Winter 12-15-2016

Discovery and characterization of a novel class of metabolic regulators in the malaria parasite Plasmodium falciparum

Ann Marie Guggisberg
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds
Part of the Biology Commons

Recommended Citation

This Dissertation is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.
Discovery and characterization of a novel class of metabolic regulators in the malaria parasite

*Plasmodium falciparum*

by

Ann M. Guggisberg

A dissertation presented to the
Graduate School of Arts & Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

December 2016
St. Louis, Missouri
© 2016, Ann Marie Guggisberg
TABLE OF CONTENTS

List of Figures ........................................................................................................ iv
List of Tables ........................................................................................................... vii
List of Abbreviations .............................................................................................. viii
Acknowledgments .................................................................................................... xiii
Abstract .................................................................................................................. xvi

Chapter 1: Introduction .......................................................................................... 1
   1.1 The malaria parasite, *Plasmodium falciparum* ........................................... 3
   1.2 Isoprenoid synthesis in the apicoplast ....................................................... 4
   1.3 The MEP pathway inhibitor fosmidomycin ............................................. 6
   1.4 Functions of isoprenoids in *P. falciparum* ............................................. 9
   1.5 Glycolysis and phosphofructokinase in the malaria parasite ............... 20
   1.6 The haloacid dehalogenase-like hydrolase (HAD) protein family .......... 23
   1.7 Aim and scope of the dissertation ............................................................ 25
   1.8 Figures ......................................................................................................... 27
   1.9 References ................................................................................................. 31

Chapter 2: The sugar phosphatase PfHAD1 is a metabolic regulator in malaria parasites...... 47
   2.1 Abstract ...................................................................................................... 50
   2.2 Introduction ................................................................................................ 50
   2.3 Methods .................................................................................................... 53
   2.4 Results ...................................................................................................... 63
   2.5 Discussion .................................................................................................. 71
   2.6 Figures ...................................................................................................... 75
   2.7 Tables ........................................................................................................ 92
   2.8 References ............................................................................................... 96

Chapter 3: Whole genome sequencing to evaluate the resistance landscape following antimalarial treatment failure with fosmidomycin-clindamycin ......................................................... 101
   3.1 Abstract .................................................................................................... 104
   3.2 Introduction ................................................................................................ 104
   3.3 Methods .................................................................................................... 106
   3.4 Results ...................................................................................................... 109
   3.5 Discussion .................................................................................................. 117
   3.6 Figures ...................................................................................................... 120
   3.7 Tables ....................................................................................................... 123
   3.8 References ............................................................................................... 132
Appendix B: Additional insights on the cellular function of PfHAD1

B.1 Methods ....................................................................................................................... 250
B.2 PfHAD1 mutants do not have a growth defect .......................................................... 253
B.3 PfHAD1 localization via electron microscopy ............................................................ 254
B.4 PfHAD1 does not behave like a membrane protein ..................................................... 256
B.5 PfHAD1 can bind to membrane lipids ......................................................................... 257
B.6 PfHAD1 is unable to utilize inositol phosphates or PIPs as substrates ..................... 259
B.7 References .................................................................................................................... 260
LIST OF FIGURES

Chapter 1: Introduction
Figure 1: Life cycle of *Plasmodium falciparum* .................................................. 27
Figure 2: Synthesis of isoprenoid products in *P. falciparum* ........................................ 28
Figure 3: Enzymatic steps of glycolysis ................................................................. 30

Chapter 2: The sugar phosphatase PfHAD1 is a metabolic regulator in malaria parasites
Figure 1: The MEP pathway for isoprenoid synthesis ................................................. 75
Figure 2: Generation of FSM-resistant *P. falciparum* ........................................... 76
Figure 3: FSM^R^ variants map to the core and active site regions of PfHAD1 .............. 77
Figure 4: PfHAD1 FSM-resistance alleles result in loss of phosphatase activity ............ 79
Figure 5: Loss of PfHAD1 is required for FSM resistance ........................................... 80
Figure 6: PfHAD1 dephosphorylates sugar phosphates, including MEP pathway intermediates 81
Figure 7: PfHAD1 is expressed in blood-stage parasites and is localized to the parasite cytoplasm .................................................. 82
Figure 8: Increased levels of MEP pathway intermediates in FSM^R^ strains with the loss of PfHAD1 function ........................................................................ 83
Figure 9: Model of PfHAD1 function ........................................................................ 84
Supplementary Figure 1: FSM^R^ strains are also resistant to the FSM-related compound, FR-900098 ............................................................. 85
Supplementary Figure 2: Relative mRNA expression levels of *PfDXS* and *PfDXR* are unchanged in FSM^R^ strains .................................................. 87
Supplementary Figure 3: Full immunoblots demonstrating expression of PfHAD1-GFP in FSM^R^ strain AM1 .............................................. 88
Supplementary Figure 4: PfHAD1-GFP localization is similar to that of PfHAD1 (Figure 7) .... 89
Supplementary Figure 5: Stereo image of the electron density maps for a representative region of PfHAD1 ........................................................ 90
Supplementary Figure 6: Loss of PfHAD1 is necessary for increased DOXP levels in FSM^R^ strain AM1 ................................................................. 91
Chapter 3: Whole genome sequencing to evaluate the resistance landscape following antimalarial treatment failure with fosmidomycin-clindamycin

Figure 1: SWGA causes enrichment of *P. falciparum* DNA from mixed samples that can be detected by PCR ................................................. 120
Figure 2: Samples cluster by patient, indicating recrudescent infections ................. 121
Figure 3: Resistance landscape before and after recrudescence .................................................. 122

Chapter 4: Fosmidomycin resistance reveals coordinated metabolic control in malaria parasites

Figure 1: HAD2 is a purine monophosphatase with a unique substrate specificity profile ....... 166
Figure 2: HAD2 and HAD1 are expressed throughout the intraerythrocytic parasite lifecycle . 168
Figure 3: FSM resistance results in a fitness cost ................................................................. 169
Figure 4: Model of parasite populations and genetic changes that modulate FSM sensitivity... 170
Figure 5: *HAD2* and *PFK9* alleles alter FSM resistance, PFK activity, and metabolite levels in *P. falciparum* ................................................................. 171
Figure 6: HAD2 is localized to the cytoplasm ................................................................. 173
Supplementary Figure 1: FSM\textsuperscript{R} strain E2 does not have increased levels of *DXS* or *DXR* transcripts ................................................................. 174
Supplementary Figure 2: HAD1 expression is unchanged in FSM\textsuperscript{R} strain E2 ................. 175
Supplementary Figure 3: HAD2 is a homolog of the MEP pathway regulator HAD1 ............. 176
Supplementary Figure 4: *P. vivax* HAD2 is structurally similar to *P. falciparum* HAD1 ........ 177
Supplementary Figure 5: AMP levels are unchanged in E2 FSM\textsuperscript{R} parasites .................. 178
Supplementary Figure 6: Map of suppressor mutations in *PFK9* ........................................... 179
Supplementary Figure 7: Assay of PFK activity from *P. falciparum* lysate is linear and specific .............................................................................................................. 180
Supplementary Figure 8: Suppressor mutations in PFK9 result in reduced activity .......... 181
Supplementary Figure 9: Rescue of HAD2 confirms role as a negative regulator ............... 182
Supplementary Figure 10: Anti-HAD1 and anti-HAD2 polyclonal antisera are specific for their target HAD antigens ......................................................................... 183
Appendix A: Cap-domain closure enables diverse substrate recognition by the C2-type haloacid dehalogenase-like sugar phosphatase *Plasmodium falciparum* HAD1

Figure A1: Model of PfHAD1 regulation ................................................................. 230
Figure A2: PfHAD1 binds and orders substrates in a closed conformation ........... 231
Figure A3: Substrates bind to the active site of PfHAD1-D27A ............................ 232
Figure A4: Distinct interactions for six-carbon and three-carbon sugar phosphates form the basis for substrate binding and specificity ................................................................. 233
Figure A5: The cap domain contains a flexible substrate-recognition element that is ordered upon substrate binding and cap closure................................................................. 235
Figure A6: Sequence alignment of substrate-binding cap-domain residues in PfHAD1 and PfHAD1 homologs from other organisms ................................................................. 236
Figure A7: Model for catalysis in PfHAD1 ................................................................ 237
Supplementary Figure A1: Stereo image of the electron density maps for representative regions of the substrate-bound PfHAD1-D27A structures ................................................................. 238
Supplementary Figure A2: PfHAD1-D27A is inactive ............................................. 239
Supplementary Figure A3: The cap domain contains a flexible substrate recognition element that is ordered upon substrate binding and cap closure................................................................. 240
Supplementary Figure A4: Sequence alignment of PfHAD1 homologs from other organisms . 241

Appendix B: Additional insights on the cellular function of PfHAD1

Figure B1: FSM\(^R\) *P. falciparum* with *PfHAD1* mutations do not have a growth defect........ 253
Figure B2: Localization of PfHAD1 by immunolabeling and electron microscopy........ 254
Figure B3: PfHAD1 does not behave like a membrane-associated protein .................. 256
Figure B4: PfHAD1 can bind membrane lipids ......................................................... 257
Figure B5: PfHAD1 is unable to utilize inositol phosphates or PIPs as substrates ....... 259
LIST OF TABLES

Chapter 2: The sugar phosphatase PfHAD1 is a metabolic regulator in malaria parasites
Table 1: Data collection and refinement statistics ................................................................. 92
Supplementary Table 1: FSM IC_{50}s, PfHAD1 alleles, and NCBI database accessions for FSM^{R} strains ........................................................................................................................................ 93
Supplementary Table 2: Kinetic parameters for PfHAD1 with the top three tested substrates and gly3P ........................................................................................................................................ 94
Supplementary Table 3: FSM^{R} strains possessed increased levels of MEP pathway metabolites ........................................................................................................................................ 95

Chapter 3: Whole genome sequencing to evaluate the resistance landscape following antimalarial treatment failure with fosmidomycin-clindamycin
Table 1: Gene ontology (GO) analysis of genes containing SNPs unique to recrudescent samples ........................................................................................................................................ 123
Supplementary Table 1: PCR primers used in this study ...................................................... 124
Supplementary Table 2: Primer sets used for SWGA reactions .......................................... 125
Supplementary Table 3: Whole genome sequencing statistics for samples sequenced in this study ........................................................................................................................................ 126
Supplementary Table 4: SNPs identified in known resistance loci .................................... 127
Supplementary Table 5: Unique, non-synonymous SNPs shared in \( \geq 50\% \) of recrudescent (post-treatment) samples ........................................................................................................................................ 128

Chapter 4: Fosmidomycin resistance reveals coordinated metabolic control in malaria parasites
Supplementary Table 1: Kinetic parameters for HAD2 ........................................................ 184
Supplementary Table 2: Genotypes of strains used to identify \( HAD2 \) and \( PFK9 \) mutations influencing FSM sensitivity ........................................................................................................................................ 185

Appendix A: Cap-domain closure enables diverse substrate recognition by the C2-type haloacid dehalogenase-like sugar phosphatase \( Plasmodium falciparum \) HAD1
Table A1: Data collection and refinement statistics .............................................................. 243
LIST OF ABBREVIATIONS

ACP, acyl carrier protein
ACT, artemisinin combination therapy
ALD, aldolase
AMP, adenosine 5’-monophosphate
Ap, apicoplast
BisP, bisphosphate
BsPFK, *Bacillus stearothermophilus* phosphofructokinase
CDP, cytidine 5’-diphosphate
CDP-ME, 4-diphosphocytidyl-2-C-methylerythritol
CDP-MEP, 4-diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate
CLN, clindamycin
CMP, cytidine 5’-monophosphate
CTP, cytidine 5’-triphosphate
dCMP, 2’-deoxycytidine 5’-monophosphate
dGMP, 2’-deoxyguanosine 5’-monophosphate
DHAP, dihydroxyacetone phosphate
DMAPP, dimethylallyl pyrophosphate
Dolichyl-PP, dolichyl pyrophosphate
DOXP, deoxyxylulose 5-phosphate
2drib5P, deoxyribose 5-phosphate
dUMP 2’-deoxyuridine 5’-monophosphate
DXR, DOXP reductoisomerase
DXS, DOXP synthase
Ery4P, erythrose 4-phosphate
FBP, fructose 1,6-bisphosphate
FPP, farnesyl pyrophosphate
FPPS, farnesyl pyrophosphate synthase
Fru1P, fructose 1-phosphate
Fru1,6bisP, fructose 1,6-bisphosphate
Fru6P, fructose 6-phosphate
FSM, fosmidomycin
FSM^R, fosmidomycin-resistant
FSM^S, fosmidomycin-sensitive
FTase, farnesyl transferase
FV, food vacuole
Gal1P, galactose 1-phosphate
GDH, glycerol 3-phosphate dehydrogenase
GDP, guanosine 5’-diphosphate
GPP, geranyl pyrophosphate
GGPP, geranylgeranyl pyrophosphate
GGTase, geranylgeranyl transferase
Glc2P, glycerol 2-phosphate
Gln1P, glucosamine 1-phosphate
Glu1P, glucose 1-phosphate
Glu6P, glucose 6-phosphate
GMP, guanosine 5’-monophosphate
GTAC, Genome Technology Access Center
HAD, haloacid dehalogenase-like hydrolase
hDHFR, human dihydrofolate reductase
HMB-PP, (E)-4-hydroxy-3-methyl-but-2-enyl pyrophosphate
Hsp70, heat shock protein 70
Hsp110, heat shock protein 110
I(1,4,5)P3, inositol (1,4,5)-triphosphate
I(3)P, inositol 3-phosphate
I(3,5)P2, inositol (3,5)-diphosphate
I(4)P, inositol 4-phosphate
I(5)P, inositol 5-phosphate
IMP, inosine 5’-monophosphate
IPP, isopentenyl pyrophosphate
IPPI, isopentenyl pyrophosphate isomerase
IspD, CDP-ME synthase
IspE, CDP-ME kinase
IspF, MEcPP synthase
IspG, HMB-PP synthase
IspH, HMB-PP reductase
K13, Kelch13
LC, lysophosphocholine
LPA, lysophosphatidic acid
Man1P, mannose 1-phosphate
Man6P, mannose 6-phosphate
MEcPP, 2-C-methyl-D-erythritol 2,4-cyclopyrophosphate
MEP, methylerthritol phosphate
MOI, multiplicity of infection
MPBQ, 2-methyl-6-phytyl-1,4-benzoquinol
MR4, Malaria Research Reference and Reagent Resource
MRM, multiple-reaction monitoring
myo-11P, myo-inositol 2-phosphate
N, nucleus
NER, nerolidol
NOR, norflurazon
OPPS, octaprenyl pyrophosphate synthase
PA, phosphatidic acid
Par, parental
PC, phosphatidylcholine
PD, phytoene desaturase
PE, phosphatidylethanolamine
PEP, phosphoenolpyruvate
2-PGA, 2-phosphoglyceric acid
3-PGA, 3-phosphoglyceric acid
PI, phosphatidylinositol
PIP, phosphatidylinositol phosphate
PI(3)P, phosphatidylinositol 3-phosphate
PI(4)P, phosphatidylinositol 4-phosphate
PI(5)P, phosphatidylinositol 5-phosphate
PI(3,4)P₂, phosphatidylinositol 3,4-bisphosphate
PI(3,5)P₂, phosphatidylinositol 3,5-bisphosphate
PI(4,5)P₂, phosphatidylinositol 4,5-bisphosphate
PI(3,4,5)P₃, phosphatidylinositol 3,4,5-trisphosphate
PK, pyruvate kinase

pNPP, para-nitrophenyl phosphate

PPR, polyprenol reductase

PS (Appendix B), phosphatidylserine

PS (Chapter 1), phytoene synthase

PTI, prenyltransferase inhibitor

Pyr5P, pyridoxal 5’-monophosphate

RBC, red blood cell

Rib5P, ribose 5-phosphate

Ribu5P, ribulose 5-phosphate

Sedo7P, sedoheptulose 7-phosphate

Sorb6P, sorbitol 6-phosphate

SWGA, selective whole genome amplification

S1P, sphingosine 1-phosphate

TDP, thymidine 5’-diphosphate

TIPT, tRNA isopentenyltransferase

TMP, thymidine 5’-monophosphate

TMT, tRNA methylthiolase

TPI, triose phosphate isomerase

Tre6P, trehalose 6-phosphate

UA, usnic acid

UDP, uridine 5’-diphosphate

UMP, uridine 5’-monophosphate

VCF, variant call format

XMP, xanthosine 5’-monophosphate
ACKNOWLEDGMENTS

I have been incredibly fortunate to find my journey as a student and scientist rich with supporters, advocates, and mentors. My thesis work, and in particular the process of writing my dissertation, has provided me an opportunity to reflect on these contributions.

First and foremost, I am incredibly grateful to my thesis mentor, Audrey. Audrey has impressed on me her unending enthusiasm for discovery and the thoughtfulness with which she communicates her science. She has been an unwavering source of confidence in the face of the self-doubt that plagues a Ph.D. student. It is from Audrey that I have learned to celebrate success and to learn from, not fear, failure.

I am grateful to my thesis committee for their time, guidance, support, and valuable feedback. I was so fortunate to have a group of talented, engaged, and thoughtful scientists guiding me through my thesis work.

The Odom lab was an incredibly engaged, supportive, and entertaining place to learn and work. Thank you for countless brainstorming sessions, helpful feedback, and making lab feel like home. I am especially grateful to my bay-mate and friend Rachel Edwards for her mentorship and many great conversations. Thank you to Megan Kelly and Andreas Mitchell for generating resistant parasites for me to study! I am also thankful for Aakash Gandhi, an undergraduate student, for his enthusiasm and hard work on this project. Thank you to my other bay-mate, Andrew Jezewski, for many thoughtful discussions and his continued work on metabolic regulation in the parasite. Many thanks to the Odom lab manager, Dana Hodge, for her excellent support and teaching me everything I needed to know about parasites!
Many thanks to the Department of Pediatrics for excellent administrative and technical support.

