 15 minutes. After removing the staining solution, gels were rinsed for 30 seconds in ultrapure water. Each gel was then incubated in 100ml of developing solution (2% sodium carbonate and 0.04% formalin (37% formaldehyde)) for 4-8 minutes until bands appeared and the desired band intensity is reached. Once the appropriate intensity is achieved, 10ml of stopper solution (5% glacial acetic acid) was immediately added directly to the gel immersed in developing solution. After being incubated for 10 minutes, the solution was decanted, the gel was washed for 10 minutes with ultrapure water, and the gel was imaged. Once the presence of protein was confirmed in the elution for the HA-tagged PfPK6 and HA-tagged Pfmrk co-immunoprecipitation experiments and
the lack of protein in the elution for the 3D7 wild type control, the remaining eluent for each was analyzed by mass spectrometry analysis.

4.2.8.1 Original Scoring System Eliminates Non-Specific Interactions

In order to identify true interacting partners of PfPK6 and Pfmrk while eliminating non-specific interactors, a novel scoring system was generated based on the spectral counts obtained and Shannon entropy. Two scores were calculated for each protein identified by mass spectrometry. The first score was the Shannon entropy, which is based on the number of spectral counts observed in a pair of replicates. To calculate the first score, 0.5 was added to spectral counts for both replicate 1 and replicate 2 spectral counts (because the entropy score is undefined when the value is zero). Then the fraction of spectral counts observed in each replicate was calculated. For calculations,

rep1 = replicate 1 and rep2 = replicate 2

\[ P_1 = \frac{(rep1 + 0.5)}{(rep1 + rep2 + 1)} \]

\[ P_2 = \frac{(rep2 + 0.5)}{(rep1 + rep2 + 1)} \]

Next the entropy is calculated using the following formula:

\[ \text{Entropy} = -1 \times (P_1 \times \log_2 (P_1) + P_2 \times \log_2 (P_2)) \]

An example for protein PF10_0141 is below:

\[ P_1 = \frac{27.5}{(27.5 + 15.5)} \]

\[ P_2 = \frac{15.5}{(27.5 + 15.5)} \]

Entropy = 0.943069

For score 1, the larger values imply that spectra are consistent between replicates for a given protein. The cutoff value for score 1 was set to 0.75. Score 2 was calculated by taking the average spectral counts for each protein across replicates, for each bait.
Average = (rep1 + rep2)

An example for protein PF10_0141 is below:

Average 1 was calculated from the Pfmrk IP and average 2 was calculated from the control 3D7 IP.

Average 1 = (27.5 + 15.5) / 2 = 21.5

Average 2 = (0.5 + 0.5) / 2 = 0.5

Next the entropy was calculated using the formula discussed above:

Entropy 1 = 21.5 / (21.5 + 0.5) = 0.976

Entropy 2 = 0.5 / (21.5 + 0.5) = 0.023

For score 2, we attempted to find proteins whose spectral counts were skewed between the two conditions rather than evenly distributed between the experimental and control samples, indicating a search for smaller values. A cutoff of 0.6 was set for score 2.

Overall, a protein was called an interactor if it met 3 conditions: 1) Score 1 entropy > 0.75, implying even expression between replicates for the IP, 2) Score 2 entropy < 0.6, implying skewed expression when comparing the IP with the control, and 3) the average spectral counts for the IP were greater than the average spectral counts in the control. This scoring system was validated by the identification of two known interactors of Pfmrk, CDK-activating kinase assembly factor, putative MAT-1 like protein (PFE0610c) and Cyclin 1 homologue (PF14_0605). The identification of these two known interactors of Pfmrk validated this novel scoring system and allowed it to be used to identify novel interacting partners of both Pfmrk and PfPK6.
4.2.9 Over Expression of Recombinant Interactors Using TnT SP6 Wheat Germ Expression System

In order to confirm the interactions of PfPK6 with the proteins identified by co-immunoprecipitation we first expressed the proteins using the Promega TnT SP6 High-Yield Wheat Germ Expression System. The TnT system couples transcription and translation in order to express the proteins of interest. The six proteins with the strongest interaction with PfPK6 as determined by our novel scoring system of the co-immunoprecipitation were generated using this method. The proteins selected are PF07_0114, PF11_0154, PFF0695, PFF0625, PF11_0259, and PF11_0395. When performing PCR to generate the DNA template, the primers had the SP6 promoter sequence, Kozak region, and hybridization region added. After PCR and confirmation of the correctly sized product by agarose electrophoresis, the plasmid DNA was purified using the PureYield Plasmid Miniprep System following the manufacturer’s instructions. Next, the DNA template was used in the translation protocol. Each gene was set up in the following reaction:

- TnT SP6 High-Yield Wheat Germ Master Mix 30μl
- DNA Template 5-8μl
- (35S) Methionine 2-4μl
- Nuclease Free Water Bring Final Volume to 50μl

After all reagents were added and mixed, the translation reaction was incubated at 25°C for 2 hours. The success of the transcription/translation reaction was verified by running 2μl of the final product on an SDS-PAGE gel. Once the gel was run to completion, the gel was fixed in a 50% methanol and 7% acetic acid solution for 15 minutes. The gel was then incubated in a 7% glycerol, 7% methanol, and 7% acetic acid solution for 10 minutes. Lastly, the gel was incubated for 5 minutes in 10% glycerol. In order to dry the gel, the gel is placed on a piece of Whatman 3MM
filter paper, wrapped in plastic wrap, and dried at 80°C for 2 hours. The dried gel was then exposed to a phosphorimaging screen and imaged to detect the presence of the (35S) Methionine and determine if the translation was successful.