Thank you to the many collaborators that made this work possible. Special thanks to Jooyoung Park and Niraj Tolia for their collaboration on the HAD1 project. Many thanks to the Goldberg lab for their assistance and support.

I am grateful to the Molecular Genetics and Genomics graduate program. Thank you to the MGG program coordinators, James Skeath and Tim Schedl, and program coordinator, Melanie Relich, for your assistance and support.

I am grateful for the three years of funding support in the form of a Monsanto Excellence Fund graduate fellowship. It was through this fellowship that I was able to intern at Monsanto. I am grateful to Barry Goldman for facilitating the fellowship and my internship and for his continued personal mentorship. I am grateful to Jim Baum and the Insect Control research group for taking me in and providing me with the opportunity to experience scientific research in a new context.

I deeply appreciate the many friends I have made during my time as a graduate student. They were there to study, troubleshoot, and commiserate. We have celebrated each other during our transformation into independent scientists. In particular, I am grateful for my best friend and fellow MGG student, Breanne Harty. I know I could not have done this without you.

I am incredibly fortunate to have such a supportive family. My parents, Ron and Joanne, encouraged me to become an enthusiastic learner, to work hard, and to always do my best. As the youngest of four, I am also thankful for my siblings, who have been wonderful role models and supporters in more ways than they know. Thank you to my in-laws, Stephen and Janet, who have always cheered me on and asked “what did you discover this week?” Finally, I could not have
done with this without my husband, Philip Ruzycki. He continues to inspire me to be my best self and is a never-ending source of love and support.

Dedicated to my husband and fellow scientist, Philip

How special it has been to learn alongside you.
ABSTRACT OF THE DISSERTATION

Discovery and characterization of a novel class of metabolic regulators in the malaria parasite

*Plasmodium falciparum*

Ann M. Guggisberg

Doctor of Philosophy in Biology and Biomedical Sciences

Molecular Genetics and Genomics

Washington University in St. Louis, 2016

Dr. Audrey R. Odom John, Chairperson

The malaria parasite, *Plasmodium falciparum*, infects hundreds of millions of people per year and causes hundreds of thousands of deaths. Within the host red blood cell, the parasite relies on glycolysis for energy and synthesis of essential biomolecules. One such anabolic fate of glucose is the synthesis of isoprenoids, a broad and essential class of compounds that participate in a variety of cellular functions. In the face of ever-evolving drug resistance, new inhibitors and better understanding of parasite metabolism are required. The antibiotic fosmidomycin (FSM) targets the methylerthritol phosphate pathway for isoprenoid synthesis and is a well-validated inhibitor of *P. falciparum* growth. A forward selection for FSM resistance generated a number of parasite strains with increased drug tolerance. We identify mutations in two members of the haloacid dehalogenase-like hydrolase (HAD) superfamily, *PfHAD1* and *PfHAD2*, as causal for resistance. Enzymatic characterization and metabolic profiling reveal that these mutations are deleterious and confirm the role of *PfHAD1* and *PfHAD2* as novel negative regulators of glucose and isoprenoid metabolism. Despite their homology and shared role in FSM resistance, *PfHAD1*, a sugar phosphatase, and *PfHAD2*, a purine nucleotidase, appear to mediate FSM resistance via
distinct enzymatic mechanisms. To further understand the role of PfHADs as metabolic regulators, we harness a growth defect in FSM-resistant PfHAD2 mutants to select for suppressors of FSM resistance. We identify suppressor mutations in the key glycolytic enzyme phosphofructokinase (PfPFK9) and describe the effect of these mutations on enzyme function and parasite metabolism.

Given its safety in humans and its specificity as a MEP pathway inhibitor, FSM is a strong candidate for clinical development and is currently being evaluated in clinical trials as part of an antimalarial combination therapy. Unfortunately, previous studies have observed high rates of recrudescence following FSM treatment when paired with the antibiotic clindamycin (CLN). To understand whether any genetic changes correlate with recrudescence, we performed whole genome sequencing on patient parasite populations before and after recrudescence. We demonstrate the use of a selective amplification method to amplify and sequence parasite genomes from blood spots. Our genotyping does not reveal any genetic changes responsible for recrudescence, but rather support the hypothesis that FSM-CLN treatment failure is due to formulation or partner drug selection. We encourage further development of FSM and other MEP pathway inhibitors as antimalarial therapies.
Chapter 1

Introduction
PREFACE

With the exception of sections 1.3.2, 1.5, 1.6, and 1.7, this chapter is reproduced and adapted from its original publication [Guggisberg AM, Amthor RE, Odom AR. (2014). Isoprenoid synthesis in *Plasmodium falciparum*. *Eukaryotic Cell*, 13(11):1348-59]. Reproduction of this article is permitted as part of the author reuse guidelines of American Society for Microbiology.

We thank Rachel L. Edwards and Wandy Beatty (Molecular Microbiology Imaging Facility, Washington University) for providing electron micrographs. We thank Aakash Y. Gandhi and Chad Schaber for critical reading of the manuscript.

This work was supported by grant MD-LI-2011-171 from the Children's Discovery Institute of Washington University and St. Louis Children's Hospital, grant R01AI103280 from the NIH/NIAID, a Basil O'Connor Starter Scholar Research Award from the March of Dimes, and a Clinical Scientist Development award from the Doris Duke Charitable Foundation. A.M.G. is supported by training grant 5T32GM007067 from the NIGMS and a Monsanto Excellence Fund Graduate Fellowship.
1.1 THE MALARIA PARASITE, *PLASMODIUM FALCIPARUM*

Severe malaria remains a threat to human health worldwide, with over 250 million cases per year. Malaria is a leading cause of death in children, with almost one million deaths each year [1,2]. Despite ongoing and intensive control efforts, malaria remains endemic on five continents. Widespread resistance to former first-line agents, most notably chloroquine, has severely limited malaria control efforts [2]. Currently, the recommended standard of care for malaria infection is combination therapy using artemisinin-based therapeutics. However, decreased sensitivity to artemisinin has been recognized in the field, particularly in Southeast Asia. The spread of artemisinin resistance threatens the progress that has been made in control of malaria, particularly in sub-Saharan Africa [3–5]. New antimalarial agents, particularly agents with novel mechanisms of action, are urgently needed.

Malaria is caused by infection with protozoan parasites in the genus Plasmodium. Most cases of life-threatening malaria are attributable to infection with a single species, *Plasmodium falciparum*, although *P. vivax* and *P. knowlesi* have also been associated with severe disease [6–9]. Plasmodium infection is transmitted through the bite of anopheline mosquitoes (Figure 1 depicts their life cycle). Expelled from mosquito salivary glands, malaria sporozoites first traffic to the liver, where 10 to 100,000 daughter parasites are generated from a single invading cell. Upon egress from the liver, the parasite enters the host bloodstream. There, the malaria parasite begins an asexual cycle of growth and development within erythrocytes. This intraerythrocytic cycle leads to the signs and symptoms associated with malaria infection, including fever, anemia, and multiorgan dysfunction due to vascular adherence of parasitized red blood cells. New
antimalarials must therefore target this pathogenic stage of parasite development. A small proportion of asexual-stage parasites leave the asexual cycle and commit to the production of sexual forms, known as gametocytes. Upon a new blood meal, gametocytes return to the mosquito midgut, where they complete sexual development and begin the life cycle anew.

1.2 ISOPRENOID SYNTHESIS IN THE APICOPLAST

One cellular peculiarity of Plasmodium species, as well as other apicomplexan parasites, such as Toxoplasma and Babesia species, is the presence of an unusual plastid organelle, the apicoplast (Figure 2A and B). The apicoplast is surrounded by four membranes, suggesting an ancient secondary endosymbiotic event between a protozoan parasite ancestor and red algae, similar to that of the chloroplast [10–12]. While the apicoplast was previously believed to be of green algal origin, the recent discovery and genome sequencing of the alveolate Chromera velia has revealed C. velia as an evolutionary link between apicomplexans and their red algal ancestors [11,12]. C. velia can potentially serve as a useful tool to study the evolution of plastid pathways in apicomplexan parasites. While photosynthetic capabilities have been lost over time, the malaria parasite has retained some plantlike metabolic pathways that hold particular value as targets for antimalarial drug development, since these pathogen-specific processes are not present in humans.

Key among apicoplast metabolic pathways is that of isoprenoid biosynthesis. Isoprenoids comprise a very large and diverse group of biomolecules derived from the sequential assembly of two 5-carbon isomers, isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate
(DMAPP). Chains of isoprene units are subsequently modified through cyclizations, oxidations, reductions, and additions to generate the array of over 25,000 isoprenoids found in nature [13]. In humans (as well as fungi, archeabacteria, cytoplasm of plants, and other metazoans), the isoprenoid building blocks IPP and DMAPP are produced through a mevalonate-dependent pathway from acetyl-coenzyme A (CoA). The rate-limiting step in the mevalonate pathway is the conversion of 3-hydroxy-3-methyl-glutaryl (HMG)-CoA to mevalonic acid by the enzyme HMG-CoA reductase; this enzyme is the target for the widely used statin class of cholesterol-lowering drugs [14].

In Plasmodium species, IPP and DMAPP are produced via an alternative biosynthetic route that does not utilize mevalonate [15,16]. This pathway, also called the MEP (2-C-methyl-D-erythritol 4-phosphate) pathway or DOXP (1-deoxy-D-xylulose 5-phosphate) pathway, converts glyceraldehyde 3-phosphate and pyruvate to IPP and DMAPP through seven enzymatic steps (Figure 2C). At least two enzymes of this pathway catalyze rate-limiting steps in IPP production: DOXP synthase (DXS) (EC 2.2.1.7, PlasmoDB identifier [ID] PF3D7_1337200) converts glyceraldehyde 3-phosphate and pyruvate to DOXP, and DOXP reductoisomerase (DXR) (EC 1.1.1.267, PF3D7_1467300) converts DOXP to MEP [17,18]. The mechanism by which IPP and DMAPP are exported from the apicoplast for use in the cytoplasm remains unknown. Given the structural diversity of isoprenoids, it is not surprising that these molecules serve diverse cellular functions. Plants in particular elaborate an incredible range of specialized isoprenoid end products, including pharmacologically active compounds like paclitaxel (originally named taxol) [19] and artemisinin [20], as well as terpenes, volatile isoprenoids that confer the characteristic odors, flavors, and colors of plants. Other roles of isoprenoids include
regulation of cell growth and energy production, intracellular signaling, and membrane structural support [15,21,22].

1.3 THE MEP PATHWAY INHIBITOR FOSMIDOMYCIN

1.3.1 Fosmidomycin as a chemical tool

An important reagent in the study of the MEP pathway has been the selective MEP pathway inhibitor, fosmidomycin. Fosmidomycin is a small, three-carbon phosphonate compound that was first identified from *Streptomyces lavendulae* by its antibacterial properties [23]. Subsequent in vitro studies revealed that fosmidomycin competitively inhibits DXR, the first dedicated enzyme of the MEP pathway [24–26] (Figure 2C). The charged nature of fosmidomycin means that this compound is typically excluded from cells unless actively imported, which has limited its utility against many organisms, including the apicomplexan *Toxoplasma gondii* [27] and the agent of tuberculosis, *Mycobacterium tuberculosis* [28]. Intraerythrocytic malaria parasites elaborately remodel the host red blood cell, significantly increasing the cellular uptake of many nutrients [29–31]. These so-called new permeability pathways likely facilitate the uptake of fosmidomycin, as fosmidomycin is excluded from uninfected red blood cells but inhibits the growth of Plasmodium and a related, tick-borne intraerythrocytic apicomplexan pathogen, *Babesia divergens* [32]. It remains unclear what cellular machinery is required for fosmidomycin uptake into *P. falciparum* cells.

Fosmidomycin is well validated as a specific inhibitor of DXR. Analysis of MEP pathway intermediates in bacteria and *P. falciparum* has established that fosmidomycin reduces the
intracellular levels of downstream MEP pathway metabolites and isoprenoid products [33–35]. In addition, the growth inhibitory effects of fosmidomycin are chemically rescued in bacteria and malaria parasites through supplementation of the medium with IPP or unphosphorylated isoprenols (farnesol and geranylgeraniol). The 50% inhibitory concentration (IC₅₀) for fosmidomycin increases 10-fold when the medium is supplemented with farnesol or geranylgeraniol [33,36]. Supplementation of the medium with geranylgeraniol also rescues protein mislocalization and the organelle disruption effects of fosmidomycin treatment [37]. Treatment with high concentrations of fosmidomycin is not completely rescued by prenyl alcohol supplementation, perhaps due to the toxicity of these compounds at high concentrations [38].

In asexual parasites, the MEP pathway may be the only essential function of the apicoplast organelle in which it resides. Treating parasites with inhibitors of apicoplast replication forces *P. falciparum* to lose its apicoplast genome and structure. These parasites nonetheless survive when supplemented with exogenous IPP [36].

Small-molecule inhibitors that target apicoplast replication often result in a delayed-death phenotype in *P. falciparum*, in which drug-treated parasites complete the first cell cycle after treatment and arrest in the second [39]. In contrast, fosmidomycin treatment inhibits intraerythrocytic growth of *P. falciparum* during the first cell cycle. Interestingly, fosmidomycin-treated parasites develop within the red blood cell, begin hemoglobin digestion, and initiate DNA replication prior to cell cycle arrest as multinucleate schizonts [37]. The requirement of the new permeability pathways mentioned above for fosmidomycin import into *P. falciparum* cells may
explain this delayed action, as these pathways are not fully developed until the trophozoite stage [29,30]. Liver-stage parasites are also sensitive to fosmidomycin. Treatment of liver-stage *Plasmodium berghei* inhibits the development of the apicoplast and reduces the number of meroosomes, the result of liver-stage replication. Thus, the MEP pathway appears to be required for optimal growth in hepatocytes [27]. Little is known about isoprenoid synthesis in the gametocyte and mosquito stages of the parasite life cycle, although proteomics studies have identified MEP pathway enzymes expressed in late gametocytes [40].

### 1.3.2 Fosmidomycin as a clinical antimalarial

Fosmidomycin was originally clinically developed for its antibacterial properties [41]. Its use as an antimalarial was first demonstrated in 1999 [42]. It is currently being evaluated as a partner agent part of an antimalarial combination therapy. Given its specificity for the MEP pathway, fosmidomycin is well-tolerated in humans [43,44]. Studies have evaluated the synergy of FSM with other antimalarials, including chloroquine, artemisinin, atovaquone, proguanil, and clindamycin [45]. This study identified synergy between FSM and the antibiotic clindamycin (CLN) [45,46]. Like FSM, clindamycin also has a short half life (1-3 h) [41,47] and thus was an unusual choice of partner drug for FSM. Nonetheless, several clinical trials have evaluated the efficacy of FSM-CLN in various populations (clinical trial identifiers NCT02198807, NCT01361269, NCT01002183, NCT00214643, and NCT00217451) [43,48]. The combination initially showed promise, with high cure rates [46,49], although recrudescence has been observed in recent pediatric studies [50]. Currently, a Phase II clinical trial is underway to test the efficacy of fosmidomycin paired with a new partner, the antimalarial piperaquine (NCT02198807).
1.4 FUNCTIONS OF ISOPRENOIDS IN P. FALCIPARUM

1.4.1 5-carbon isoprenoids: IPP and DMAPP

The most proximally produced compounds of the MEP pathway are the end products and isoprenoid building blocks, isopentenyl pyrophosphate (IPP) and its isomer dimethylallyl pyrophosphate (DMAPP). In bacteria, DMAPP is used as a substrate for tRNA isopentenylation [51]. In this process, an isopentenyl group is added to an adenosine in the tRNA, targeting the tRNA to the ribosome and improving translation fidelity. Evidence suggests that P. falciparum produces isopentenylated tRNAs, as its apicoplast genome encodes four tRNAs that represent probable candidates for isopentenylation [52]. P. falciparum encodes a homologue of the Escherichia coli tRNA isopentenyltransferase MiaA (EC 2.5.1.75, PF3D7_1207600) [52]. P. falciparum also possesses a homolog of the isopentenyl-adenosine tRNA methylthiolase MiaB (EC 2.8.4.-, PF3D7_0622200), an enzyme whose bacterial homologs participate in downstream tRNA isopentenylation steps [52,53].

As previously described, supplementation of medium with farnesol or geranylgeraniol rescues fosmidomycin treatment of malaria parasites. These isoprenols are presumed to be phosphorylated intracellularly by nonspecific kinases to generate their cognate diphosphates, farnesyl pyrophosphate (FPP) and geranylgeranyl pyrophosphate (GGPP). Fosmidomycin-treated parasites supplemented with farnesol (15-carbon) or geranylgeraniol (20-carbon) therefore do not have a known source of 5- or 10-carbon isoprenoids but are still capable of intraerythrocytic growth [33]. These studies cannot rule out a role for tRNA isopentenylation in maintaining apicoplast function, but they do suggest there may not be additional roles for tRNA
isopentenylation outside the maintenance of isoprenoid biosynthesis within the apicoplast.