4.2.10 Validation of the PfPK6 Interactome through Pull-Down Analysis

To validate the interactors of PfPK6 identified by co-immunoprecipitation of HA-tagged *Plasmodium falciparum* lysates, we performed pull-down analysis of recombinant GST-PfPK6 and identified interactors generated through the TnT SP6 Wheat Germ Expression System. For this validation, we used the Pierce GST Protein Interaction Pull-Down Kit from Thermo Scientific and followed the provided instructions with some adaptations for optimization purposes. The pull-down validation was performed for four of the interactors listed above (PF11_0154, MAL8P1.127, PF11_0259, and PF11_0305) with and without GST-PfPK6. Afterward the columns were labeled and a 1:1 wash solution was prepared of TBS: Pull-down lysis buffer, 50μl of re-suspended glutathione agarose resin slurry was added to each column using a wide bore pipette tip. The resin was washed with 400μl of wash solution followed by centrifuging at 1,250 x g for 30 seconds. The washes were repeated four times for a total of five washes. In order to prevent non-specific binding of the interactors with the GST agarose resin, 3% BSA in wash solution was added to each column and incubated for 1 hour at 4°C with end-over-end mixing. After centrifuging to remove the blocking solution, the bait GST-PfPK6 was immobilized to the GST agarose resin by adding 25μg of purified GST-PfPK6 in a total of 500μl of wash buffer. For the control pull-down reactions not containing GST-PfPK6, 500μl of wash buffer was added. The GST-PfPK6 bait was incubated with the GST agarose resin for 1 hour at 4°C and was then centrifuged as before. Next, the resin was washed five times with 500μl of wash solution followed by centrifuging at 1,250 x g for 30
seconds. Following the immobilization of the bait protein and thorough washing, the prey protein was captured. A total of 10μl of the TnT reaction for each interactor was added in 500μl of wash buffer. The prey incubated on the GST agarose resin for 6 hours at 4°C with end-over-end mixing. After centrifuging at 1,250 x g for 30 seconds, the samples were washed ten times with 500μl of wash solution with 50mM NaCl added. After thorough washing, the bait-prey was eluted by adding 20μl of elution buffer (10mM glutathione in TBS pH 8.0), incubating for 5 minutes with gentle mixing, and centrifuging at 1,250 x g for 30 seconds. The pull-down samples were analyzed by SDS-PAGE and autoradiography.

4.3 Results and Discussion

4.3.1 Our Global Phosphoproteomic Analysis Identifies Putative MAPK/CDK Substrates

The intraerythrocytic cell cycle of the malaria parasite digresses from the traditional eukaryotic cell cycle model. The six identified Plasmodium CDK-related protein kinases [21, 112] are likely to be the key regulators of the unique cell cycle progression. To validate this prediction and gain insight into the physiological functions of Plasmodium CDK-related kinases, it is essential to identify their cellular substrates. CDKs are proline-directed serine/threonine protein kinases that have a strong requirement for basic amino acids, such as arginine and lysine, at the +3 site [126]. The global phosphoproteomics study performed in our lab resulted in the identification of 206 proline-directed serine/threonine phosphorylated proteins [1]. Further filtering was performed to isolate phosphorylated proteins with the strong CDK consensus sequence pS/T-P-X-K/R, which narrowed the phosphorylated proteins to 45 proteins that are putative Plasmodium CDK substrates. Among the proteins that may be phosphorylated by Plasmodium CDKs are homologs of origin recognition complex subunit 2 (Orc2) (MAL1P1.21), Orc4 (PF13_0189), and Orc1/CDK6.
(PFE0155w), which are all components of the pre-replication complex and are known CDK substrates in other eukaryotes. Once mammalian Orc proteins are phosphorylated by CDKs, this prevents helicase loading beyond G1 phase of the cell cycle and prevents re-replication [127, 128]. Among those proteins phosphorylated at the strong CDK consensus sequence are significant proteins potentially involved in regulating the parasite’s DNA replication, transcription, or mitosis such as homologs of cell division cycle protein 48 (PFF0940c) and regulator of chromosome condensation (MAL7P1.38). PTMScan® Direct was utilized to further identify putative Plasmodium CDK substrates. This technique isolates proteins with specific post-translational modifications, in our case, proteins that are phosphorylated at the strong CDK consensus sequence. Using this antibody-based pull-down of phospho-motif peptides in duplicate experiments, we identified a total of 48 unique phosphorylation events. Significantly, the PTMScan® immunoprecipitation identified two phosphorylation events, homologs of cell division cycle protein 48 (PFF0940c) (S512) and regulator of chromosome condensation (MAL7P1.38) (S602) that were predicted MAPK/CDK substrates from the global phosphoproteomic analysis. It would be crucial to identify which substrates the kinase is acting on to delineate the role of PfPK6 in the Plasmodium cell cycle. Five of the predicted CDK substrates identified by the global phosphoproteomics study or by PTMScan® Immunoprecipitation were investigated further. These five putative substrates are listed in Table 4 along with their peak phosphorylation stage as well as the biological process and molecular function of the phosphorylated protein.

4.3.2 Characterization of PfPK6 and Identification of Novel PfPK6 Substrates

In an attempt to further unravel PfPK6’s role in Plasmodium falciparum we performed immunofluorescence assays to determine its subcellular localization. Stage-specific examination of
the kinase revealed an extra-nuclear cellular localization (Fig. 25A) as evident from PfPK6 not co-localizing with the nuclear stain DAPI (Fig. 25B). Although PfPK6 is a CDK-like kinase, it is clearly not nuclear and may function by phosphorylating proteins that are exported to and/or from the nucleus. PfPK6 was further characterized by determining its stage-specific profile. Western blot analysis of 8 hour time point lysates reveals little PfPK6 expression at the late ring and early trophozoite stages, while expression reaches its peak at the late intraerythrocytic stages (Fig. 25C).