1.4.2 Monoterpenes (10-carbon isoprenoids)

Condensation of two isoprene units produces the 10-carbon isoprenoid geranyl pyrophosphate (GPP). GPP is utilized by monoterpane synthases and monoterpane cyclases to produce 10-carbon monoterpenes. Monoterpenes are the most abundant compounds found in plant essential oils. Terpene mixtures such as citronellal (citronella) and citral (lemon) are produced by plants to deter herbivores [54]. Some monoterpenes have been found to have antimicrobial properties; thymol, a monoterpane component of the essential oil in thyme, has been shown to decrease counts of Salmonella enterica serovar Typhimurium and Staphylococcus aureus [55]. To date, no studies have identified evidence of monoterpane synthesis in P. falciparum. Homology-based searches do not identify potential monoterpane synthases or monoterpane cyclases in the P. falciparum genome, although this class of enzymes is remarkably diverse [56].

1.4.3 Sesquiterpenes and diterpenes (15- and 20-carbon isoprenoids)

FPP and GGPP are also used to produce sesquiterpenes and diterpenes. In P. falciparum, the condensation of GPP and IPP (to produce FPP) and of FPP and IPP (to produce geranylgeranyl diphosphate [GGPP]) appears to be catalyzed by a bifunctional FPP-GGPP synthase (PF3D7_1128400) [57]. The crystal structure of the P. vivax enzyme has been solved [58,59]. This bifunctional enzyme is sensitive to bisphosphonates, such as zoledronate and risedronate [57]. These compounds bind to bone minerals and are traditionally used to inhibit bone resorption in the treatment of diseases like osteoporosis [60]. Metabolic-labeling studies using $[^{14}\text{C}]$mevalonate, $[^{14}\text{C}]$IPP, and $[^{14}\text{C}]$DMAPP validate FPP synthesis as the target of
bisphosphonate inhibition [61]. Bisphosphonates, as well as their analogs, have been shown to bind FPP-GGPP synthase in the active site of both the human and parasite enzymes [59]. Bisphosphonates compete with GPP for binding, and their efficacy is enhanced by IPP stabilization of the enzyme inhibitor complex [62,63]. Treatment with bisphosphonates inhibits parasite growth and decreases protein prenylation [58,59,64–66]. Partial rescue of parasite growth can be achieved by the addition of FPP or GGPP to the culture medium, indicating that bisphosphonates target FPP and GGPP synthesis in *P. falciparum* [64].

*P. falciparum* utilizes sesquiterpenes and diterpenes for protein prenylation. In protein prenylation, lipophilic farnesyl (15-carbon) and geranylgeranyl groups (20-carbon) are attached to C-terminal cysteines, which results in protein association with membranes. Prenylation is crucial for the function of a variety of membrane-bound enzymes, such as the Ras, Rho, and Rab families of small GTPases. Farnesyl transferase (EC 2.5.1.58, PF3D7_1242600 [α subunit] and PF3D7_1147500 [β subunit]) and geranylgeranyl transferase type I (EC 2.5.1.59, PF3D7_1242600 [α subunit] and PF3D7_0602500 [β subunit]) transfer FPP and GGPP moieties to the target protein via recognition of a C-terminal CaaX motif [67]. This motif is composed of a cysteine (C), two aliphatic amino acids (aa), and the C-terminal amino acid (X). These two prenyltransferases share an α subunit but have distinct β subunits [68]. Geranylgeranyl transferase type II (EC 2.5.1.60, Rab geranylgeranyltransferase, PF3D7_1442500 [α subunit] and PF3D7_1214300 [β subunit]) utilizes a different mechanism of substrate recognition, which requires a Rab escort protein (REP) [69]. The REP binds Rab proteins, facilitates their prenylation, and delivers them to their target membrane. REPs are part of the REP/GDI superfamily, which also includes GDP dissociation inhibitor (GDI) proteins. GDI proteins are
involved in cycling the Rab between membranes and the cytosol [67]. Two members of the REP/GDI superfamily are found in the *P. falciparum* genome (PF3D7_1242800 and PF3D7_1038100) [70]. Further studies are required to determine if these proteins function as REPs or GDIs.

Protein prenylation appears to be an essential use of isoprenoids in *P. falciparum*, as the parasite is sensitive to chemical inhibition of protein prenylation. Prenyltransferase inhibitors, major candidates for anticancer therapy, have shown potent antiplasmodial activity [71–73]. These include a number of peptidomimetics of the CaaX motif [71,74,75]. Additionally, certain monoterpenes have been shown to inhibit the growth of *P. falciparum* via inhibition of prenylation [38,76]. Limonene has been shown to inhibit the prenylation of 21 to 26 kDa proteins in mammalian cell culture and in vitro [77–79]. *P. falciparum* parasites treated with limonene are unable to progress from the ring to the trophozoite stage [76].

Feeding *P. falciparum* with labeled [³H]FPP and [³H]GPP identifies 21 to 24 kDa and 50 kDa prenylation target proteins that are differentially labeled by FPP and GPP [76]. Overall, it appears that the majority of prenylated proteins are small and preferentially geranylgeranylated in vivo, with the exception of a single 50 kDa protein [73]. Bioinformatic methods have been used to compile a list of predicted prenylation targets in *P. falciparum* [68]. These include a number of GTP-binding proteins, such as Rab2 and Rab11a.

A number of studies have identified specific prenylation targets in *P. falciparum*. The localization of the small GTPase Rab7 to endosomal vesicles was shown to be prenylation
dependent. These vesicles are predicted to participate in endosomal trafficking [80]. The SNARE protein Ykt6.1 of *P. falciparum* (PfYkt6.1) has been shown to be a farnesyltransferase substrate in vitro, and its localization is disrupted when lacking a CaaX motif, suggesting it is also a prenylation target in vivo [81]. Similarly, the *P. falciparum* tyrosine phosphatase PfPRL is a substrate for farnesylation in vitro [82].

Inhibition of isoprenoid synthesis by fosmidomycin produces prenylation phenotypes in *P. falciparum*. Similar to metabolic labeling, probing for prenylation using a prenylation-specific antibody identifies proteins of 21 to 24 kDa and 50 kDa [37,77]. Prenylation of these targets is reduced in fosmidomycin-treated parasites, confirming that prenyl groups are indeed products of the MEP pathway. The *P. falciparum* geranylgeranyltransferase substrates PfRab5a and PfRab5c mislocalize from hemoglobin-containing vesicles to the host cell membrane upon treatment with fosmidomycin, and this mislocalization correlates with changes to food vacuole morphology and integrity. Proper localization is restored by geranylgeraniol supplementation [37]. Geranylgeraniol supplementation also substantially rescues growth inhibition by fosmidomycin, suggesting that geranylgeranylation may be the only essential form of protein prenylation in *P. falciparum* [33,36].

The group of compounds collectively known as vitamin E (tocopherols and tocotrienols) function as antioxidants and membrane stabilizers [83–85]. Recently, a metabolic labeling study using [³H]FPP and [³H]GPP identified de novo vitamin E synthesis by *P. falciparum*. Parasite growth is sensitive to usnic acid, an inhibitor of vitamin E biosynthesis. Growth is partially rescued by α-tocopherol, indicating that vitamin E synthesis is essential in malaria parasites [86].
In plants, the homogentisic acid head group of vitamin E is synthesized via the shikimate pathway. This head group is then prenylated with phytol diphosphate (20-carbon) or geranylgeranyl diphosphate to generate 2-methyl-6-phytyl-1,4-benzoquinol, the first committed intermediate for the synthesis of tocopherols and tocotrienols. This prenylation is catalyzed by homogentisate prenyltransferases (EC 2.5.1.115 and EC 2.5.1.116) [87]. As no obvious homogentisate prenyltransferase homologs exist in the *P. falciparum* genome, further work will be required to understand the mechanism by which the parasite synthesizes vitamin E.

### 1.4.4 Sterols

Sterols are 30-carbon isoprenoids that are ubiquitous among eukaryotes and are utilized for a variety of cellular functions. In particular, cholesterol is essential for membrane architecture in eukaryotes, and its production is tightly regulated [88,89]. Cholesterol is also a precursor for signaling molecules, such as sex steroids and mineralocorticoids in mammals and brassinosteroids in plants. Squalene synthase (EC 2.5.1.21) commits the isoprenoid pathway to sterol biosynthesis by converting two molecules of FPP to squalene; squalene then serves as the backbone for subsequent modifications [90].

The animal host synthesizes cholesterol de novo and is also able to import it from dietary sources [91]. Radioactive labeling experiments show no evidence for cholesterol biosynthesis in *P. falciparum* [92,93]. Homology searches do not identify a squalene synthase in the *P. falciparum* genome. Instead, Plasmodium spp. appear to obtain cholesterol from the host cell. *P. knowlesi* was shown to import host-derived $^{14}\text{C}$-labeled cholesterol, and cellular uptake by the host cell
itself was also increased upon infection with malaria parasites [94–96]. *P. falciparum* import of cholesterol has been studied within hepatocytes, a site of high cholesterol synthesis and parasite replication early in infection. Inhibition of host cell isoprenoid synthesis decreases sterol levels in the liver-stage parasite [97]. While cholesterol is essential for the maintenance of parasite membrane stability [98], these studies suggest that cholesterol synthesis does not occur in *P. falciparum* and is therefore not an essential function of de novo isoprenoid synthesis by the parasite.

### 1.4.5 Carotenoids

Carotenoids are 40-carbon isoprenoids derived from the condensation of two GGPP molecules by phytoene synthase (EC 2.5.1.32, PF3D7_0202700). Carotenoids are synthesized by plants and algae, as well as some bacteria and fungi. In plants, algae, and photosynthetic bacteria, carotenoids like carotene, lycopene, xanthophyll, and lutein function in photosynthesis and protect against free radical damage [99–103]. In plants, carotenoid synthesis occurs in the chloroplast [104,105]. Fungi also utilize carotenoid pigments for protection against free radicals [106,107]. In animals, which cannot synthesize carotenoids, dietary carotenoids are used for the synthesis of vitamin A [99]. An exception is found in insects that have acquired carotenoid synthesis from fungi through lateral gene transfer [108,109]. Commercial synthesis of carotenoids is of interest for their use as nutraceuticals, dietary supplements, and pigments [110,111].

Carotenoids have recently been detected in the intraerythrocytic stages of *P. falciparum*; schizonts contain the highest concentrations, indicating that carotenoid synthesis begins in the
ring stage and builds during the schizont stage [112]. Geranyl pyrophosphate serves as a substrate for carotenoid synthesis by phytoene synthase (EC 2.5.1.32, PF3D7_0202700). Phytoene is then converted to carotenoid products by phytoene desaturase (EC 1.3.99.30, locus unknown). *P. falciparum* is sensitive to the small molecular herbicide norflurazon, which inhibits phytoene desaturase. Norflurazon treatment causes an accumulation of phytoene and a decrease in carotenoid content. Inhibition by norflurazon can be partially rescued with lycopene [112]. While carotenoids serve important functions in plants, algae, bacteria, and fungi, it is not yet known what physiological role they play in *Plasmodium*. As in other organisms, they may play a role in the cellular response to oxidative stress.

In plants, the phytohormone abscisic acid is also produced from carotenoid intermediates [113]. The *Plasmodium* relative and apicomplexan parasite, *Toxoplasma gondii*, has been shown to produce abscisic acid to control calcium signaling for processes like protein secretion and parasite egress. The abscisic acid response genes identified in *T. gondii* are conserved in *P. falciparum*, but it is not known whether *P. falciparum* also synthesizes this isoprenoid product. The route for abscisic acid synthesis from isoprenoid precursors in apicomplexan parasites remains unknown, as no clear biosynthetic route is readily identified bioinformatically [114].

### 1.4.6 Coenzyme Q (ubiquinone)

In most eukaryotes, mitochondria are the site of energy generation through oxidative phosphorylation. Within the mitochondrial matrix, the tricarboxylic acid cycle uses 2-carbon metabolites generated from the breakdown of glucose, amino acids, and fatty acids to produce high-energy electron carriers. In the inner mitochondrial membrane, the electron transport chain
uses high-energy electrons to harness energy in the form of ATP. In the mitochondria of asexual Plasmodium parasites, however, the electron transport chain is not a primary source of ATP and the parasite instead relies on glycolysis for most of its ATP production [115]. Indeed, little of the parasitic glucose supply is completely oxidized, and glucose is instead excreted as lactic acid [116,117]. Additionally, the parasites show relatively little oxygen consumption, consistent with minimal respiration [118]. However, ATP generation by the electron transport chain may be essential for parasite stages within the mosquito host, where extracellular glucose levels are lower and the parasite cannot rely solely on glycolysis for ATP production [119,120].

In asexual-stage parasites, the electron transport chain operates to provide a continuous supply of reduced coenzyme Q. Coenzyme Q, or ubiquinone, typically functions as an electron acceptor in the electron transport chain. It is maintained in asexual Plasmodium parasites as an electron acceptor for dihydroorotate dehydrogenase (DHODH) (EC 1.3.98.1, PF3D7_0603300), an enzyme required for pyrimidine synthesis. DHODH is essential for survival, as the parasite is incapable of pyrimidine salvage, and small molecules targeting DHODH have potent antimalarial activity [37,121–124]. Ubiquinone levels have been shown to peak at the beginning of schizogony and are sensitive to fumidomycin treatment [34].

Synthesis of coenzyme Q requires the addition of an isoprenyl side chain to a benzoquinone ring. The parasite possesses an octaprenyl pyrophosphate synthase (EC 2.5.1.90, PF3D7_0202700) that is capable of synthesizing these side chains. This multifunctional enzyme produces 40-carbon, 45-carbon, and 55-carbon isoprenoid products and has been shown to also have phytoene synthase activity [112,125]. The addition of these isoprenoids to 4-hydroxybenzoate is
performed by 4-hydroxybenzoate octaprenyltransferase (EC 2.5.1.39, PF3D7_0607500) [126,127]. Labeling studies in *P. falciparum* identify coenzyme Q isoforms coenzyme Q8 and Q9, which have 8 and 9 isoprene units (40-carbon and 45-carbon), respectively, in their side chains [128,129]. Incorporation of labeled FPP results in the detection of coenzyme Q8, and incorporation of labeled GGPP detects coenzyme Q9 [128]. In another study using labeled p-hydrobenzoic acid, coenzyme Q8 was found to be the dominant form of coenzyme Q [129].

Nerolidol, a sesquiterpene alcohol, was found to inhibit the synthesis of the isoprenyl side chain destined for coenzyme Q, likely because of its structural similarity to FPP. Treatment with nerolidol inhibits the intraerythrocytic development of *P. falciparum* [125,128].

### 1.4.7 Dolichols

Dolichols are long-chain hydrocarbon compounds made of various numbers of isoprene units. In the form of dolichyl phosphate or pyrophosphate, dolichols are essential for the transfer of sugars onto proteins, i.e., dolichylation, O-linked glycosylation, N-linked glycosylation, and the production of glycoposphatidylinositol (GPI) anchors, which are essential for successful infection [130].

Multiple studies have demonstrated the presence of dolichols and their intermediates in *P. falciparum*, specifically, those composed of 11 and 12 isoprene units (55 and 60 carbons). Labeling experiments demonstrate that these dolichols are formed from FPP and GPP, respectively [131,132]. As expected, dolichol synthesis is also sensitive to fosmidomycin treatment [35].
Dolichyl pyrophosphate is produced from IPP by a polyprenol reductase (EC 1.3.1.94) and from GGPP via a dehydrodolichol pyrophosphate intermediate. Dolichyl pyrophosphate is then converted to dolichyl phosphate by dolichyldiphosphatase (EC 3.6.1.43). Dolichyl phosphate is utilized for glycosylation and synthesis of GPI anchors. *P. falciparum* possesses homologs of both polyprenol reductase (PF3D7_1455900) and dolichyldiphosphatase (PF3D7_0805600).

Posttranslational addition of dolichols to proteins has been demonstrated in *P. falciparum*. Labeling using [3H]FPP and [3H]GGPP identified a dolichol with 11 isoprene units attached to 21 to 28 kDa protein(s). The target proteins and enzyme(s) responsible for dolichylation of proteins in *P. falciparum* remain unknown [132].

*P. falciparum* synthesizes GPI anchors for protein modification [133]. The parasite is sensitive to known inhibitors of GPI synthesis, such as the mannose analogue 2-deoxyglucose [134,135]. Studies have identified a number of parasite proteins as targets for GPI addition, including merozoite surface antigens, a serine protease, and a heat shock 70 protein [136,137]. Merozoite surface antigens appear to be the primary targets of GPI anchor addition and are of great interest as antigens for vaccine development [138]. The enzyme GPI1 (EC 2.4.1.198, PF3D7_0618900) transfers N-acetylglucosamine to phosphatidylinositol in GPI biosynthesis. The *P. falciparum* GPI1 homolog was shown to complement a yeast gpi1 mutant, confirming its function [139]. Dolichol phosphate mannose synthase (EC 2.4.1.83, PF3D7_1141600) catalyzes the addition of sugar moieties in both GPI anchor glycosylations and N-linked glycosylations. The *P. falciparum* enzyme has been shown to be of a novel clade distinct from animal or yeast synthases.
of the same type [140].

While GPI anchors appear to constitute most of the glycosylation in \textit{P. falciparum}, there is also evidence for O- and N-linked glycosylation of proteins [141–143]. The presence of N-linked glycosylation has long been debated [144,145]. However, the parasite is sensitive to tunicamycin, an inhibitor of N-linked glycosylation [143], and parasite proteins have been shown to be capable substrates of N-linked glycosylation when expressed in heterologous systems [146].

1.5 GLYCOLYSIS AND PHOSPHOFRUCTOKINASE IN MALARIA PARASITES

Isoprenoid synthesis via the MEP pathway is one of many of the anabolic fates of glucose and begins with the import of two glycolytic metabolites, phosphoenolpyruvate (PEP) and dihydroxyacetone phosphate (DHAP), into the apicoplast. Two apicoplast transporters, PfiTPT (PF3D7_0530200) and PfoTPT (PF3D7_0508300), have been found to be responsible for import of PEP and DHAP into the organelle [147,148] (Figure 2C). Within the apicoplast, PEP and DHAP are converted into the DXS substrates pyruvate and glyceraldehyde 3-phosphate by apicoplast-localized pyruvate kinase (PK, PF3D7_1037100) [149] and triosephosphate isomerase (TPI) [52], respectively.