After examining the phosphoproteins with phosphorylation events occurring at the strong CDK consensus sequence, we generated His-tagged recombinant forms of five proteins identified through our previous global phosphoproteomics analysis or previous PTMScan® Immunoprecipitation (Table 4). One of the chosen proteins, regulator of chromosome condensation (MAL7P1.38) was identified through both types of analysis. To investigate the role of PfPK6 in *Plasmodium falciparum* we performed *in vitro* kinase assays to identify substrates of this CDK-like kinase. Fig. 26A and 26C show 1μg of recombinant PfPK6 and 5μg of substrate was used for each assay condition. After four independent kinase assays, it was clear that PfPK6 was able to phosphorylate putative regulator of chromosome condensation (MAL7P1.38) to a greater extent than the control substrate histone or greater than it’s own auto-phosphorylation (Fig. 26B). PfPK6 was also able to phosphorylate RNA binding protein (PF10_0047) to a much lesser extent (Fig. 26D). The ability of PfPK6 to phosphorylate these substrates was determined by normalizing signal intensity values to the master mix control (0%) and PfPK6 phosphorylation of histone (100%). PfPK6 phosphorylation of MAL7P1.38 was 276%, which is much greater than PfPK6’s auto-phosphorylation or phosphorylation of the control histone at 100% (Fig. 26E). PfPK6’s phosphorylation of PF10_0047 was 50% (Fig. 26E). The ability of PfPK6 to phosphorylate a putative regulator of chromosome condensation and a RNA binding protein is suggestive of its
nuclear function although the nuclear localization of PfPK6 is not evident. It is possible that these substrates of PfPK6 shuttle in and out of the nucleus.

### 4.3.3 Key T-loop Residues Greatly Affect PfPK6 Auto-phosphorylation and Activity Towards Substrates

PfPK6 demonstrates homology to both CDKs and MAPKs; however, molecular modeling of PfPK6 suggests its structure is more closely related to CDKs. The regulatory phosphorylation sites, typically found in CDKs, are conserved in PfPK6 (Thr14, Thr15, and Thr160 in CDK2); however, PfPK6 lacks the canonical cyclin binding PSTAIRE motif, which is replaced by a SKCILRE sequence [19]. Further evidence supporting the cyclin-independent activity of PfPK6 is the fact that the kinase shows significant auto-phosphorylation and phosphorylation of the substrate histone in the absence of a cyclin binding partner and PfPK6 activity is not stimulated by incubation with a cyclin [19]. By examining the T-loop of PfPK6 and comparing it to eukaryotic CDKs, it would be predicted that Thr173 phosphorylation is required for full activation of PfPK6. To investigate the role of T173 of PfPK6 in its catalytic activity, we tested the property of a T173A mutant. However, in vitro kinase assays using a T173A mutated PfPK6 resulted in no effect on PfPK6 auto-phosphorylation, phosphorylation of the control substrate histone, or phosphorylation of the novel substrate MAL7P1.38 (Fig. 27A, Fig. 27B, and Fig. 27E). The fact that Thr173 phosphorylation has no effect on PfPK6 activity suggests that this Thr residue has no role in the catalytic activity of PfPK6 unlike CDKs. Furthermore, in MAP kinases the activation loop contains a signature TxY motif, whereas in PfPK6 it is TxT. Our results thus supports the notion that PfPK6 is a unique “cyclin-independent” CDK.
Because there are two additional Thr residues adjacent to Thr173, we investigated the role of Thr175 and Thr178 by mutating them to Ala. Both of these threonine residues are also located in the T-loop and may have an effect on PfPK6’s phosphorylation. The *in vitro* kinase assays revealed the T175A mutation had no effect on PfPK6 auto-phosphorylation or substrate phosphorylation. In fact, the mutation of Thr175 to an alanine increased its activity. PfPK6 phosphorylation of the control histone by T175A was 127% compared to the 100% of histone phosphorylation in the wild type, and was 162% for the novel substrate MAL7P1.38 (Fig. 27E). The absence of an effect in PfPK6 activity when Thr175 is mutated suggests similar to Thr173, this residue has no influence on the catalytic function of PfPK6. Next, we investigated the role of Thr178 residue within the T-loop for its significance in PfPK6 auto-phosphorylation or substrate phosphorylation. When Thr178 was mutated to an alanine residue, PfPK6 exhibited a total loss of both auto-phosphorylation and substrate phosphorylation (Fig. 27C, Fig. 27D, and Fig. 27E). This result suggests that Thr178 is crucial for both auto-phosphorylation and substrate phosphorylation. The phospho-mimic mutants T178D or T178E were unable to restore PfPK6 activity (Fig. 27C, 27D, and 27E). Overall the site-directed mutagenesis showed the significance of phosphorylation at Thr178 in leading to full activation of PfPK6, while Thr173 and Thr175 phosphorylation had no effect on activity.

**4.3.4 Co-Immunoprecipitation Identifies the Interactome of PfPK6 and Pfmrk**

In addition to identifying substrates, the role a protein kinase plays can be determined by detecting interacting partners. To identify substrate/interactors of *Plasmodium* CDK-like kinases, we used HA-tagged PfPK6 and Pfmrk lines where endogenous protein kinase loci were replaced with epitope-tagged genes by homologous recombination. Western blot analysis using an anti-HA antibody confirmed the presence of the HA-tag in both the PfPK6 and Pfmrk cell lines and the
absence of HA in the 3D7 wild type control (Fig. 28A). After confirming the presence of the HA-tagged kinases in their respective cell lines, co-immunoprecipitation was performed with HA-tagged PfPK6 and Pfmrk as baits with highly specific anti-HA antibody cross-linked to agarose beads, followed by mass spectrometric analyses of tryptic peptides (Fig. 28B-28D). The co-immunoprecipitation identified the components of the protein interaction complexes of these kinases. Non-specific interactions were eliminated through the use of a novel filtering system based on Shannon entropy and the number of spectral counts observed in a pair of replicates. The validity of our scoring system is underscored by the identification of Pfcyc-1 and PfMAT1 as specific interactors of Pfmrk, which have been previously reported. Our analyses of HA-PfPK6 and HA-Pfmrk immunoprecipitates have identified 15 and 21 proteins in the interaction complex, respectively. The top interactors of Pfmrk are listed in Table 5A, while the top interactors of PfPK6 are listed in Table 5B. Interestingly, all of the proteins identified as interactors of either PfPK6 or Pfmrk were only found as interactors of that particular kinase. There was no crossover of identified interactors between the co-immunoprecipitation experiments. Among the proteins identified in the PfPK6 interaction complex are many with nuclear roles e.g. Nucleolar GTP-binding protein 1 (PFF0625w) and an uncharacterized protein (PFF0695w) with a role in mRNA processing and RNA splicing. Components of the Prmrk interaction complex appear to contain proteins with diverse functions such as a DNA excision-repair helicase homolog (PFI1650w) and a methyltransferase (PF07_0015).

In an effort to validate the interactors identified by mass spectrometry of the immunoprecipitated complex of HA-PfPK6 we tested binary interactions between GST-PfPK6 as the bait and the interactors identified above as the prey. The TnT SP6 wheat germ protein expression system was used to produce the previously identified radiolabeled interactors of PfPK6
(PF11_0154, MAL8P1.127, PFF0625w, PF11_0259, and PF11_0305) (Fig. 29A). We were successful in expressing four of the interactors. GST-PfPK6 was immobilized to anti-GST agarose beads followed by incubation with the interactors. The pull-down analysis confirmed the interaction of PfPK6 with these four proteins (Fig. 29B). When comparing the amount of protein pulled-down in the presence and absence of PfPK6, it is clear that PfPK6 does indeed interact with these proteins as is evident by the enrichment of prey protein when PfPK6 was included (Fig. 29C). We used the identified interactors and substrates of PfPK6 in this study as well as previously established interactors published on PlasmoDB to map the preliminary interactome of PfPK6 (Fig. 30).

4.4 Conclusions

The malaria parasite, *Plasmodium falciparum*, undergoes complex developmental cycles in its human and insect hosts, which involves fluctuations between stages of intense cell proliferation and stages of withdrawal from proliferation. In traditional eukaryotic cells, members of the cyclin-dependent kinase family as well as other protein kinases regulate progression of the cell cycle, which is expected to be the case for *Plasmodium* cell division as well. Several CDK homologues, e.g. PfPK5, PfPK6, and Pfmrk, have been identified in *Plasmodium falciparum* [19]. Of these CDK-like malaria parasite kinases, PfPK6 possesses novel features: (a) it exhibits significant homology to MAP kinases although the PfPK6 molecular model shows structural similarity to the CDKs; (b) the canonical PSTAIRE cyclin binding motif is changed to SKCILRE in PfPK6; (c) the recombinant protein shows robust auto-phosphorylation activity and unlike PfPK5 and Pfmrk, PfPK6 activity is not stimulated by any cyclin; (d) similar to tyrosine kinases, PfPK6 prefers Mn$^{2+}$ over Mg$^{2+}$ as the divalent cation; and (e) the sensitivity of PfPK6 to the CDK inhibitor roscovitine is more analogous to MAP kinases rather than CDK1, CDK2, or CDK5 [19, 37].
To gain a better understanding of the cellular function of this atypical CDK-like kinase, we investigated the substrates and interactors of PfPK6. This study utilized our phosphoproteomic dataset and PTMScan® Direct to identify putative CDK/MAPK *Plasmodium* substrates; and it revealed the ability of PfPK6 to directly phosphorylate putative regulator of chromosome condensation (MAL7P1.38) and RNA binding protein (PF10_0047). The putative regulator of chromosome condensation MAL7P1.38 has approximately 50% similarity to the regulator of chromosome condensation RCC1. The RCC1 protein functions as a RAN GTPase activating protein and is involved in mitotic spindle assembly, nuclear envelope formation, and nucleocytoplasmic transport [149]. The RNA binding protein PF10_0047 is a homologue of serine-arginine (SR) rich proteins involved in pre-mRNA splicing and are known to be involved in nucleocytoplasmic transport, thus PfPK6 can phosphorylate these proteins while localized in the cytoplasm.

In addition to identifying substrates, we used co-immunoprecipitation of HA-PfPK6 with a novel stringent scoring system to detect the interacting partners of PfPK6. Among the proteins identified in the PfPK6 interaction complex are many with nuclear roles e.g. Nucleolar GTP-binding protein 1 (PFF0625w) and an uncharacterized protein (PFF0695w) with a role in mRNA processing and RNA splicing. Therefore, the identification of these proteins through co-immunoprecipitation studies also supports the role of PfPK6 in nuclear function. Additionally, the data from our study has enabled us to construct a preliminary interactome of the kinase, shedding light on one of the many signaling pathways of the *Plasmodium* cell cycle. This is the first interactome developed for any of the protein kinases of *Plasmodium*.

Our study also defined the catalytic mechanism of PfPK6 through specification of the amino acid residue involved in auto-phosphorylation. This analysis highlights the importance of T178 in
the T-loop of PfPK6 for catalytic activity and substrate phosphorylation. It is to be noted that phosphomimic residues, glutamate or aspartate, did not rescue PfPK6 activity. Our previous molecular modeling [19] suggests that the Thr178 residue hydrogen bonds to Asp143 in the catalytic loop. However, the mutants T178A, T178E, and T178D were not able to do so.

In conclusion, this study has made significant progress in defining catalytic mechanism and the cellular role(s) of PfPK6.
Figure 24: PfPK6 Localization and Expression Pattern

(A) Immunofluorescence assay depicting the cellular localization of PfPK6 in green and nuclear localization in blue. Cellular localization of PfPK6 at the ring, trophozoite, schizont, and segmenter stages were all examined. (B) Immunofluorescence assay illustrating the extra-nuclear localization of PfPK6. PfPK6 does not co-localization with the DAPI nuclear stain. (C) Stage-specific profile of PfPK6 expression determined from western blot analysis of 8 hour 3D7 lysate time points.
Table 4: Identification of Putative Plasmodium CDK Substrates