It has been established that Plasmodium parasites are ravenous consumers of glucose. Infected erythrocytes import nearly 100-fold more glucose than uninfected erythrocytes [150–152]. The
parasite is able to import glucose from the host cell via at least one hexose transporter (PF3D7_0204700) [153]. As it is expected that host (erythrocyte) glucose is readily phosphorylated to glucose 6-phosphate (glu6P), the parasite may employ a phosphatase to convert glu6P to glucose for import [115,154].

In the intraerythrocytic asexual stage parasite forms, the parasite relies on anaerobic glycolysis as the primary source of ATP [155]. As previously mentioned, only a small proportion of the parasite glucose is oxidized [116,117] and the parasite shows relatively little oxygen consumption, consistent with a lack of aerobic respiration for ATP generation [118]. Most parasite glucose is converted to and excreted as lactate [119,156]. Labeling studies confirm that flux of glucose into the TCA cycle is low during the asexual stages, but increases in the gametocyte (sexual) stages of the parasite lifecycle [157].

The dependence of *P. falciparum* on glucose is underscored by the parasite’s lack of gluconeogenesis. Many organisms, including the related apicomplexan *T. gondii*, metabolize other carbon sources (acetate, amino acids, etc.) and convert them to glucose via gluconeogenesis [158]. Recent work has shown that *T. gondii* regulates constitutive expression of gluconeogenic enzymes to fine-tune its metabolism [158]. Notably, the *P. falciparum* genome does not appear to encode a fructose bisphosphatase (FBPase), an enzyme essential for gluconeogenesis [159]. However, early labeling studies indicate that Plasmodium parasites have a limited capacity to utilize metabolites such as pyruvate, glutamate, and lactate [160–162].

The parasite maintains and expresses all the canonical enzymes of glycolysis (Figure 3)
Imported glucose is converted to glucose 6-phosphate by hexokinase (EC 2.7.1.1, PF3D7_0624000). As glucose 6-phosphate is also shunted into the pentose phosphate pathway and fructose 6-phosphate is used for the synthesis of N-glycans and dolichols, the glycolytic enzyme phosphofructokinase (PFK) (EC 2.7.1.11, PF3D7_0915400) catalyzes the first committed enzymatic step of glycolysis [164].

Studies of glycolysis in other organisms have identified and established the enzymatic step catalyzed by PFK as a key regulatory step of glycolysis. ATP-utilizing PFKs (such as the human enzyme) are usually allosterically inhibited by ATP and citrate. This inhibition is reversed by the allosteric activators AMP, ADP, and fructose 2,6-bisphosphate [165]. PFKs are often comprised of alpha and beta domains, with the alpha domain usually considered regulatory and the beta domain catalytic [165–168].

PP\(_i\)-utilizing PFKs have been found in photosynthetic plants, some bacteria, and the parasitic amoeba *Entamoeba histolytica*, with many of these organisms possessing both a PP\(_i\) and ATP-PFK [169–172]. PP\(_i\)-PFK regulation appears to differ in some aspects when compared to ATP-PFKs. PP\(_i\)-PFKs are generally insensitive to regulation by ATP, ADP, and AMP. Plant PP\(_i\)-PFKs are sensitive to regulation by fructose 2,6-bisphosphate [173]. Some PP\(_i\)-PFKs have been shown to be regulated via differential oligomerization of their subunits [174,175]. The PFKs of only a few apicomplexans have been studied, although sequence alignments have shown that they are related to plant PP\(_i\)-PFKs [176]. PFK from the rodent parasite *Plasmodium berghei* is ATP-dependent [177]. However, *T. gondii* and Cryptosporidium possess PP\(_i\)-PFKs, neither of which are activated by fructose 2,6-bisphosphate [178,179]. *P. falciparum* possess two PFK-encoding
ORFs, designated by their chromosome locations as PfPFK9 (PF3D7_0915400) and PfPFK11 (PF3D7_1128300). It appears that PfPFK9 is the main PFK for asexual stage parasites, as PfPFK11 is expressed primarily in gametocytes [163] and has been speculated to be missing key residues for catalysis [176]. One study, which measured PFK activity from parasite cell lysate, found PfPFK9 to be ATP-utilizing, despite its overall homology to PP\textsubscript{i}-PFKs. PfPFK9 activity from lysate was sensitive to regulation by ADP, but not citrate, fructose 2,6-bisphosphate, or PEP [176]. However, this work was only performed on the beta subunit, as full-length recombinant PfPFK9 could not be purified. Further studies are required to understand the regulation and behavior of the full-length PfPFK9 enzyme and the possible regulatory role of the PfPFK9 alpha subunit.

1.6 THE HALOACID DEHALOGENASE-LIKE HYDROLASE (HAD) PROTEIN FAMILY

Given its obvious importance for energy and the synthesis of essential biomolecules, microbes have evolved many ways to regulate sugar metabolism. One such set of regulators is the haloacid dehalogenase-like hydrolase (HAD) superfamily of enzymes. The HAD superfamily (Pfam PF08282, Interpro IPR023214) [180,181] is comprised of thousands of proteins found in organisms in all three kingdoms of life [182]. The large size of the HAD superfamily has allowed its members to evolve diverse functions [183]. While HADs perform a variety of chemical reactions and participate in many biological processes, they share similar domain organization and structural features, such as a Rossmann-like fold and active site aspartate nucleophile [182]. Structurally, HAD proteins are characterized by the insertion of various cap
domains, which vary in size and insert location in the protein sequence. C0 caps are small and relatively simple, while C1 and C2 caps are larger [183].

The HAD protein family is dominated by phosphatases (>75%) [183], including the Cof-like hydrolase subfamily of phosphatases (IPR000150) [181]. Modification of the C2 cap has allowed for diverse substrate utilization within this group. Most members of this subfamily are found in bacteria and, similar to other members of the HAD superfamily, remain uncharacterized. However, studies have uncovered the functions of a number of bacterial members of this subfamily. A genome-wide screen for HAD enzyme activity in *E. coli* revealed that members of this subfamily (YidA, YbhA, YigL, YbiV, YbjI) possess broad sugar phosphatase activity, with substrates participating in key metabolic pathways such as glycolysis and the pentose phosphate pathway. Notably, most of these HADs possess a relatively low affinity (K_m > 1 mM) for their substrate(s), drawing into question whether or how this substrate utilization is physiologically relevant [184].

While a number of bacterial Cof-like hydrolases have been studied, the expansion of the HAD protein family in bacteria and the overlapping substrate profiles of its members have presented a challenge in determining their in vivo functions. *E. coli* YidA was identified as a potential positive regulator in a screen for increased lycopene production [185]. Bacterial strains with attenuated YbiV have been patented for increased amino acid production [186]. A bioengineering study showed that YidA and YbiV expression in *E. coli* resulted in increased ribose production, supporting these proteins’ role in sugar phosphate metabolism [187]. YbhA is a member of the molybdate transport operon in *E. coli* [188]. YigL expression has been shown to
be repressed under anerobic conditions and increased under phosphosugar stress [189,190]. While further studies are require to confidently define the in vivo function of these bacterial HADs, collective evidence points to their function as metabolic regulators through hydrolysis of sugar phosphate substrates.

Given their expansion in bacteria, it is perhaps not surprising that HAD family members are found in organisms that possess a plastid organelle (plants and apicomplexan parasites). HAD proteins that have been characterized from plants have functions similar to their bacterial homologs. The Cof-like hydrolase YbeY from Arabidopsis thaliana functions as a chloroplast endoribonuclease [191] and PvNTD1 from the French bean Phaseolus vulgaris is a nucleotidase [192]. The sugar phosphatases AtGpp1, AtGpp2, and AtSgpp from A. thaliana show the broad substrate specificity and low substrate affinity (K_m > 1 mM) seen in bacterial HADs and are hypothesized to be involved in stress response and phosphate homeostasis [193,194].

1.7 AIM AND SCOPE OF THE DISSERTATION

Similar to what has been proposed in bacteria and other systems [182], the domain organization and broad substrate utilization of the HAD superfamily lends itself to adaptation to diverse metabolic niches. One such metabolic niche is that of P. falciparum, an obligate intracellular parasite dependent on constant glucose consumption for production of ATP and essential biomolecules, including isoprenoids. Despite the essentiality of this metabolism and the need to identify and characterize additional antimalarial targets, relatively little is known regarding metabolic regulation in the parasite.
We hypothesized that resistance to inhibition of isoprenoid synthesis would uncover novel regulators of this critical pathway. We utilize a forward genetics approach and the isoprenoid synthesis inhibitor FSM to identify the PfHADs, a class of novel regulators whose loss confers resistance to FSM. We provide the first characterization for this protein family in *P. falciparum* and elucidate the effects of its members on parasite metabolism, fitness, and drug tolerance.

We also evaluated parasite survival following FSM treatment in clinical populations and find that there does not appear to be a genetic biomarker of recrudescence following FSM treatment failure. Together, this work expands our understanding of both metabolic regulation and drug resistance in a clinically important pathogen.
Figure 1. Life cycle of *Plasmodium falciparum*.

Infection begins with the injection of sporozoites into the host bloodstream by the bite of an Anopheles mosquito. Parasites multiply in the liver and are released back into the host bloodstream as merozoites, where they begin the intraerythrocytic developmental cycle (RBCs, red blood cells). Inside the erythrocyte, parasites grow into large trophozoites. They eventually divide to become multinucleate schizonts, which erupt from the host cell and reenter the blood as merozoites. A proportion of these blood-stage parasites become gametocytes and are taken up by the mosquito vector, where they complete sexual replication.
**Figure 2. Synthesis of isoprenoid products in *P. falciparum*.**

Electron micrograph of a *P. falciparum* cell, with labels showing the red blood cell (RBC), nucleus (N), food vacuole (FV), and apicoplast (Ap). Scale bar represents 500 nm. (B) The *P. falciparum* apicoplast is the site of isoprenoid synthesis by the MEP pathway. It is surrounded by four membranes, indicative of secondary endosymbiotic origins. Scale bar represents 100 nm. (C) Isoprenoid products produced by *P. falciparum*. Abbreviations used: phosphoenol pyruvate (PEP), dihydroxyacetone phosphate (DHAP), pyruvate kinase (PK), triose phosphate isomerase (TPI), 1-deoxy-D-xylulose 5-phosphate synthase (DXS), 1-deoxy-D-xylulose 5-phosphate (DOXP), DOXP reductoisomerase (DXR), fosmidomycin (FSM), 2-C-methyl-D-erythritol 4-phosphate (MEP), MEP cytidyltransferase (IspD), 4-diphosphocytidyl-2-C-methylerythritol (CDP-ME), CDP-ME kinase (IspE), 4-diphosphocytidyl-2-C-methylerythritol 2-phosphate (CDP-MEP), 2-C-methyl-D-erythritol 2,4-cyclopyrophosphate (MEcPP), MEcPP synthase (IspF), (E)-4-hydroxy-3-methyl-but-2-enyl pyrophosphate (HMB-PP), HMB-PP synthase (IspG), HMB-PP reductase (IspH), dimethylallyl pyrophosphate (DMAPP), isopentenyl pyrophosphate (IPP), isopentenyl pyrophosphate isomerase (IPPI), tRNA isopentenyltransferase (TIPT), tRNA methylthiolase (TMT), farnesyl pyrophosphate synthase (FPPS), bisphosphate (BisP), geranyl pyrophosphate (GPP), polyisoprene reductase (PPR), farnesyl pyrophosphate (FPP), octaprenyl pyrophosphate synthase (OPPS), nerolidol (NER), farnesyl transferase (FTase), prenyltransferase inhibitors (PTI), geranylgeranyl transferase (GGTase), geranylgeranyl pyrophosphate (GGPP), phytoene synthase (PS), phytoene desaturase (PD), norflurazon (NOR), dolichyl pyrophosphate (dolichyl-PP), 2-methyl-6-phytyl-1,4-benzoquinol (MPBQ), and usnic acid (UA).
Figure 3. Enzymatic steps of glycolysis.

Abbreviations used: HK, hexokinase; glu6P, glucose 6-phosphate; PGI, phosphoglucone isomerase; fru6P, fructose 6-phosphate; PFK, phosphofructokinase; FBP, fructose 1,6-bisphosphate; ALD, aldoase; gly3P, glyceraldehyde 3-phosphate; DHAP, dihydroxyacetone phosphate; TPI, triose phosphate isomerase; PGK, phosphoglycerate kinase; 1,3BPG, 1,3-bisphosphoglycerate; PGM, phosphoglycerate mutase; 3-PGA, 3-phosphoglyceric acid; ENO, enolase; 2-PGA, 2-phosphoglyceric acid, PK, pyruvate kinase, PEP, phosphoenolpyruvate, LDH, lactate dehydrogenase.
1.9 REFERENCES


63. Kavanagh KL, Guo K, Dunford JE, et al. The molecular mechanism of nitrogen-


103. Gruszecki WI, Strzalka K. Carotenoids as modulators of lipid membrane physical


142. Dieckmann-Schuppert A, Bause E, Schwarz RT. Studies on O-glycans of Plasmodium-


168. Theodorou ME, Cornel FA, Duff SM, Plaxton WC. Phosphate starvation-inducible


193. Caparrós-Martín JA, McCarthy-Suárez I, Culiáñez-Macià FA. HAD hydrolase function

Chapter 2

The sugar phosphatase PfHAD1 is a metabolic regulator in malaria parasites
PREFACE

The following work was performed by me, Jooyoung Park, Rachel L. Edwards, Megan L. Kelly, Dana M. Hodge, Niraj H. Tolia, and Audrey R. Odom. I quantified FSM resistance, prepared samples for genome sequencing and variant discovery, performed parasite transfections and genetic complementation, performed enzymatic characterization, and performed live fluorescent microscopy, and wrote the manuscript. J.P. performed crystallization, data collection, and structure solution, and wrote the manuscript. R.L.E. prepared samples for mass spectrometry. M.L.K. generated drug-resistant parasite strains. D.M.H. performed confocal microscopy and quantified FSM resistance. A.R.O. and N.H.T. designed, supervised, and analyzed the studies and wrote the manuscript.

With the exception of the data presented in Supplementary Figure 6, this chapter is published in its entirety: [Guggisberg AM*, Park J*, Edwards RL, Kelly ML, Hodge DM, Tolia NH**, Odom AR**. (2014). A sugar phosphatase regulates the methylerthritol phosphate (MEP) pathway in malaria parasites. Nature Communications, 5:4467. (**/** co-first/co-last authors)] and is available online at: [http://www.nature.com/ncomms/2014/140724/ncomms5467/full/ncomms5467.html]. Reproduction of this article is permitted as part of our ownership of copyright and the author reuse guidelines of Nature Publishing Group. Data presented in Section 2.4.2 and Supplementary Table 1 has been updated from the published form to reflect additional findings since publication.

We thank Andreas Mitchell for excellent technical assistance in the generation of FSM\textsuperscript{R} strains. We also thank J. Nix and ALS Beamline 4.2.2 (contract DE-AC02-05CH11231) for assistance
with X-ray data collection. We are grateful to Daniel Goldberg (Washington University) for critical reading of the manuscript. We thank the Proteomics and Mass Spectrometry Facility at the Donald Danforth Plant Science Center, St. Louis, MO for assistance with LC-MS/MS. We thank the Malaria Research Reference and Reagent Resource (MR4) for providing us with reagents contributed by DJ Carucci (MRA-102), AF Cowman (MRA-568), and John Adams (MRA-912).

This work was supported by the Children’s Discovery Institute of Washington University and St Louis Children’s Hospital (MD-LI-2011-171, to AO and NT), NIH/NIAID R01AI1103280 (to AO), a March of Dimes Basil O’Connor Starter Scholar Research Award (to AO) and a Doris Duke Charitable Foundation Clinical Scientist Development award (to AO). A.M.G. is supported by an NIGMS Training grant (5T32GM007067) and a Monsanto Excellence Fund Graduate Fellowship. J.P. is supported by a Sondra Schlesinger Graduate Student Fellowship from Washington University. M.L.K. was supported by an HHMI Summer Scholars in Biology and Biomedical Research grant and a Washington University Summer Undergraduate Research Fellowship. M.L.K. was also supported by an ASM Undergraduate Research Fellowship and the Amgen Scholars Program at Washington University.
2.1 ABSTRACT

Isoprenoid biosynthesis through the methylerythritol phosphate (MEP) pathway generates commercially important products and is a target for antimicrobial drug development. MEP pathway regulation is poorly understood in microorganisms. Here we employ a forward genetics approach to understand MEP pathway regulation in the malaria parasite, *Plasmodium falciparum*. The antimalarial fosmidomycin inhibits the MEP pathway enzyme deoxyxylulose 5-phosphate reductoisomerase (DXR). Fosmidomycin-resistant *P. falciparum* are enriched for changes in the PF3D7_1033400 locus (hereafter referred to as *PfHAD1*), encoding a homologue of haloacid dehalogenase (HAD)-like sugar phosphatases. We describe the structural basis for loss-of-function *PfHAD1* alleles and find that PfHAD1 dephosphorylates a variety of sugar phosphates, including glycolytic intermediates. Loss of PfHAD1 is required for fosmidomycin resistance. Parasites lacking PfHAD1 have increased MEP pathway metabolites, particularly the DXR substrate, deoxyxylulose 5-phosphate. PfHAD1 therefore controls substrate availability to the MEP pathway. Because *PfHAD1* has homologs in plants and bacteria, other HAD proteins may be MEP pathway regulators.