<table>
<thead>
<tr>
<th>Annotation</th>
<th>PlasmoDB ID</th>
<th>Phosphorylated Peptide</th>
<th>Site Position</th>
<th>Biological Process</th>
<th>Molecular Function</th>
<th>Peak Protein Expression</th>
<th>Peak Phosphorylation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regulator of Chromosome Condensation, putative</td>
<td>MAL7P1.38</td>
<td>SYLSNESPTRKQP</td>
<td>602</td>
<td>Cell Cycle and Chromosome Condensation</td>
<td>Chromatin Binding</td>
<td>Global</td>
<td>Schizont</td>
</tr>
<tr>
<td>Conserved Plasmodium protein</td>
<td>PFE0155w</td>
<td>PKQYYESPFKVLK</td>
<td>730</td>
<td>DNA Replication</td>
<td>ATP Binding</td>
<td>Ring</td>
<td>Schizont</td>
</tr>
<tr>
<td>Conserved Plasmodium protein</td>
<td>PF11_0332</td>
<td>ELTKRGSPGELAK</td>
<td>33</td>
<td>Unclassified</td>
<td>Nucleic Acid Binding</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>RNA Binding Protein, putative</td>
<td>PF10_0047</td>
<td>GGDEINLSKYNKN</td>
<td>685</td>
<td>Unclassified</td>
<td>Nucleic Acid Binding</td>
<td>Ring/Trophozoite</td>
<td>Trophozoite/Schizont</td>
</tr>
<tr>
<td>DNA-Directed RNA Polymerase II, putative</td>
<td>PFC0805w</td>
<td>PAYILQSPVQIKQ</td>
<td>2410</td>
<td>Transcription from RNA polymerase II promoter</td>
<td>DNA Binding, DNA-Directed RNA Polymerase Activity, Protein Binding</td>
<td>Ring/Schizont</td>
<td>N/A</td>
</tr>
<tr>
<td>DNA-Directed RNA Polymerase II, putative</td>
<td>PFC0805w</td>
<td>NAHQMSPAYILQ</td>
<td>2403</td>
<td>Transcription from RNA polymerase II promoter</td>
<td>DNA Binding, DNA-Directed RNA Polymerase Activity, Protein Binding</td>
<td>Ring/Schizont</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Five proteins that are predicted *Plasmodium* CDK substrates were selected for further investigation to determine the ability of PfPK6 to phosphorylate them. The table lists the chosen proteins as well as the detected phosphorylation event occurring at the strong CDK consensus sequence, the site position of the phosphorylation event, the biological process and molecular function of the phosphorylated protein as determined by GO term analysis, and the peak protein expression and phosphorylation stage as revealed in our global phosphoproteomic analysis [1].
Figure 25: Identification of Novel Substrates of PfPK6
(A) Coomassie blue stained gel of the *in vitro* kinase assays. Lane 1: Master mix, Lane 2: Histone, Lane 3: PfPK6, Lane 4: PfPK6 + Histone, Lane 5: MAL7P1.38, Lane 6: PfPK6 + MAL7P1.38, Lane 7: PFE0155w, Lane 8: PfPK6 + PFE0155w. (B) Phosphoimage of the *in vitro* kinase assay gel depicted in (A). (C) Coomassie blue stained gel of the *in vitro* kinase assays. Lane 1: PF11_0332, Lane 2: PfPK6 + PF11_0332, Lane 3: PF10_0047, Lane 4: PfPK6 + PF10_0047. (D) Phosphoimage of the *in vitro* kinase assay gel depicted in (C). (E) Normalized signal intensity values. Master mix alone was set as the 0% value and PfPK6 auto-phosphorylation was set as the 100% value.
Figure 26: Key T-loop Residues Greatly Affect PfPK6 Auto-Phosphorylation and Activity Towards Substrates
(A) Coomassie blue stained gel of the *in vitro* kinase assays. Lane 1: Master mix, Lane 2: Histone, Lane 3: MAL7P1.38, Lane 4: PfPK6 wild type (WT), Lane 5: PfPK6 WT + Histone, Lane 6: PfPK6 WT + MAL7P1.38, Lane 7: PfPK6 T173A, Lane 8: PfPK6 T173A + Histone, Lane 9: PfPK6 T173A + MAL7P1.38. (B) Phosphoimage of the *in vitro* kinase assay gel depicted in (A). (C) Coomassie blue stained gel of the *in vitro* kinase assays. Lane 1: PfPK6 T175A, Lane 2: PfPK6 T175A + Histone, Lane 3: PfPK6 T175A + MAL7P1.38, Lane 4: PfPK6 T178D, Lane 5: PfPK6 T178D + Histone, Lane 6: PfPK6 T178D + MAL7P1.38, Lane 7: PfPK6 T178E, Lane 8: PfPK6 T178E + Histone, Lane 9: PfPK6 T178E + MAL7P1.38. (D) Phosphoimage of the *in vitro* kinase assay gel depicted in (C). (E) Normalized signal intensity values. Master mix alone was set as the 0% value and PfPK6 auto-phosphorylation was set as the 100% value.
Table 5: Identification of the (A) Pfmrk and (B) PfPK6 Interactomes

Table 5A: Identification of the Pfmrk Interactome

<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
<th>mrkE</th>
<th>mrkWT</th>
<th>pk6Int</th>
<th>Ratio E/WT</th>
</tr>
</thead>
<tbody>
<tr>
<td>PFE0610c</td>
<td>CDK-activating kinase assembly factor, putative</td>
<td>0.999</td>
<td>0.078</td>
<td>0</td>
<td>12.78</td>
</tr>
<tr>
<td>PF1650w</td>
<td>DNA excision-repair helicase, putative</td>
<td>0.989</td>
<td>0.110</td>
<td>0</td>
<td>8.94</td>
</tr>
<tr>
<td>PF10_0141</td>
<td>Cdk7, putative</td>
<td>0.943</td>
<td>0.156</td>
<td>0</td>
<td>6.02</td>
</tr>
<tr>
<td>PF14_0605</td>
<td>Cyclin homologue</td>
<td>0.811</td>
<td>0.139</td>
<td>0</td>
<td>5.82</td>
</tr>
<tr>
<td>PF14_0618</td>
<td>Putative uncharacterized protein</td>
<td>0.811</td>
<td>0.258</td>
<td>0</td>
<td>3.14</td>
</tr>
</tbody>
</table>