2.2 INTRODUCTION

Isoprenoids are the largest and most diverse class of biomolecules [1]. Isoprenoids perform essential cellular functions such as respiration (ubiquinone) and membrane localization of proteins (prenylation). There is considerable commercial interest in the production of isoprenoid-derived secondary metabolites as pharmaceuticals (for example, artemisinin and taxol) and
biofuels. Isoprenoids are synthesized from two five-carbon precursor molecules, isopentenyl pyrophosphate and its isomer, dimethylallyl pyrophosphate. Two metabolically and enzymatically distinct pathways exist for the synthesis of isopentenyl pyrophosphate and dimethylallyl pyrophosphate. Archaea, fungi and animals utilize a mevalonate-dependent pathway. In contrast, plastid-containing eukaryotes and most bacteria use an alternate route that proceeds through the key metabolite methylerthritol phosphate (MEP) (Figure 1) [2].

In the MEP pathway, two glycolytic intermediates (glyceraldehyde 3-phosphate (gly3P) and pyruvate) are combined to generate deoxyxylulose 5-phosphate (DOXP) by DOXP synthase (DXS; E.C. 2.2.1.7). DOXP is subsequently reduced and isomerized to MEP in the first dedicated step of the MEP pathway, catalyzed by deoxyxylulose phosphate reductoisomerase (DXR; E.C. 1.1.1.267). These first two steps of the MEP pathway are rate limiting in MEP-dependent isoprenoid biosynthesis [3–5]. Organisms may increase flux through the MEP pathway by increasing DXS and DXR production, or by increasing substrate availability for these enzymes [6].

The phosphonic acid antibiotic, fosmidomycin (FSM), competitively inhibits DXR in vitro [7]. Metabolic profiling of FSM-treated cells confirms that FSM inhibits MEP pathway metabolism within cells [8]. The growth effects of FSM are rescued by supplementation with downstream isoprenoids, confirming that the biological effects of FSM are mediated through isoprenoid biosynthesis inhibition [8–10]. These studies validate the use of FSM as a specific tool to probe the MEP pathway. Because chlorophyll biosynthesis requires MEP pathway-derived isoprenoids, FSM is a herbicide [11]. FSM resistance in the model plant, *Arabidopsis thaliana*, has been used
to gain insight into MEP pathway regulation, and these studies have revealed post-transcriptional control of MEP pathway enzymes through an RNA metabolism protein (Rif10) [12].

In contrast, there is little understanding of the MEP pathway regulation in non-model organisms. The malaria parasite, *Plasmodium falciparum*, depends on *de novo* isoprenoid biosynthesis through the MEP pathway [9,13]. Since drug resistance is one of the greatest challenges in malaria eradication, the MEP pathway is an attractive parasite-specific target for antimalarial drug development. As in plants, the MEP pathway in apicomplexan parasites is localized to a plastid-like organelle. The parasite organelle is called the apicoplast [14], and isoprenoid synthesis may be the only required function of the apicoplast during blood-stage development [9,10]. Though not photosynthetic, the apicoplast shares a similar endosymbiotic origin with the plant chloroplast [15]. This uniquely positions *P. falciparum* in studies of MEP pathway biology as a plastid-containing eukaryote that is also a globally important pathogen.

Here, to gain insight into how malaria parasites regulate isoprenoid precursor synthesis, we screen for *P. falciparum* strains able to survive MEP pathway inhibition by FSM. We uncover the first regulator of MEP pathway metabolism in malaria parasites, PF3D7_1467300 (*PfHAD1*), a member of the haloacid dehalogenase (HAD) superfamily. Using detailed structural and biochemical studies, we show that FSM resistance is correlated with changes in PfHAD1 that ablate function. Further, we find that PfHAD1 is a cytosolic sugar phosphatase that dephosphorylates the intermediates of glycolysis. Loss of PfHAD1 function increases substrate availability to the MEP pathway and increases levels of isoprenoid precursors, thus conferring FSM resistance. Since close homologs of *PfHAD1* are present in other MEP pathway-containing
organisms, sugar phosphatases may represent a common strategy to regulate MEP pathway flux.

2.3 METHODS

2.3.1 Reagents

Reagents were purchased from Sigma-Aldrich unless otherwise indicated.

2.3.2 Maintenance of P. falciparum cultures

The P. falciparum strain used for FSM selection is derived from genome reference strain 3D7 [16]. The ACP_L-GFP (expresses apicoplast-targeted GFP) strain used for immunofluorescence microscopy is described in Waller, et al. [17]. 3D7 and ACP_L-GFP were obtained from the Malaria Research and Reference Reagent Resource Center (strains MRA-102 and MRA-568, respectively, MR4, ATCC, Manassas, Virginia). Unless otherwise stated, P. falciparum strains were cultured at 37 °C in a 2% suspension of human erythrocytes in RPMI-1640 medium (Sigma-Aldrich, SKU R4130) supplemented with 27 mM sodium bicarbonate, 11 mM glucose, 5 mM HEPES, 1 mM sodium pyruvate, 0.37 mM hypoxanthine, 0.01 mM thymidine, 10 µg/ml gentamycin and 0.5% albumax (Life Technologies) in a 5% O₂/5% CO₂/90% N₂ atmosphere, as previously described [8,18]. Culture growth was monitored by microscopic analysis of Giemsa-stained blood smears.

2.3.3 Generation of FSM\textsuperscript{R} P. falciparum

A 3D7 parental strain was cloned by limiting dilution and used as the parental strain for all
selections. Independent selections are defined as those cultures that were cultured in separate wells before and during selection. The parental strain was at 4% parasitemia in 4 ml cultures at the initiation of FSM selection. Parasites were initially cultured in media containing 0.5–1 µM FSM and gradually scaled to final concentrations of 2–4 µM FSM. The following FSM\textsuperscript{R} strains were cloned by limiting dilution: E1, D6, MK1, D3, AM1 and AM2.

### 2.3.4 Quantification of FSM resistance

Asynchronous \textit{P. falciparum} cultures were diluted to 1% parasitemia and were treated with FSM (Life Technologies) at concentrations ranging from 0.025–500 µM. Growth inhibition assays were performed in opaque 96-well plates at 100 µl culture volume. After 3 days, parasite growth was quantified by measuring DNA content using Picogreen (Life Technologies), as previously described [19]. Picogreen fluorescence was measured on a FLUOstar Omega microplate reader (BMG Labtech) at 485 nm excitation and 528 nm emission. IC\textsubscript{50} values were calculated by nonlinear regression analysis using GraphPad Prism software.

### 2.3.5 Whole genome sequencing and variant discovery

Genomic DNA was isolated from \textit{P. falciparum} cells by a standard phenol-chloroform extraction and ethanol precipitation protocol [20]. Sequencing libraries were prepared and sequenced by the Washington University Genome Technology Access Center (GTAC). Briefly, 5 µg of genomic DNA was sheared to an average size of 175 bp. Fragments were end repaired and an ‘A’ base was added to the 3’ end. Adapters containing index sequences were ligated to the fragments. Resulting libraries were sequenced on an Illumina HiSeq 2000 to obtain 101 bp paired end reads. Bioinformatic analyses were performed by GTAC. Reads were not merged, but were aligned to
the *P. falciparum* reference genome using Novoalign (Novocraft Technologies). PCR duplicates were removed using Picard and variants were analyzed using SAMtools. SNPs were called if present in >50% of the reads. For all samples, at least 75% of coding regions were sequenced at ≥5X coverage. All whole genome sequencing data is available in the NCBI BioProject database and Sequence Read Archive.

### 2.3.6 Sanger sequencing of PfHAD1

The region of interest was amplified from *P. falciparum* genomic DNA using the following gene-specific primers: 5′-GGTTCAAGGGTGATAGATAGGA-3’, 5′-CGAAGGTCCAACATAAGCAG-3’, 5′-ATGCACGAAATTGTAGATAAGAATG-3’, 5′-AATTCATGTATCTCCATTTTCAAGTC-3’ and 5′-GATCACTACATTTTGGACGTACAC-3′. Purified PCR products were sequenced by the Washington University Protein and Nucleic Acid Laboratory using BigDye Terminator v3.1 Cycle Sequencing reagents (Life Technologies). Chromatogram files were analyzed using DNAStar SeqMan software. Representative traces for all strains are available through the NCBI Trace Archive.

### 2.3.7 Recombinant expression and purification of PfHAD1 variants

All PfHAD1 (PlasmoDB ID PF3D7_1033400) alleles were amplified from *P. falciparum* genomic DNA using the following primers: 5′–

CTCACCCACCACCACCACCATATGCACGAAATTGTAGATAAGA-3’ and 5′-ATCCTATCTTACTCATTATGTGCACGAGAATGTCTTCA-3’. The PCR product was cloned by ligation-independent cloning into vector BG1861 [21], which introduces an N-terminal 6xHis tag and the construct was verified by Sanger sequencing. BG1861:6xHis-PfHAD1 was
transformed into BL21(DE3)pLysS *Escherichia coli* cells (Life Technologies). Following induction with isopropyl-β-D-thiogalactoside, cells were collected by centrifugation and stored at −20 °C. Pellets were resuspended in lysis buffer containing 10 mM Tris HCl (pH 7.5), 20 mM imidazole, 01 mM MgCl2, 01 mM dithiothreitol, 1 mg/ml lysozyme, 100 U benzonase and Complete Mini EDTA-free protease inhibitor tablets (Roche Applied Science). 6xHis-PfHAD1 was bound to nickel agarose beads (Gold Biotechnology), eluted in 300 mM imidazole, 20 mM Tris HCl (pH 7.5) and 150 mM NaCl, and dialyzed in buffer lacking imidazole before analysis. Protein was flash frozen and stored at −80 °C.

### 2.3.8 Crystallization and structure determination

PfHAD1 was further purified by gel chromatography using a HiLoad 16/600 Superdex 200 pg column (GE Healthcare) equilibrated with 10 mM Tris (pH 8.0), 150 mM NaCl and 5 mM dithiothreitol. The fractions containing PfHAD1 were pooled and concentrated to 20 mg/ml using a centrifugal filter.

Crystals of PfHAD1 were obtained by vapor diffusion using hanging drops equilibrated at 18 °C against 500 µl of a reservoir containing 0.1 M HEPES (pH 7.5) and 20–25% PEG 8000. Crystals were cryo-protected with 0.1 M HEPES (pH 7.5) and 30% PEG 8000 before flash-freezing under liquid nitrogen for data collection. X-ray data were collected from a single crystal using a wavelength of 1 Å at synchrotron beamline 4.2.2 of the Advanced Light Source in Berkeley, CA. Data were processed with XDS [22] and the coordinates from PDB 2B30 were used as a search model for phase determination by molecular replacement using PHENIX [23]. Rigid-body refinement followed by automatic model rebuilding were performed in PHENIX [23].
Subsequent iterated manual building/rebuilding and refinement of models were performed using Coot [24] and PHENIX [23], respectively. The structure validation server MolProbity [25] was used to monitor refinement. The final refined model results in a Ramachandran plot with 96.75% of residues in the favored region and 3.25% in the allowed region. A complete summary of the final crystallographic refinement statistics are given in Table 1 and a stereo image of a representative region of the electron density map is shown in Supplementary Figure 5. Figures were prepared using PyMOL Molecular Graphics System, Version 0.99rc6, Schrödinger, LLC. Structure data is deposited in the RSCB Protein Structure Database.

2.3.9 pTEOE110: PfHAD1-GFP plasmid construction

The construct used to express PfHAD1 in FSM\textsuperscript{R} strains, pTEOE110, was a gift from Daniel Goldberg (Washington University, St. Louis). The construct contains the heat shock protein 110 (Hsp110, PF3D7\_0708800) 5′ UTR [26] and a carboxyl-terminal GFP tag. The human dihydrofolate reductase selectable marker confers resistance to WR92210. These sites are flanked by inverted terminal repeats [27] that allow for integration of the segment containing the Hsp110 5′ UTR, the gene of interest, GFP and human dihydrofolate reductase into the genome, facilitated by a co-expression from a plasmid encoding the piggyBac transposase (pHTH, MRA912, MR4, ATCC, Manassas, Virginia).

PfHAD1 was amplified using the following primers: 5′-
CATGCTCGAGATGCACGAAATTGTAGATAAGAA-3′ and 5′-
CATGCCTAGGTATGTCACAGAATGTCTTCAAG-3′. The resulting PCR product was cloned into the XhoI and AvrII sites in pTEOE110 to generate pTEOE:PfHAD1. Successful insertion
was confirmed by colony PCR and Sanger sequencing.

2.3.10 *P. falciparum* transfection

Seventy-five micrograms of each plasmid (pTEOE110: PfHAD1 and pHTH) was combined, precipitated and resuspended in 400 µl Cytomix (120 mM KCl, 0.15 mM CaCl₂, 2 mM EGTA, 5 mM MgCl₂, 10 mM K₂HPO₄, 25 mM HEPES, pH 7.6).

A 4-ml culture of FSM⁺ strain AM1 was synchronized by treatment with 5% sorbitol to obtain ring-stage parasites at >4% parasitemia. Cells were washed with Cytomix and resuspended in 400 µl DNA/Cytomix solution. Cells were electroporated using a Biorad GenePulser II electroporator set to 950 µF capacitance and 0.31 kV. Transfected cells were washed with media and returned to the previously described culture conditions. Culture media was replaced daily for one week. Cultures were maintained under selection with 5 nM WR92210 (Jacobus Pharmaceutical Co.) beginning 48 h post transfection. Parasite growth was monitored by microscopy. Successful transfectants were cloned by limiting dilution. The resulting strain was verified by microscopy, western blot and sequencing of the endogenous *PfHAD1* locus.

2.3.11 Generation of PfHAD1 polyclonal antisera

PfHAD1 rabbit polyclonal antisera was produced by Cocalico Biologicals (Reamston, PA) using their standard protocol. Purified recombinant 6XHis-PfHAD1 was used as an antigen. Titermax was used as an adjuvant.
2.3.12 Immunoblotting

*P. falciparum* cells were obtained by treatment of infected erythrocytes with 0.1% saponin. Lysates were separated on a polyacrylamide gel and transferred to polyvinylidene fluoride membrane. For the detection of PfHAD1, the membrane was blocked and probed with 1:20,000 anti-PfHAD1 polyclonal antisera (Cocalico Biologicals), followed by 1:20,000 HRP-conjugated goat anti-rabbit IgG polyclonal antibody (Life Technologies). The membrane was stripped by treatment with 200 mM glycine, 0.1% sodium dodecyl sulfate, 1% Tween-20 (pH 2.2) and reprobed with 1:5,000 anti-heat shock protein 70 antibody (Agrisera) as a loading control.

2.3.13 Enzymatic assays

All assays were performed in clear 96-well half-area plates using a FLUOstar Omega microplate reader (BMG Labtech) at 37 °C. Reaction rates were determined using GraphPad Prism software.

Phosphatase activity was measured using the substrate *p*NPP (New England Biolabs). Reactions were performed in 50 µl volumes with 10 mM *p*NPP, 50 mM Tris HCl (pH 7.5), 5 mM MgCl₂ and 0.5 mM MnCl₂. Reactions contained 2 µg purified recombinant enzyme. *Para*-nitrophenyl production was quantified by absorbance at 405 nm.

Enzyme activity against phosphorylated sugar substrates was measured using the EnzChek Phosphate Assay Kit (Life Technologies) according to supplier instructions. Each assay contained 200 ng recombinant purified enzyme and 1 mM substrate. Substrates were purchased from Sigma-Aldrich, except for DOXP and MEP (Echelon Biosciences), fructose 1-phosphate (Santa Cruz Biotechnology) and FSM (Life Technologies).
For the determination of kinetic parameters, assays were performed using 50 ng 6XHis-PfHAD1 and 0.03–12 mM substrate. Data were fit to Michaelis–Menten curves using GraphPad Prism software.

### 2.3.14 Microscopy

For immunofluorescence microscopy, cells at ~10% parasitemia were fixed in 4% paraformaldehyde in PBS. Fixed cells were washed with 50 mM ammonium chloride, permeabilized by treatment with 0.075% NP-40 in PBS and blocked using 2% bovine serum albumin. Cells were incubated with the following antibody dilutions: 1:10,000 rabbit polyclonal anti- PfHAD1 (described above) and 1:1,000 mouse monoclonal anti-GFP (Life Technologies #A11120) for the detection of GFP in the ACP₄-GFP strain (MRA-568). Hoechst 33258 (Life Technologies) was used as a nuclear counterstain. Secondary antibodies used were 1:200 dilutions of Alexa Fluor 568 goat anti-rabbit IgG (Life Technologies #A11011) and Alexa Fluor 488 goat anti-mouse IgG (Life Technologies #A11029). Images were obtained on an Olympus Fluoview FV1000 confocal microscope.

For imaging of live cells, cells were stained with 10 ng/µl Hoechst 33258 for 10 min and mounted under coverslips on polysine adhesion slides. Cells were visualized on an Olympus BH8 microscope at room temperature for no longer than 15 min.

For all microscopy, minimal adjustments in brightness and contrast were applied equally to all images.
2.3.15 Sample preparation for mass spectrometry analysis

*P. falciparum* were cultured in 75 ml volumes in 150 × 25 mm tissue culture dishes at 4% hematocrit until >6% parasitemia was reached. Cultures were synchronized by one to two treatments with 5% sorbitol until >8% parasitemia and >75% of parasites were in ring-stage growth. Once these criteria were met, 75 ml cultures were split into 3 × 25 ml samples. Ring-stage cultures were then treated +/- 5 µM FSM (Life Technologies) for either 0 or 10 h. At each time point, parasite-infected erythrocytes were lysed with 0.1% saponin, the parasite pellets washed with PBS and the pellets stored at −80 °C until extraction. Samples were extracted and analyzed by liquid chromatography-mass spectrometry as previously described [8]. Briefly, samples were extracted in chloroform, methanol and acetonitrile (2:1:1, v/v/v) and homogenized. Water was added and the sample was homogenized again. The polar phase was separated by centrifugation, lyophilized and resuspended in water for analysis.

2.3.16 Quantitative PCR of *PfDXS* and *PfDXR* transcripts

Infected erythrocytes were lysed by treatment with 0.1% saponin. Parasite pellets were stored at −80 °C until extraction. Total RNA was isolated using the Ambion Purelink RNA Mini kit (Life Technologies) according to the supplier’s instructions. After the first wash step, samples underwent an on-column DNase treatment (Qiagen) for 30 min. RNA was eluted in RNase-free water, quantified and stored at −80 °C until further use. At least 100 ng of RNA was used to synthesize cDNA using the Quantitect Reverse Transcription kit (Qiagen), according to the supplier’s instructions.
Amplification of cDNA was performed using an Applied Biosystems Fast RT PCR System and Applied Biosystems Fast SYBR Green PCR Master Mix (Life Technologies). PfDXS (PlasmoDB ID PF3D7_1337200) was amplified using the following primers: 5’-AACGTGGATAAAAGTACACATTGC-3’ and 5’-TGATATACCTACGGCATTTGTTGG-3’ and PfDXR (PlasmoDB ID PF3D7_1467300) was amplified using the following primers; 5’-ACATGGCCTGATAGAATAAAAACA-3’ and 5’-TTCATTTGACGCATTTAGTACAGTT-3’ (Sigma-Aldrich). The gene encoding beta tubulin (PF3D7_1008700) was used as a control and was amplified using the following primers: 5’-ATCCCATCACCACGTTTACATT-3’ and 5’-TCCTTTGTGGACATTC-TTCCTC-3’ (Eurofins MWG Operon). Standard curves were used to verify primer efficiency (r^2 >0.97, E >80%). In addition to SYBR Master Mix, reactions contained 10 ng cDNA and 300 nM of each primer. All reactions were performed in MicroAmp Fast Optical 96-well plates (Life Technologies).

Thermocycling parameters were as follows: 95 °C for 30 s, 40 cycles of 95 °C for 3 s and 60 °C for 30 s. Agarose gel electrophoresis, as well as melt-curve analysis, with temperatures ranging from 60 to 95 °C, confirms production of a single product from each primer pair. Controls lacking either reverse transcriptase or template produced no significant signals. C_t values were generated using 7500 Fast System software (Applied Biosystems).

Relative PfDXS and PfDXR expression levels were calculated using the ∆∆C_t method. ∆C_t values were calculated by subtracting the C_t of the reference gene (beta tubulin) from the target gene (PfDXS or PfDXR). ∆∆C_t values were then calculated by normalizing to the ∆C_t of the parental sample. Relative expression changes were calculated as 2^{-\Delta\Delta C_t}. 

62
2.4 RESULTS

2.4.1 Selection of FSM-resistant (FSM\(^\text{R}\)) \(P. falciparum\)

The MEP pathway is essential for the synthesis of isoprenoids in \(P. falciparum\) [8,9,13]. The phosphonic acid antibiotic FSM specifically inhibits this pathway, primarily via competitive inhibition of PfDXR [7] (Figure 1). A FSM-sensitive 3D7 parental strain was subjected to FSM treatment at ~two to fourfold the wild-type half maximal inhibitory concentration (IC\(_{50}\); Figure 2A, Supplementary Table 1). This selection generated 20 independent FSM\(^\text{R}\) strains of \(P. falciparum\) with increased IC\(_{50}\)s for FSM ranging from 1.5–8.1 \(\mu\)M (Figure 2B, Supplementary Table 1). FSM\(^\text{R}\) strains are also resistant to the related antibiotic, FR-900098 [28] (Supplementary Figure 1).

FSM resistance has been associated with copy number variations in the FSM target, \(PfDXR\) (PF3D7_1467300) [29]. However, copy number changes that result in drug resistance often revert when selection is relieved [30]. To discourage \(PfDXR\) copy number variants, FSM\(^\text{R}\) strains were cycled on and off drug treatment. All strains maintain their resistance phenotype after prolonged growth without FSM, and no changes in \(PfDXR\) transcript levels are observed in FSM\(^\text{R}\) strains (Supplementary Figure 2).

To control for genetic changes arising from prolonged culture or culture stresses, we performed control selections in which the parental 3D7 strain was cultured continuously for \(>3\) months and, in some cases, was subjected to culture stresses such as depleted or incomplete media and aged
 (>90 days) erythrocytes. These strains retain their FSM sensitivity as measured by IC$ _{50}$ (mean IC$ _{50}$ 1.8 ± 0.4 µM).

### 2.4.2 FSM$ ^R $ strains are enriched for genetic changes in PfHAD1

To uncover genetic changes correlated with FSM resistance, we genotyped the parent strain, six FSM$ ^R $ strains, and two control strains by whole genome sequencing with at least 15X genome coverage. Sanger sequencing was used to verify single-nucleotide polymorphisms (SNPs) and insertion–deletions (indels) and to genotype loci of interest in strains whose genomes were not sequenced. We prioritized genetic changes found in coding regions and did not evaluate changes in highly variable loci (such as var and rifin genes) [31,32]. A list of SNPs uniquely identified in the FSM$ ^R $ strains is found in Supplementary Data 1 (available online [33]). The resulting data reveal a significant enrichment for nonsynonymous changes in PF3D7_1033400 in FSM$ ^R $ strains compared with control strains (Figure 3A). We hereafter refer to PF3D7_1033400 as PfHAD1. Seventeen of the twenty FSM$ ^R $ strains have a nonsynonymous change in PfHAD1, while seven control laboratory strains do not have any sequence changes in PfHAD1 (P<0.01, Fisher’s Exact Test).

Thirteen alleles were represented in the seventeen strains with PfHAD1 changes. Of these alleles, four indels and two nonsynonymous substitutions result in a truncated amino acid sequence (Supplementary Table 1). Six alleles are nonsynonymous substitutions that result in single amino acid changes, once of which results in a loss of the start codon. Five of these alleles are predicted to have deleterious effects on protein function by Polyphen-2, an algorithm for predicting the probability of functional effects of missense mutations (Polyphen-2 score >0.9).
We also identify one strain with a whole-gene deletion of PfHAD1 (Supplementary Table 1). Overall, FSM\textsuperscript{R} strains are significantly enriched for likely deleterious changes in PfHAD1, a previously uncharacterized gene, compared with FSM-sensitive control strains.

In addition to the 3D7 PfHAD1 allele that we designate as wild-type, a nonsynonymous N70S variant has also been reported in sequenced clinical and laboratory isolates of \textit{P. falciparum} [35,36]. This allele is not predicted to be deleterious (Polyphen-2 score=0.007) [34]. Two strains that possess the N70S allele, HB3 and Dd2, are not resistant to FSM (FSM IC\textsubscript{50}s of 1.3 ± 0.2 µM and 0.6 ± 0.09 µM, respectively). Five additional strains whose sequence data are available on public databases (7G8, D10, GHANA1, K1 and V1_S, plasmodb.org, accessed 4 November 2013) also possess the N70S allele. The minimal sequence variation in this gene in natural populations suggests that PfHAD1 has an important cellular function in \textit{P. falciparum}.

2.4.3 PfHAD1 is a Cof-like hydrolase member of the HAD superfamily

Sequence homology places PfHAD1 in the HAD superfamily (Pfam PF08282) [37]. Members of this large superfamily are found in all three kingdoms of life and catalyse carbon- or phosphoryl-transfer reactions [38]. The promiscuous substrate profiles and low levels of sequence similarity between HADs have made determining the \textit{in vivo} function of many superfamily members difficult [39].

Within the HAD superfamily, PfHAD1 has the greatest sequence similarity to the Cof-like hydrolase subfamily (Interpro IPR000150) [40]. This family is named for an \textit{Escherichia coli} enzyme involved in thiamin biosynthesis [39,41]. Most members of the Cof subfamily are
bacterial and most remain uncharacterized. Biochemical characterization suggests that members of this family utilize phosphorylated sugar substrates [39,42].

### 2.4.4 Sequence changes map to core and active site of PfHAD1

*PfHAD1* changes found in FSM<sup>R</sup> strains map along the entire gene body (Figure 3A). To define the structural basis for the role of these point mutations in PfHAD1 function, we solved the crystal structure of PfHAD1 to 2.05 Å (Figure 3B, Table 1). Clear electron density was not observed for the 6xHis tag in either chain, or the amino-terminal (N-terminal) region in chain B (residues 1–19), and were therefore not modeled. Like other members of the HAD superfamily, PfHAD1 consists of two distinct domains: a catalytic core domain (residues 1–106 and 212–288) and a cap domain (residues 107–211). The cap domain is inserted into the linker immediately following the third beta-strand.

The four sequence motifs required for the catalysis [38] of HAD superfamily members are structurally conserved in PfHAD1: motif I—DXD sequence at the end of strand I; motif II—a conserved threonine/serine at the end of strand II; motif III—a conserved lysine at the N terminus of the helix before strand IV; motif IV—a DD, GDxxxD or GDxxxxD on strand IV (Figure 3C) [38]. The sequence 27-DLD-29 in PfHAD1 corresponds to motif I, Thr-61 corresponds to motif II, Lys-215 corresponds to motif III and the sequence 237-GDGEND-242 corresponds to the GDxxxD signature of motif IV. These structural motifs make up the substrate-binding site, located on the interface between the core and cap domains. These motifs are conserved among PfHAD1 homologs found in MEP pathway-containing organisms, including *E.*
coli, Mycobacterium tuberculosis, the alga Chlamydomonas reinhardtii and Arabidopsis thaliana (Figure 3C).

Catalysis by PfHAD1 is metal–ion dependent, and the side chains of Asp-27 and Asp-238, and the backbone carbonyl oxygen of Asp-29, coordinate a magnesium ion in the structure (Figure 3C). On the basis of homology to other HAD family phosphatases [38], it is predicted that the phosphoryl group of sugar phosphate compounds binds by coordinating the magnesium ion and the side chains of Thr-61, Lys-215 and Asp-29. Asp-27 is predicted to perform nucleophilic attack of the phosphate group. As PfHAD1 was crystallized in the absence of a sugar phosphate substrate, a chloride ion occupies the likely phosphate-binding site based on structural similarity to other HAD sugar phosphatases.

The point mutations identified in FSM\textsuperscript{R} strains (T26R, G30E, A60E, W130R and Y148C) were mapped onto the PfHAD1 crystal structure (Figure 3B). Trp-130 and Tyr-148 are located in a tightly packed hydrophobic inner region of the cap domain. Mutations of these aromatic residues to a charged arginine and a small polar cysteine residue, respectively, likely result in the misfolding of the cap domain. Similarly, Thr-26 and Ala-60 are located in the hydrophobic inner region of the core domain, and mutation of these residues to a charged glutamate and arginine, respectively, are predicted to result in the misfolding of the core domain. Finally, Gly-30 is located in the active site of PfHAD1, and a mutation to a glutamate with a much larger side chain is predicted to interfere with substrate binding. Together, the structural mapping of FSM\textsuperscript{R} PfHAD1 mutations suggests that, similar to nonsense mutations, the FSM\textsuperscript{R} nonsynonymous mutations result in the loss of function of PfHAD1 activity through drastic changes in the core or
catalytic regions of PfHAD1. The N70S variant is not suspected to have changes in activity, as Asn-70 is located on the surface of PfHAD1 and away from the substrate-binding site.

2.4.5 FSM\textsuperscript{R} PfHAD1 alleles result in loss of phosphatase activity

We next investigated the enzymatic activity of PfHAD1. Because PfHAD1 has sequence homology to known phosphatases [39] (Figure 3C), we predicted that PfHAD1 is also a phosphatase. We expressed and purified recombinant wild-type PfHAD1, the N70S variant and three full-length variants (G30E, A60E and Y148C) found in FSM\textsuperscript{R} strains. Recombinant PfHAD1 is active against a non-specific phosphatase substrate, \textit{para}-nitrophenyl phosphate (\textit{pNPP}) (Figure 4). This compound has been used in phosphatase screens as an effective indicator of phosphatase activity [39,43]. Activity against \textit{pNPP} is significantly reduced in each of the FSM\textsuperscript{R} PfHAD1 variants (Figure 4). The N70S variant found in sequenced isolates has activity similar to the wild-type enzyme (\(P=0.59\), unpaired Student’s \(t\)-test).

2.4.6 PfHAD1 complementation restores FSM sensitivity

Because our FSM\textsuperscript{R} strains are enriched for changes that result in loss of PfHAD1 function, we confirmed that restoring PfHAD1 expression would restore sensitivity to FSM. Using a piggyBac transposon system [27], functional PfHAD1 was expressed in FSM\textsuperscript{R} strain AM1, which possesses a premature truncation at amino acid 208 of the \textit{PfHAD1} locus and lacks detectable PfHAD1 (Figure 5A). The expression construct produces green fluorescent protein (GFP)-tagged PfHAD1. The rescued strain, AM1 Hsp110: PfHAD1-GFP, retains its original mutation at the endogenous locus and successfully expresses PfHAD1-GFP. The expression pattern of \textit{Hsp110} is similar to that of \textit{PfHAD1} [44]. The expression of functional PfHAD1
restores FSM sensitivity in AM1. While the FSM IC_{50} of AM1 is 5.5 ± 1.5 µM, AM1 Hsp110: PfHAD1-GFP has an IC_{50} of 1.4 ± 0.2 µM, similar to the parental IC_{50} of 1.1 ± 0.2 µM, indicating that loss of PfHAD1 function leads to FSM resistance in our strains (Figure 5B).

2.4.7 PfHAD1 is a sugar phosphatase

Within the HAD superfamily, PfHAD1 has the greatest sequence similarity to the Cof-like hydrolase subfamily of sugar phosphatases. We assayed recombinant wild-type PfHAD1 with a number of phosphorylated sugar substrates (Figure 6). PfHAD1 has a wide substrate utilization profile and low substrate specificity, consistent with other members of this enzyme family [39,45,46]. PfHAD1 does not directly utilize FSM as a substrate.

*In vitro*, PfHAD1 is most active against monophosphorylated three to six-carbon monosaccharides, including intermediates of glycolysis, the pentose phosphate pathway and the MEP pathway (Figure 6). Kinetic parameters for the top three substrates as well as gly3P, a MEP pathway precursor, can be found in Supplementary Table 2. The substrate profile of PfHAD1 is similar to that of related HADs in *E. coli* and *A. thaliana* [39,45]. Like other HADs, PfHAD1 dephosphorylates a variety of substrates *in vitro*. Catalytic efficiencies (k_{cat}/K_{m}) for the four measured substrates were within an order of magnitude and do not clearly indicate a preferred *in vivo* substrate, which is typical of this enzyme class. None of the purified recombinant FSM^{R} PfHAD1 variants have significant activity against sugar phosphate substrates (maximum mean activity <0.9 µmol/min/mg for all variants against man6P, glu6P, fru6P, rib5P, gly3P and DHAP).
2.4.8 PfHAD1 is localized to the cytoplasm

MEP pathway enzymes are localized to the apicoplast in *P. falciparum* [14,47]. To help elucidate the mechanism by which PfHAD1 affects MEP pathway function, we examined the subcellular localization of PfHAD1. We raised specific polyclonal rabbit antisera against recombinant PfHAD1 and performed confocal immunofluorescence to localize PfHAD1. PfHAD1 is highly expressed in blood-stage parasites [48]. PfHAD1 appears to be cytoplasmic and has minimal overlap with the apicoplast marker ACP-L-GFP (described by Waller, et al. [17]; Figure 7). Live fluorescent localization of PfHAD1-GFP (in strain AM1 Hsp110: PfHAD1-GFP) was similar to that of PfHAD1 immunolocalization in the fixed parental strain (Supplementary Figure 4). In addition, PfHAD1 lacks an apicoplast localization signal as predicted by the PlasmoAP and PATS algorithms [49,50]. Given the localization of PfHAD1, the enzyme likely utilizes phosphorylated sugars available in the cytoplasm, such as glycolytic intermediates.

2.4.9 FSM\textsuperscript{R} strains increase levels of MEP pathway intermediates

Since PfHAD1 dephosphorylates glycolytic intermediates, we predicted that the loss of PfHAD1 function would increase substrate availability to the MEP pathway, which requires gly3P and pyruvate. We evaluated the levels of MEP pathway intermediates in eight FSM\textsuperscript{R} *P. falciparum* strains with *PfHAD1* changes, using our previously described quantitative liquid chromatography-mass spectrometry method (Figure 8, Supplementary Table 3) [8]. Parental 3D7 and eight FSM\textsuperscript{R} strains were treated +/- FSM for 10 h and the levels of MEP intermediates were quantified in each strain. Our method quantifies cellular levels of 1-deoxy-D-xylulose 5-phosphate (DOXP), 2-C-methylerythritol 4-phosphate (MEP), 4-diphosphocytidyl-2-C-methylerythritol and 2-C-methyl-D-erythritol 2,4-cyclopyrophosphate (MEcPP).
FSM\textsuperscript{R} strains with \textit{PfHAD1} mutations show an overall increase in levels of MEP pathway intermediates. Compared with the parental strain, FSM\textsuperscript{R} strains have a 55.2 ± 5.9-fold increase in the levels of DOXP, the substrate for PfDXR (median $P<0.01$, unpaired Student’s $t$-test). This increase is maintained during treatment with FSM (78.7 ± 10.1-fold increase, median $P<0.01$; Figure 8, Supplementary Table 3) and would allow for DOXP to compete with FSM binding to PfDXR. Increased DOXP levels are not due to the increased transcript levels of \textit{PfDXS}, the enzyme that produces DOXP (Supplementary Figure 2). Loss of PfHAD1 is necessary for this increase in DOXP, as strain AM1 Hsp110:PfHAD1-GFP has near-parental levels of DOXP (Supplementary Figure 6). FSM-treated FSM\textsuperscript{R} strains also have an increase in MEcPP levels compared with the parental strain (2.7 ± 0.43-fold, median $P<0.01$).