Table 5B: Identification of the PfPK6 Interactome

<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
<th>mkInt</th>
<th>pk6E</th>
<th>pk6WTE</th>
<th>Ratio of E to WT</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF07_0114</td>
<td>Uncharacterized protein</td>
<td>0</td>
<td>0.994</td>
<td>0.097</td>
<td>10.15</td>
</tr>
<tr>
<td>MAL13P1.185</td>
<td>Protein kinase 6</td>
<td>0</td>
<td>0.967</td>
<td>0.135</td>
<td>7.16</td>
</tr>
<tr>
<td>PF11_0154</td>
<td>Conserved Plasmodium protein</td>
<td>0</td>
<td>0.942</td>
<td>0.195</td>
<td>4.80</td>
</tr>
<tr>
<td>MAL8P1.127</td>
<td>Uncharacterized protein</td>
<td>0</td>
<td>0.964</td>
<td>0.468</td>
<td>2.05</td>
</tr>
<tr>
<td>PFF0695w</td>
<td>Uncharacterized protein</td>
<td>0</td>
<td>0.988</td>
<td>0.503</td>
<td>1.96</td>
</tr>
<tr>
<td>PFF0625w</td>
<td>Nucleolar GTP-binding protein 1, putative</td>
<td>0</td>
<td>0.896</td>
<td>0.503</td>
<td>1.78</td>
</tr>
<tr>
<td>PFL1800w</td>
<td>Conserved Plasmodium protein</td>
<td>0</td>
<td>0.940</td>
<td>0.543</td>
<td>1.72</td>
</tr>
<tr>
<td>PF11_0259</td>
<td>Nuclear preribosomal assembly protein, putative</td>
<td>0</td>
<td>0.940</td>
<td>0.543</td>
<td>1.72</td>
</tr>
</tbody>
</table>

The top interactors identified in the HA-tagged co-immunoprecipitation are listed. (A) The top seven interactors of Pfmrk including PFE0610c and PF14_0605 that are previously demonstrated
interactors of Pfmrk. The mrkE column shows the normalized spectral count for a particular protein in the HA-Pfmrk cell line, while the mrkWT column shows the normalized spectral count for the same protein in the wild type 3D7 cell line. The next column, pk6Int, indicates whether a protein was considered to be an interactor for PfPK6 as well as Pfmrk. Finally, the Ratio E/WT column calculates how enriched a particular protein was in the HA-Pfmrk cell line compared to the wild type 3D7 cell line. For this particular column, the higher the value, the stronger the interaction between the specific protein and Pfmrk. (B) The top eight interactors of PfPK6 as revealed after the co-immunoprecipitation. The mrkE column indicates whether a protein was considered to be an interactor for Pfmrk as well as PfPK6. The Pk6E column shows the normalized spectral count for a particular protein in the HA-PfPK6 cell line, while the PK6WTE column shows the normalized spectral count for the same protein in the wild type 3D7 cell line. The next column, the Ratio of E to WT column calculates how enriched a particular protein was in the HA-Pfmrk cell line compared to the wild type 3D7 cell line.
Figure 27: Identification of the PfPK6 Interactome

(A) Western blot analysis showing the absence of the HA tag in the 3D7 wild type control cell line and the presence of the HA-tagged proteins in either the HA-tagged PfPK6 or HA-tagged Pfmrk
cell lines. (B) Silver stained SDS-Page image showing 10% of the final elution from the 3D7 wild type control co-immunoprecipitation. (C) Silver stained SDS-Page image showing 10% of the final elution from the HA-tagged PfPK6 co-immunoprecipitation. (D) Silver stained SDS-Page image showing 10% of the final elution from the HA-tagged Pfmrk co-immunoprecipitation.
Figure 28: Verification of the PfPK6 Interactome

Pull-down analysis using GST-PfPK6 was used to confirm the interactions identified through co-immunoprecipitation. (A) Phosphoimage displaying the PfPK6 interactors generated with the TnT SP6 Wheat Germ Expression System. (B) Phosphoimage showing the ability of GST-PfPK6 to pull-down the interactors identified in the co-immunoprecipitation experiment. (C) Histogram depicting the extent of interaction of PfPK6 with the proteins identified in the co-immunoprecipitation.
The interactors of PfPK6 identified in this study as well as previously identified interactors were mapped to reveal the PfPK6 preliminary interactome. The *Plasmodium* proteins in blue are the interactors of PfPK6 identified through HA-tagged co-immunoprecipitation and confirmed by pull-down analysis. The proteins in green were top PfPK6 interactors as revealed through HA-tagged co-immunoprecipitation, but have yet to be confirmed through pull-down analysis. Proteins in orange are previously published substrates of PfPK6. The yellow protein was a novel substrate of PfPK6 classified in this study. All of the red proteins were proven interactors of their respective proteins through yeast-two hybrid experiments published on PlasmoDB.
CHAPTER 5: GENERAL DISCUSSIONS AND CONCLUSIONS

The aim of this work was to investigate the signaling mechanisms that regulate intraerythrocytic progression of *Plasmodium falciparum*. Because reversible phosphorylation plays a key role in intracellular signaling we focused on the role of protein kinases in the atypical intraerythrocytic life cycle. We defined the constitutive and regulated expression of the *Plasmodium* proteome during the intraerythrocytic developmental cycle, which provided insight into the dynamics of phosphorylation during asexual cycle progression. We identified 2,767 proteins, 1,337 phosphoproteins, and 6,293 phosphorylation sites through quantitative phosphoproteomics analyses and tracked their individual changes in expression and/or phosphorylation as the malaria parasite progressed from one intraerythrocytic stage to the next. We showed that 34% of identified proteins and 75% of phosphorylation sites are dynamic and exhibit changes in abundance as the intraerythrocytic cycle progresses. This study is the most comprehensive and the only quantitative phosphoproteomic analyses to date examining progression through the entire intraerythrocytic cell cycle, marking a tremendous step forward in delineating the molecular mechanisms governing the *Plasmodium* intraerythrocytic cell cycle.