In other systems, FSM resistance is due to the lack of FSM uptake or FSM efflux [51–53]. Metabolic profiling of FSM\textsuperscript{R} parasites suggests that is not the mechanism of resistance in our strains. FSM treatment of FSM\textsuperscript{R} strains results in metabolic changes that are still consistent with inhibition of PfDXR, the target of FSM. Specifically, the levels of the upstream metabolite DOXP increase, and levels of distal metabolites, such as MEcPP, decrease (Figure 8, Supplementary Table 3) [8]. Our data are most consistent with a change in a metabolic regulator outside the MEP pathway that influences production of pathway intermediates, namely DOXP.

\textbf{2.5 DISCUSSION}

In this study, we describe the genetic, biochemical, structural and metabolic basis for the
resistance to the MEP pathway inhibitor FSM in the malaria parasite, *Plasmodium falciparum*. Our study represents a powerful use of next-generation sequencing to facilitate a forward genetics approach in a non-model organism to identify a new regulator of an essential biochemical pathway, the MEP pathway of isoprenoid biosynthesis.

We identify a member of the HAD superfamily, PfHAD1, whose loss-of-function confers FSM resistance in *P. falciparum*. To our knowledge, PfHAD1 is the first cellular factor that regulates the MEP pathway in *P. falciparum*. PfHAD1 has structural similarity to other members of the HAD superfamily, retains the motifs necessary for catalysis, and is active against sugar phosphates. Loss of PfHAD1 function causes dramatically increased levels of MEP pathway metabolites, particularly DOXP, the substrate of PfDXR. Our work suggests a model of PfHAD1 function, shown in Figure 9. We propose that PfHAD1 is a negative regulator of the MEP pathway. It typically catalyses the dephosphorylation of cytosolic phosphometabolites, such as glycolytic intermediates, to decrease the availability of substrates to the MEP pathway. Loss of sugar phosphatase activity in FSM\(^R\) malaria parasites results in increased pools of DOXP, which overcomes competitive inhibition of PfDXR by FSM.

FSM\(^R\) malaria strains are highly enriched for genetic changes in the *PfHAD1* locus. In our study, we identified 11 separate deleterious genotypic changes in *PfHAD1*. If FSM is approved for clinical use, these changes may represent the genetic biomarkers of FSM resistance. While other genetic changes are also present in the FSM\(^R\) strains (Supplementary Data 1, available online [33]), our data indicate that the loss of PfHAD1 is the primary driver of resistance. In addition, changes in *PfHAD1* genotype correlate not only with FSM resistance, but also with increased
cellular levels of isoprenoid precursor metabolites. Complementation of a deleterious *PfHAD1* allele restores FSM sensitivity. These findings strongly support that the primary biochemical change responsible for these phenotypes is the loss of PfHAD1 function.

*PfHAD1* is a member of the large, diverse HAD superfamily of aspartate nucleophile hydrolases, and belongs to the subfamily of Cof-like hydrolases (Interpro IPR00150). Many HAD family members have been structurally and enzymatically characterized [39,42,45,46]. However, the cellular functions of this group of enzymes are largely unknown. The Cof-like hydrolases are absent from animal and fungal genomes, but are highly expanded in bacteria. In addition, Cof-like hydrolases are uniformly present in eukaryotes that possess a plastid-like organelle, such as chloroplast-containing plants, and the apicomplexan parasites, such as *P. falciparum*. This phylogeny parallels that of the MEP pathway, which is also exclusive to many bacteria and plastid-containing eukaryotes, and absent in animals or fungi. Since few studies have addressed the potential biological function of these enzymes, additional work is required to determine whether *PfHAD1* homologs represent regulators of isoprenoid biosynthesis in other MEP pathway-containing organisms. Evidence suggests that this may be the case in bacteria, since the *E. coli* *PfHAD1* ortholog *YidA* was identified in a colorimetric screen of *E. coli* for increased lycopene production [54].

During its intraerythrocytic development cycle, glycolysis is the primary source for ATP generation in *P. falciparum* [55]. As a result, the malaria parasite has great capacity for glucose uptake and carbon flux through glycolysis. PfHAD1 has a broad substrate profile, and readily dephosphorylates a range of three to six-carbon monophosphorylated substrates, including
glycolytic and pentose phosphate intermediates, as well as MEP pathway metabolites. Given this promiscuity, it is unclear which metabolite/s represent the natural *in vivo* substrate(s) for PfHAD1 and other Cof-like subfamily members. Thus, our studies do not distinguish which upstream metabolic step is regulated by PfHAD1. In addition, it is unclear why malaria parasites produce an enzyme whose activity would decrease glycolysis and ATP production. Genetic conservation of *PfHAD1* in clinical and laboratory isolates of *P. falciparum*, as well as among *PfHAD1* homologs in divergent organisms, suggests that these proteins have important cellular functions under typical growth conditions.

The MEP pathway is an important target for the development of small-molecule inhibitors, since this pathway is absent in animals but essential in plants and key human pathogens, including malaria parasites and *M. tuberculosis*. Our studies highlight what may be a particular challenge in developing novel MEP pathway inhibitors. Even though FSM\(^R\) strains are capable of growth in the presence of FSM, drug treatment still substantially reduces levels of the distal MEP pathway metabolite, MEcPP. Following FSM treatment, MEcPP levels in FSM\(^R\) strains are reduced to ~20% of the concentrations present in untreated, wild-type parasites. This suggests that, at least in malaria parasites, cells direct more flux through the MEP pathway than is absolutely required for development, and therefore, they are capable of growth even when MEP pathway metabolism is substantially reduced. As novel MEP pathway inhibitors are developed, it appears that these compounds must achieve a sobering near-complete inhibition of the cellular MEP pathway in order to inhibit malaria parasite growth.
2.6 FIGURES

Figure 1. The MEP pathway for isoprenoid biosynthesis.

The MEP pathway for the synthesis of isoprenoids is specifically inhibited by FSM. FSM competitively inhibits PfDXR. CDP-ME, 4-diphosphocytidyl-2-C-methylerythritol; CDP-MEP, 4-diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate; DMAPP, dimethylallyl pyrophosphate; DXR, DOXP reductase; IPP, isopentenyl pyrophosphate; IspD, CDP-ME synthase; IspE, CDP-ME kinase; IspF, MEcPP synthase.
Figure 2. Generation of FSM-resistant (FSM\textsuperscript{R}) \textit{P. falciparum}.

(a) Schematic of selection of FSM\textsuperscript{R} strains. (b) FSM\textsuperscript{R} strains exhibit a range of resistance phenotypes. FSM\textsuperscript{R} strains have an average FSM IC\textsubscript{50} of 4.8 ± 0.4 μM compared with the wild-type IC\textsubscript{50} of 1.1 ± 0.2 μM. A complete list of strains and their IC\textsubscript{50} values are found in Supplementary Table 1.
Figure 3. FSM\(^{R}\) variants map to the core and active site regions of PfHAD1.

(a) Schematic of PfHAD1 variants found in FSM\(^{R}\) strains. Six of the alleles result in premature stop codons. Five alleles produce full-length protein. N70S is a nondeleterious allele reported from sequenced clinical and laboratory isolates. Conserved active site motifs are shown as black boxes [38].

(b) Overall structure of PfHAD1 with the core domain in blue, cap domain in green and polymorphic residues mapped in yellow. Residues W130 and Y148 are located in the hydrophobic inner region of the cap domain. Residues T26 and A60 are located in the hydrophobic inner region of the core domain. Residue G30 is located in the substrate-binding site.

(c) The backbone carbonyl oxygen of Asp-29 and the side chains of Asp-27 and Asp-238 coordinate a magnesium ion to form the active site. The side chains of Asp-29, Thr-61, and Lys-
215 coordinate a chloride ion, and are correctly positioned to coordinate a phosphate group on substrate binding. These active-site residues are conserved in \textit{PfHAD1} homologs from organisms possessing the MEP pathway. Alignment was produced in T-Coffee [56] using default parameters. Accession codes for NCBI protein sequences: \textit{E. coli} YidA, EOU46719; \textit{M. tuberculosis} Rv3813c, NP_218330; \textit{C. reinhardtii} FER_156463, ADF43173; \textit{A. thaliana} At2g25870, ABO38782.
**Figure 4.** PfHAD1 FSM-resistance alleles result in loss of phosphatase activity.

Displayed are the means ± S.E.M. of enzyme activity from at least three independent experiments. P-values were determined using a one-way analysis of variance (Tukey’s post test, α=0.05) in GraphPad Prism software. ‘NS’=not significant (P>0.05). ‘No enz.’=no enzyme. We were unable to obtain soluble T26R and W130R PfHAD1 variants.
Figure 5. Loss of PfHAD1 is required for FSM resistance.

Data shown are representative of at least three independent experiments. (a) Immunoblot of the parent strain, FSM strain AM1 and AM1 Hsp110: PfHAD1-GFP. Full blots are shown in Supplementary Figure 3. Marker units are kilodaltons (kDa). Top panel was probed with anti-PfHAD1 antisera, and the bottom panel was probed with anti-heat shock protein 70 (hsp70) antisera as a loading control. Expected approximate protein masses: native PfHAD1, 33 kDa; PfHAD1-GFP, 60 kDa; hsp70, 74 kDa. (b) FSM IC_{50}s of the parent strain, FSM strain AM1, and AM1 Hsp110: PfHAD1-GFP. Displayed are the means ± S.E.M. The parent strain has a FSM IC_{50} of 1.1 ± 0.2 µM, while AM1 has an IC_{50} of 5.5 ± 1.5 µM. Expression of PfHAD1-GFP in AM1 results in an IC_{50} of 1.4 ± 0.2 µM. Data shown are mean and are normalized such that 100% growth is defined by that of untreated cells and 0% growth is defined as the smallest value in each data set.
Figure 6. PfHAD1 dephosphorylates sugar phosphates, including MEP pathway intermediates.

Displayed are the means ± S.E.M. of the enzyme activity from at least three independent experiments. DHAP, dihydroxyacetone phosphate; Ery4P, erythrose 4-phosphate; Fru1P, fructose 1-phosphate; Fru6P, fructose 6-phosphate; Fru1,6bisP, fructose 1,6-bisphosphate; Gal1P, galactose 1-phosphate; Glc2P, glycerol 2-phosphate; Gln1P, glucosamine 1-phosphate; Glu1P, glucose 1-phosphate; Glu6P, glucose 6-phosphate; Man1P, mannose 1-phosphate; Man6P, mannose 6-phosphate; PEP, phosphoenolpyruvate; Rib5P, ribose 5-phosphate; Ribu5P, ribulose 5-phosphate; Sedo7P, sedoheptulose 7-phosphate; Sorb6P, sorbitol 6-phosphate; Tre6P, trehalose 6-phosphate; 2drib5P, deoxyribose 5-phosphate; 2-PGA, 2-phosphoglyceric acid; 3-PGA, 3-phosphoglyceric acid.
Figure 7. PfHAD1 is expressed in blood-stage parasites and is localized to the parasite cytoplasm.

Immunofluorescence confocal microscopy of ACP$_L$-GFP trophozoite and schizont [17], stained with αGFP and αPfHAD1 antibodies and Hoechst 33258 nuclear stain. Scale bars, 2 μm.
Figure 8. Increased levels of MEP pathway intermediates in FSM\textsuperscript{R} strains with the loss of PfHAD1 function.

Ring-stage parental and FSM\textsuperscript{R} \textit{P. falciparum} parasites were treated +/− 5 µM FSM for 10 h. Displayed are the means ± S.E.M. of the metabolite levels from at least three independent experiments. Representative FSM\textsuperscript{R} strain E1 is shown, which possesses the K10X \textit{PfHAD1} allele. The complete dataset is found in Supplementary Table 3. CDP-ME, 4-diphosphocytidyl-2-C-methylerythritol.
Figure 9. Model of PfHAD1 function.

PfHAD1 (blue arrows) may dephosphorylate the intermediates of glycolysis. Loss of PfHAD1 may increase local levels of sugar phosphates and increase the substrate availability to the apicoplast MEP pathway (green). The use of PEP as a substrate is unlikely, given the poor PfHAD1 activity against this substrate in vitro (Figure 6). Green reactions and substrates are part of the MEP pathway. PfTPT and PfoTPT are transporters responsible for the import of PEP and DHAP into the apicoplast [57,58]. The conversion of PEP and DHAP into pyruvate and gly3P, respectively, are catalyzed by apicoplast-localized pyruvate kinase (PK) [59] and triosephosphate
isomerase (TPI) [60]. DHAP, dihydroxyacetone phosphate; DMAPP, dimethylallyl pyrophosphate; IPP, isopentenyl pyrophosphate; PEP, phosphoenolpyruvate.
Supplementary Figure 1. FSM\textsuperscript{R} strains are also resistant to the FSM-related compound, FR-900098. Displayed are the means ± S.E.M. of at least three independent experiments. The parental strain has an IC\textsubscript{50} of 0.23 ± 0.03 µM against FR-900098. FSM\textsuperscript{R} strains AM1, E1, and D6 have FR-900098 IC\textsubscript{50}s of 3.2 ± 0.24 µM, 1.7 ± 0.001 µM, and 2.8 ± 0.17 µM, respectively.
Supplementary Figure 2. Relative mRNA expression levels of *PfDXS* and *PfDXR* are unchanged in FSM\textsuperscript{R} strains. Data represent means and standard error of the mean (S.E.M.) of at least two independent experiments. “N.s” = not significant. P-values are >0.3 (unpaired Student’s *t*-test) for comparisons shown.
Supplementary Figure 3. Full immunoblots demonstrating expression of PfHAD1-GFP in FSM$^R$ strain AM1. Samples shown are the parental strain, FSM$^R$ strain AM1, and AM1 Hsp110:PfHAD1-GFP. Marker units are kilodaltons. The blot was probed with anti-PfHAD1 antisera (left), stripped, and re-probed with anti-heat shock protein 70 antisera (right). Expected approximate protein masses: native PfHAD1, 33 kDa; PfHAD1-GFP, 60 kDa; hsp70, 74 kDa. Data are representative of at least three independent experiments.
Supplementary Figure 4. PfHAD1-GFP localization is similar to that of PfHAD1 (Figure 7).

Shown is live microscopy of AM1 Hsp110:PfHAD1-GFP trophozoite and schizont, stained with Hoechst 33258 nuclear stain. Scale bars represent 2 μm.
Supplementary Figure 5. Stereo image of the electron density maps for a representative region of PfHAD1. (a) The 2Fo-Fc electron density map contoured at 1.0 σ is colored blue. (b) The composite simulated annealing omit map contoured at 1.0 σ is colored grey.
Supplementary Figure 6. Loss of PfHAD1 is necessary for increased DOXP levels in FSM<sup>R</sup> strain AM1. Shown are attograms/cell DOXP as measured by LC-MS/MS in the parental strain, FSM<sup>R</sup> strain AM1, and AM1 Hsp110:PfHAD1-GFP. Displayed are the means and S.E.M. from three independent parasite cultures.
### 2.7 TABLES

Table 1. Data collection and refinement statistics.

<table>
<thead>
<tr>
<th>TABLE 1. Data collection and refinement statistics</th>
<th>PfHAD1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P2₁</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
</tr>
<tr>
<td>a, b, c, (Å)</td>
<td>77.90, 43.80, 83.90</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>90.00, 101.10, 90.00</td>
</tr>
<tr>
<td>Resolution (Å)*</td>
<td>20-2.05 (2.15-2.05)*</td>
</tr>
<tr>
<td>R&lt;sub&gt;meas&lt;/sub&gt;</td>
<td>7.5% (72.4%)</td>
</tr>
<tr>
<td>I/σI</td>
<td>13.84 (2.31)</td>
</tr>
<tr>
<td>Total no. of reflections</td>
<td>130,290 (16,948)</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>35,166 (4,630)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.5% (99.4%)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>3.70 (3.66)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>20-2.05</td>
</tr>
<tr>
<td>No. of reflections</td>
<td>35,143</td>
</tr>
<tr>
<td>R&lt;sub&gt;work&lt;/sub&gt;/R&lt;sub&gt;free&lt;/sub&gt;</td>
<td>20.19/24.18</td>
</tr>
<tr>
<td>No. of atoms</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>4,445</td>
</tr>
<tr>
<td>Mg</td>
<td>2</td>
</tr>
<tr>
<td>Cl</td>
<td>2</td>
</tr>
<tr>
<td>Water</td>
<td>232</td>
</tr>
<tr>
<td>B-factors</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>39.90</td>
</tr>
<tr>
<td>Ligands</td>
<td>51.70</td>
</tr>
<tr>
<td>Water</td>
<td>37.30</td>
</tr>
<tr>
<td>R.m.s. deviations</td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.003</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>0.647</td>
</tr>
</tbody>
</table>

* Highest resolution shell is shown in parenthesis.
Supplementary Table 1. FSM IC$_{50}$s, PfHAD1 alleles, and NCBI database accessions for FSM$^R$ strains.