Our success with comprehensive quantitative phosphoproteomics analyses, prompted us to apply this approach to delineate the signaling network where one of the orphan *Plasmodium* kinases, PfPK7, functions. For this analysis we compared the phosphoproteome profiles of the wild type *P. falciparum* 3D7 and the pfpk7 knockout lines. In the absence of a reliable conditional knock out system, generating kinase knockout cell lines in the haploid genome of the *Plasmodium* intraerythrocytic cell cycle stages is particularly challenging. Fortunately, the knock out of *pfpk7* is not lethal, and exhibits a phenotype of a slower growth rate as a result of reduced number of
merozoites per schizont [27]. Since PfPK7 has a known effect in *Plasmodium* and is a non-essential gene, we were able to use quantitative phosphoproteomics on a PfPK7− cell line and compare it to the wild type 3D7 cell line to reveal the potential substrates of this atypical kinase. This analysis identified 146 putative PfPK7 substrates; three of these were proven to be direct substrates of the kinase, while multiple were not directly phosphorylated. This suggests that PfPK7 may function as an upstream regulatory kinase, acting on down-stream protein kinases, which may lead to the decreased phosphorylation of the 146 proteins observed in this study. This was the first use of quantitative phosphoproteomic analyses to identify substrates of a protein kinase, which may now be used to investigate the role of other protein kinases within *Plasmodium*.

Finally, this dissertation focused on elucidating the function of the atypical CDK-like kinase PfPK6 through characterization of its substrates and interactors. Our phosphoproteomic dataset was analyzed for phosphorylation events occurring at the stringent CDK phosphorylation motif in order to identify putative CDK substrates. The ability of PfPK6 to phosphorylate these putative substrates was tested and resulted in the identification of two novel PfPK6 substrates. We also uncovered the preliminary PfPK6 interactome by detecting the direct interacting partners of PfPK6. A summation of our findings is discussed below in the context of their relevance and significance to our understanding of the molecular operations regulating the *Plasmodium falciparum* cell cycle.

### 5.1 The *Plasmodium falciparum* Proteome and Phosphoproteome

*Plasmodium falciparum* exhibits an unusual asexual intraerythrocytic cell cycle, which diverges from the traditional paradigm of cell cycle typically seen in eukaryotes. Specifically, *P. falciparum* undergoes multiple distinct rounds of DNA replication and nuclear division without
cytokinesis. Gaining a better understanding of the molecular instruments coordinating the multitude of events that progress the parasite through the intraerythrocytic developmental stages is of fundamental importance for rational design of novel intervention strategies.

To date, there have been no reports on the analysis of the dynamics of protein expression and phosphorylation during the entire intraerythrocytic developmental cycle of *Plasmodium*. In this study, we present the most comprehensive analysis of protein expression and phosphorylation across the three major developmental stages in the *P. falciparum* intraerythrocytic asexual cycle using isobaric tagging for relative and absolute quantitation [58, 87]. We identified 2,767 proteins, 1,337 phosphoproteins, and 6,293 phosphorylation sites from the ring, trophozoite, and schizont stages, the three major stages in the asexual cycle of malaria parasites in red blood cells (for more information on the *Plasmodium* life cycle, see reference [150]). This comparative analysis revealed stage-specific profiles of protein expression and phosphorylation. A wealth of insight is contained within these quantitative profiles of *P. falciparum* protein and phosphoprotein expression. Importantly, the 2,914 unique proteins detected in these proteomic and phosphoproteomic analyses comprise approximately 55% of the current proteome prediction [20]. Since many proteins of the *Plasmodium* proteome are exclusively expressed in the mosquito stages, as reported earlier [84, 85] and many others are restricted to liver and gametocyte stages during infection of the human host, we propose that our present dataset likely encompasses the majority of intraerythrocytic stage-specific proteins. This data obtained in our study have significantly enhanced numbers of *Plasmodium* protein identifications compared to previous proteomics reports [83, 91]. The greater depth of our analysis compared to previous attempts reflects multiple factors. First, by simultaneously surveying all intraerythrocytic developmental stages, each with its own protein complement, our study can encompass a larger fraction of the proteomic landscape. Second, the
continuous improvements in instrument sensitivity and methodology resulted in increased number of protein and phosphorylation identifications than that are typically obtained via proteomics analysis [92].

An intriguing aspect of intraerythrocytic differentiation of *Plasmodium* is the mechanism by which parasites regulate gene expression. Based on transcriptome analysis, it was initially believed that the majority of genes are globally controlled in a stage-dependent manner [105, 106], although it is now generally accepted that post-transcriptional and post-translational regulatory mechanisms have more dominant roles in regulation of intraerythrocytic gene expression [79, 80]. Our study supports the idea that post-transcriptional regulation of gene expression plays a major role in *Plasmodium*. Our analysis revealed that over 60% of all identified proteins are constitutively expressed at relatively constant amounts during all three intraerythrocytic stages.

Another interesting aspect of our phosphoproteome analysis is the discovery of tyrosine phosphorylation of a large number of *Plasmodium* proteins in the absence of true tyrosine kinase homologues. This underscores the importance of dual-specificity kinases in the parasite or the role of host cell tyrosine kinase. Our study also revealed that peak protein expression and phosphorylation are unlinked.

This research also identified 43 distinct phosphorylation motifs and a range of potential MAPK/CDK substrates. Among the proteins that may be phosphorylated by *Plasmodium* CDKs are homologs of origin recognition complex subunit 2 (Orc2) (MAL7P1.21), Orc4 (PF13_0189), and Orc1/CDC6 (PFE0155w), components of the pre-replication complex that are known CDK substrates in other eukaryotes. Additional pre-replicative complex proteins that are phosphorylated at the weak CDK consensus motif (pS/T-P) are MCM4, and MCM5. Mammalian ORC and MCM subunits are phosphorylated by CDKs to prevent helicase loading beyond G1 phase of the cell.
cycle, thus preventing re-replication [127, 128]. Other proteins of significance in potentially regulating the parasite’s DNA replication, transcription, or mitosis that are phosphorylated at the strong CDK motif are homologs of cell division cycle protein 48 (PFF0940c) and regulator of chromosome condensation (MAL7P1.38). Closer examination of phosphorylation events identified 30 protein kinases with 126 phosphorylation sites within the kinase domain or in N- or C-terminal tails. Many of these phosphorylation events are likely CK2-mediated. Through this global phosphoproteomics analyses we defined the constitutive and regulated expression of the Plasmodium proteome during intraerythrocytic development, offering an insight into the dynamics of phosphorylation during asexual cycle progression. Our system-wide comprehensive approach is a major step toward defining kinase-substrate pairs operative in various signaling networks in the parasite.

A major challenge in the post-genome sequencing era is to understand the cellular roles of the vast number of proteins whose function is not deducible from sequence analysis. Achieving a comprehensive functional understanding of the Plasmodium falciparum proteome is particularly challenging, given that 60% of putative proteins are Plasmodium-specific proteins that have no known orthologs in other organisms. This work brings significant contribution to the ongoing global efforts at implementing functional genomics studies of Plasmodium falciparum, the end goal of which is to facilitate the discovery of novel interference drug targets. Our global phosphoproteomic analysis is the only quantitative study that compares dynamics of phosphorylation events as the parasite progresses through its asexual developmental stages in the host erythrocyte. Our study also represents a step toward elucidation of the kinase-substrate pairs that will eventually form the basis of a comprehensive Plasmodium signaling network.
5.2 The Use of Kinase Negative Phosphoproteomics to Delineate the Role of A Protein Kinase in Plasmodium

We demonstrated for the first time that phosphoproteomic analysis can be applied to a kinase negative cell line in order to detect novel substrates and shed light on the role a protein kinase plays in Plasmodium falciparum signaling pathways. A major challenge in understanding the unique nature of the P. falciparum cell cycle is the lack of information currently known about malaria signaling cascades. Thorough investigation of the molecular operators governing the intraerythrocytic cell cycle progression is needed. Furthermore, characterization of the protein kinases and respective substrates involved in this process are of utmost importance. Because of the novelty of Plasmodium protein kinases they could be considered to be ideal novel therapeutic targets. One atypical Plasmodium protein kinase is PfPK7, which possesses hybrid features and does not show strong homology to eukaryotic protein kinases.

This kinase negative phosphoproteomic study identified 146 putative substrates of PfPK7 as well as three phosphorylation motifs enriched among Plasmodium proteins that are down regulated in the absence of PfPK7. We were able to verify the ability of PfPK7 to directly utilize three of the putative substrates identified in the kinase negative phosphoproteomic analysis. However, PfPK7 did not utilize all of the potential substrates identified in this study, suggesting the role of PfPK7 as an upstream regulatory kinase. Our study provides the foundation for the use of quantitative phosphoproteomic analysis of kinase negative mutants in order to delineate their function and role in parasite signaling networks. This work represents a substantial advancement toward defining kinase-substrate interactions by unveiling the preliminary role of PfPK7 in Plasmodium signaling networks. Further studies will determine whether PfPK7 acts as an upstream regulatory kinase.
5.3 Elucidating the Function of Plasmodium falciparum Protein Kinases Through Substrate/Interactor Characterization

The intraerythrocytic development of the malaria parasite diverges from the eukaryotic cell cycle paradigm that has emerged largely from studies in yeast cells. In the typical eukaryotic cell cycle, chromosomes replicate only once and the transition of one cell cycle stage to the next is associated with the oscillation of CDK activity. Likewise, it is expected that proteins belonging to the CDK-related subfamily will be key regulators in the unique Plasmodium intraerythrocytic developmental cycle. Six CDK-related protein kinases have been identified in P. falciparum [19, 21]. To gain an insight into physiological functions of Plasmodium CDK-related kinases it is necessary to identify their cellular substrates and interactors. However, defining kinase-substrate pairs is one of the most challenging aspects of kinomics.

Our research made significant progress in elucidating the function of the atypical CDK-like kinase PfPK6. In order to identify putative CDK substrates, our phosphoproteomic dataset was analyzed for phosphorylation events occurring at the stringent CDK phosphorylation motif. Furthermore, PTMScan® Direct was used to identify additional CDK substrates. Following identification of potential CDK substrates, we showed the ability of PfPK6 to directly phosphorylate two of these novel substrates. We also uncovered the preliminary interaction network of PfPK6 by detecting the direct interacting partners of PfPK6 through co-immunoprecipitation of HA-tagged PfPK6. Our findings with PfPK6 are in agreement with the CDK-like nature of PfPK6 and also established a hitherto previously undefined interactome of PfPK6, while pointing toward PfPK6 having a role in nuclear function.

In conclusion, the data presented in this dissertation 1) defined the constitutive and regulated expression of the proteome during the intraerythrocytic developmental cycle, 2) demonstrated that
fluctuation and reversible phosphorylation is important for the regulation of *P. falciparum*’s unique cell cycle, 3) provided the foundation for quantitative phosphoproteomic analysis of kinase negative mutants to understand their function, 4) generated a preliminary network of interactions for PfPK6, and 5) set the stage for defining kinase-substrate pairs that are key in various signaling networks in the parasite and from this, perhaps the identification of novel therapeutic targets against malaria parasites.
LIST OF REFERENCES


28. Doerig C, Billker O, Pratt D, Endicott J: Protein kinases as targets for antimalarial intervention: Kinomics, structure-based design,


46. Xue L, Tao WA: Current technologies to identify protein kinase substrates in high throughput. *Front Biol (Beijing)* 2013, **8**:216-227.


60. Beynon RJ, Doherty MK, Pratt JM, Gaskell SJ: **Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides.** *Nat Methods* 2005, **2**:587-589.


76. Treeck M, Sanders JL, Elias JE, Boothroyd JC: The phosphoproteomes of *Plasmodium falciparum* and *Toxoplasma gondii* reveal unusual adaptations within and beyond the parasites' boundaries. *Cell Host Microbe* 2011, 10:410-419.

102. Neuwald AF: Rapid detection, classification and accurate alignment of up to a million or more related protein sequences. *Bioinformatics* 2009, **25**:1869-1875.


113. Peixoto L, Chen F, Harb OS, Davis PH, Beiting DP, Brownback CS, Ouloguem D, Roos DS: Integrative genomic approaches highlight a family