<table>
<thead>
<tr>
<th>Strain</th>
<th>FSM IC$_{50}$ (µM)</th>
<th>PfHAD1 change</th>
<th>PfHAD1 protein variant</th>
<th>Polyphen-2 score*</th>
<th>SRA accession #</th>
<th>Trace Archive TI #s#</th>
</tr>
</thead>
<tbody>
<tr>
<td>parental</td>
<td>1.1</td>
<td>none</td>
<td>wild-type</td>
<td></td>
<td>SRS561734</td>
<td>2338198571-75</td>
</tr>
<tr>
<td>E1</td>
<td>8.1</td>
<td>A28T</td>
<td>K10X</td>
<td>n/a*</td>
<td>SRS561733</td>
<td>2338198557-60</td>
</tr>
<tr>
<td>B4</td>
<td>2.8</td>
<td>A28T</td>
<td>K10X</td>
<td>n/a</td>
<td>SRS561730</td>
<td>2338198501-44</td>
</tr>
<tr>
<td>A1</td>
<td>6.1</td>
<td>C77G</td>
<td>T26R</td>
<td>0.999</td>
<td>SRS561731</td>
<td>2338198460-63</td>
</tr>
<tr>
<td>D6</td>
<td>8.7</td>
<td>G89A</td>
<td>G30E</td>
<td>1.000</td>
<td>SRS561730</td>
<td>2338198553-56</td>
</tr>
<tr>
<td>A3</td>
<td>4.6</td>
<td>ΔA112</td>
<td>Frameshift, results in L47X</td>
<td>n/a</td>
<td>2338198468-71</td>
<td></td>
</tr>
<tr>
<td>D2</td>
<td>1.5</td>
<td>C179A</td>
<td>A60E</td>
<td>1.000</td>
<td>2338198540-43</td>
<td></td>
</tr>
<tr>
<td>A2</td>
<td>4.6</td>
<td>T388C</td>
<td>W130R</td>
<td>0.941</td>
<td>2338198464-67</td>
<td></td>
</tr>
<tr>
<td>A4</td>
<td>5.8</td>
<td>A403 insertion (+A)</td>
<td>Frameshift, results in E140X</td>
<td>n/a</td>
<td>2338198472-75</td>
<td></td>
</tr>
<tr>
<td>MK1</td>
<td>3.7</td>
<td>A443G</td>
<td>Y148C</td>
<td>0.747</td>
<td>SRS561761</td>
<td>2338198569-70</td>
</tr>
<tr>
<td>AM1</td>
<td>5.5</td>
<td>T623A</td>
<td>L208X</td>
<td>n/a</td>
<td>SRS561775</td>
<td>2338198484-87</td>
</tr>
<tr>
<td>AM2</td>
<td>3.3</td>
<td>T623A</td>
<td>L208X</td>
<td>n/a</td>
<td>SRS561776</td>
<td>2338198488-91</td>
</tr>
<tr>
<td>B1</td>
<td>5.9</td>
<td>Δ681-685</td>
<td>Frameshift, results in E240X</td>
<td>n/a</td>
<td>SRS561732</td>
<td>2338198492-95</td>
</tr>
<tr>
<td>F1</td>
<td>3.7</td>
<td>Δ681-685</td>
<td>Frameshift, results in E240X</td>
<td>n/a</td>
<td>2338198565-68</td>
<td></td>
</tr>
<tr>
<td>D3</td>
<td>4.3</td>
<td>Δ681-685</td>
<td>Frameshift, results in E240X</td>
<td>n/a</td>
<td>2338198544-47</td>
<td></td>
</tr>
<tr>
<td>A5</td>
<td>5.4</td>
<td>ΔC698</td>
<td>Frameshift, results in L234X</td>
<td>n/a</td>
<td>2338198476-79</td>
<td></td>
</tr>
<tr>
<td>D4</td>
<td>2.0</td>
<td>none</td>
<td>wild-type</td>
<td></td>
<td>2338198548-52</td>
<td></td>
</tr>
<tr>
<td>A6</td>
<td>4.1</td>
<td>none</td>
<td>whole gene deletion</td>
<td>n/a</td>
<td>SRS561778</td>
<td>2338198480-83</td>
</tr>
<tr>
<td>E2</td>
<td>4.6</td>
<td>none</td>
<td>wild-type</td>
<td></td>
<td>2338198561-64</td>
<td></td>
</tr>
<tr>
<td>B2</td>
<td>5.0</td>
<td>none</td>
<td>M1I</td>
<td>n/a</td>
<td>2338198496-500</td>
<td></td>
</tr>
<tr>
<td>D1</td>
<td>4.7</td>
<td>none</td>
<td>wild-type</td>
<td></td>
<td>SRS561777</td>
<td>2338198536-39</td>
</tr>
</tbody>
</table>

* = Polyphen-2 scores cannot be calculated for truncation mutations. Polyphen-2 is an algorithm for predicting the probability of deleterious effects of missense mutations [34].

$^\$ = Whole genome sequencing data is deposited in the NCBI BioProject and Sequence Read Archive databases. Representative Sanger sequencing data of the PfHAD1 locus is deposited in the NCBI Trace Archive.

# = mutations discovered after the publication of this manuscript.
Supplementary Table 2. Kinetic parameters for PfHAD1 with the top three tested substrates and gly3P. Shown are means ± S.E.M. of at least three independent experiments.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>$K_m$ (mM)</th>
<th>$k_{cat}$ (sec$^{-1}$)</th>
<th>$k_{cat}/K_m$ (M$^{-1}$ sec$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>fru1P</td>
<td>2.5 ± 0.6</td>
<td>5.2 ± 0.9</td>
<td>2.1 x 10$^3$</td>
</tr>
<tr>
<td>glc2P</td>
<td>6.5 ± 0.8</td>
<td>8.5 ± 0.2</td>
<td>1.3 x 10$^3$</td>
</tr>
<tr>
<td>man6P</td>
<td>3.6 ± 0.8</td>
<td>3.1 ± 0.3</td>
<td>0.9 x 10$^3$</td>
</tr>
<tr>
<td>gly3P</td>
<td>4.6 ± 0.7</td>
<td>11 ± 4.3</td>
<td>2.4 x 10$^3$</td>
</tr>
</tbody>
</table>
Supplementary Table 3. FSM\textsuperscript{R} strains possessed increased levels of MEP pathway metabolites. Shown are concentrations of MEP pathway intermediates in \textit{P. falciparum} FSM\textsuperscript{R} PfHAD1 strains. Data shown are means ± S.E.M. of at least three independent experiments.

<table>
<thead>
<tr>
<th>Compound (attograms/cell)</th>
<th>DOXP</th>
<th>MEP</th>
<th>CDP-ME</th>
<th>MEcPP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strain</strong></td>
<td>- FSM</td>
<td>+ FSM</td>
<td>- FSM</td>
<td>+ FSM</td>
</tr>
<tr>
<td>parental</td>
<td>0.20 ± 0.01</td>
<td>0.25 ± 0.03</td>
<td>1.03 ± 0.36</td>
<td>1.03 ± 0.38</td>
</tr>
<tr>
<td>B4</td>
<td>6.58 ± 3.14</td>
<td>18.73 ± 1.02</td>
<td>0.87 ± 0.32</td>
<td>0.27 ± 0.02</td>
</tr>
<tr>
<td>E1</td>
<td>10.06 ± 1.41</td>
<td>17.26 ± 2.18</td>
<td>3.97 ± 0.29</td>
<td>2.94 ± 0.28</td>
</tr>
<tr>
<td>A1</td>
<td>17.12 ± 2.52</td>
<td>33.39 ± 2.99</td>
<td>2.77 ± 0.30</td>
<td>0.75 ± 0.10</td>
</tr>
<tr>
<td>D6</td>
<td>8.50 ± 1.35</td>
<td>13.98 ± 2.79</td>
<td>4.19 ± 0.53</td>
<td>3.05 ± 0.42</td>
</tr>
<tr>
<td>A3</td>
<td>10.03 ± 0.86</td>
<td>14.86 ± 2.64</td>
<td>1.51 ± 0.10</td>
<td>0.33 ± 0.10</td>
</tr>
<tr>
<td>A4</td>
<td>11.03 ± 1.54</td>
<td>14.44 ± 5.18</td>
<td>2.02 ± 0.27</td>
<td>0.42 ± 0.18</td>
</tr>
<tr>
<td>AM1</td>
<td>14.71 ± 1.28</td>
<td>28.2 ± 3.61</td>
<td>1.65 ± 0.14</td>
<td>0.38 ± 0.02</td>
</tr>
<tr>
<td>A5</td>
<td>10.31 ± 1.24</td>
<td>16.63 ± 3.28</td>
<td>1.10 ± 0.12</td>
<td>0.14 ± 0.02</td>
</tr>
</tbody>
</table>
2.8 REFERENCES


17. Waller RF, Reed MB, Cowman AF, McFadden GI. Protein trafficking to the plastid of *Plasmodium falciparum* is via the secretory pathway. EMBO J. 2000; 19:1794–802.


27. Balu B, Shoue DA, Fraser MJ, Adams JH. High-efficiency transformation of *Plasmodium*


52. Brown AC, Parish T. Dxr is essential in Mycobacterium tuberculosis and fosmidomycin resistance is due to a lack of uptake. BMC Microbiol. 2008; 8:78.


Chapter 3

Whole genome sequencing to evaluate the resistance landscape following antimalarial treatment failure with fosmidomycin-clindamycin
PREFACE

The following work was performed by me, Sesh A. Sundararaman, Miguel Lanaspa, Cinta Moraleda, Raquel González, Alfredo Mayor, Pau Cisteró, David Hutchinson, Peter G. Kremsner, Beatrice H. Hahn, Quique Bassat, and Audrey R. Odom. I performed and conceived the majority of experiments and wrote the manuscript. S.A.S. and B.H.H. provided reagents, protocols and helpful feedback. L.M., C.M., R.G., A.M., P.C., H.D., P.G.K., and Q.B designed and performed the original clinical trial in Mozambique and provided helpful feedback. A.R.O. conceived and supervised the study and wrote the manuscript.

This chapter is published in its entirety and the reproduction of this article is permitted as part of the copyright agreement with Oxford University Press. This chapter is a reproduction of a pre-copyedited, author-produced PDF of an article accepted for publication in The Journal of Infectious Diseases following peer review. The version of record [Guggisberg AM, Sundararaman SA, Lanaspa M, Moraleda C, González R, Mayor A, Cisteró P, Hutchinson D, Kremsner PG, Hahn BH, Bassat Q, Odom AR. (2016). Whole genome sequencing to evaluate the resistance landscape following antimalarial treatment failure with fosmidomycin-clindamycin. Journal of Infectious Diseases, online ahead of print: doi: 10.1093/infdis/jiw304] is available online at: [http://jid.oxfordjournals.org/content/early/2016/07/20/infdis.jiw304.long].

We thank the study participants at the Centro de Investigação em Saúde de Manhiça (CISM) in Manhiça, Mozambique for allowing us to analyze their parasite isolates.
We thank Philip Ruzycki (laboratory of Dr. Shiming Chen, Washington University) for supplying human-specific primers and for helpful discussion regarding the analysis. We also thank Andrew Jezewski (Odom laboratory, Washington University) for helpful discussion regarding the analyses.

We thank Wei (Will) Yang at the Genome Technology Access Center in the Department of Genetics at Washington University School of Medicine for assistance with the analyses. This work is supported by the Children’s Discovery Institute of Washington University and St. Louis Children’s Hospital [MD-LI-2011-171 to A.R.O.], the National Institute of Allergy and Infectious Diseases at the National Institutes of Health [R01AI103280 to A.R.O., R21AI1123808 to A.R.O., R01AI058715 to B.H.H., R01AI091595 to B.H.H., R37AI050529 to B.H.H., and T32AI007532 to S.A.S.], the March of Dimes [Basil O’Connor Starter Scholar Research Award to A.R.O.], the National Institute of General Medical Sciences at the National Institutes of Health [T32GM007067 to A.M.G.], the Washington University Monsanto Excellence Fund [graduate fellowship to A.M.G.], and the program Miguel Servet of the ISCIII [Plan Nacional de I+D+I 2008-2011, grant #CP11/00269 to Q.B.].

The Genome Technology Access Center (GTAC) at Washington University is partially supported by the National Cancer Institute at the National Institutes of Health [P30CA91842], the National Institutes of Health National Center for Research Resources [ICTS/CTSA grant UL1TR000448], and NIH Roadmap for Medical Research.
3.1 ABSTRACT

Novel antimalarial therapies are needed in the face of emerging resistance to artemisinin combination therapies. A previous study found a high cure rate in Mozambican children with uncomplicated *Plasmodium falciparum* malaria 7 days post treatment with a fosmidomycin-clindamycin combination. However, 28-day cure rates were low (45.9%), due to parasite recrudescence. We sought to identify any genetic changes underlying parasite recrudescence. To this end, we utilized a selective whole genome amplification method to amplify parasite genomes from blood spot DNA samples. Parasite genomes from pre-treatment and post-recrudescence samples were subjected to whole genome sequencing to identify nucleotide variants. We find that our data do not support the existence of a genetic change responsible for recrudescence following fosmidomycin-clindamycin treatment. Additionally, we find that previously described resistance alleles for these drugs do not represent biomarkers of recrudescence. Future studies should continue to optimize fosmidomycin combinations for use as antimalarial therapies.

3.2 INTRODUCTION

Malaria remains a serious global health concern, with ~214 million cases and 483,000 deaths due to malaria in 2015 [1]. A majority of severe malaria cases occur in pregnant women and children under five years of age and result from infection with the parasite *Plasmodium falciparum* [1]. Artemisinin combination therapies (ACTs) represent the current first-line treatment in endemic areas. Historically, the development of drug resistance has hindered malaria control. As delayed
parasite clearance has emerged for ACTs in Southeast Asia and has continued to spread [1,2], new antimalarials are urgently needed.

One such potential therapeutic is the phosphonic acid antibiotic fosmidomycin (FSM). FSM is a well-characterized inhibitor of the first committed step of isoprenoid precursor synthesis via the methylerythritol phosphate (MEP) pathway [3–5]. Because humans synthesize isoprenoids via an enzymatically distinct metabolic route (mevalonate pathway), FSM is a highly specific inhibitor of malaria parasite growth [3,5] and is well-tolerated in humans [6,7].

FSM was originally developed as an antibacterial [8] and is currently under evaluation as a partner agent for combination therapy against uncomplicated \( P. falciparum \) malaria (clinicaltrials.gov identifier NCT02198807). Several studies have paired FSM with the antibiotic clindamycin (CLN), a protein translation inhibitor (clinical trials NCT02198807, NCT01361269, NCT01002183, NCT00214643, and NCT00217451) [6,9,10]. One such Phase II clinical trial, performed in Mozambique in 2010, evaluated the efficacy of a FSM-CLN combination on uncomplicated malaria in children aged 6-35 months, the youngest cohort tested to date [11,12]. A high parasite cure rate was observed at day 7 post-treatment with FSM-CLN (94.6%). Unfortunately, the cure was not durable, and the post-treatment PCR-corrected day 28 cure rate was 45.9% (17/39), due to recrudescence of infections. Parasite recrudescence is common in clinical studies of FSM efficacy, with an overall day 28 cure rate of ~70% in adults [12].

The safety and specificity of MEP pathway-targeting therapeutics, such as FSM, are highly desirable. However, the failure of FSM treatment to result in lasting cure has led to concern
regarding the clinical utility of FSM or related antimalarials. We recently found that FSM resistance is readily achieved in culture due to mutations in the metabolic regulator PfHAD1 (PlasmoDB ID PF3D7_1033400) [13]. Additionally, CLN resistance has been reported in clinical isolates of *P. falciparum*, attributed to mutation in the apicoplast 23S rRNA [14].

In this study, we address whether observed recrudescence is produced by selection for mutations in previously identified candidate genes or novel resistance loci. Using selective whole genome amplification (SWGA) and sequencing, we characterize *P. falciparum* field isolates from blood spot samples to evaluate the genetic landscape of drug resistance before and after FSM-CLN treatment. Specifically, we evaluate for enrichment in genetic changes associated with decreased susceptibility to either FSM or CLN.

### 3.3 METHODS

#### 3.3.1 Study information

The study criteria have been previously described [11]. The study evaluated 37 children ages 6-35 months with uncomplicated malaria. Inclusion and exclusion criteria are outlined in the original study [11]. Patients were administered a three-day, twice-daily course of oral FSM-CLN. Blood spots were collected at days 0, 7, 14, and 28, and if applicable, upon recrudescence.
3.3.2 Clinical trial information

The original trial in Mozambique, sponsored by Jomaa Pharma, was conducted in 2010 according to the ICH Good Clinical Practice guidelines. The protocol was approved by the National Mozambican Ethics Review Committee and the Hospital Clinic of Barcelona Ethics Review Committee. The clinicaltrials.gov identifier is NCT01464138.

3.3.3 Selective whole genome amplification

Blood spotting and DNA isolation has been previously described [11]. SWGA of *P. falciparum* genomes was performed as described previously [15], including primer sets (Supplementary Table 2) and reaction conditions. Reactions contained 10 ng template and 3.5 µM primer set 6A. Reactions were cleaned using Ampure beads (Beckmann Coulter) at a 1:1 DNA:bead ratio. Samples underwent a second round of SWGA using primer set 8A and 15-30 ng DNA.

3.3.4 PCR of SWGA products

Selective amplification of *P. falciparum* genomes was verified by PCR using *P. falciparum*- or human-specific primers (Supplementary Table 1). Human DNA and human-specific primers were kindly provided by Shiming Chen (Washington University). Reactions contained 2 µM each primer, 25-50 ng bead-cleaned SWGA product, and BIO-X-ACT Short PCR Mix (Bioline). Vendor-recommended cycling conditions were used, with modifications: (55 °C and 60 °C annealing for *P. falciparum*- and human-specific primers, respectively; 68 °C extension). Amplicons were visualized by agarose gel electrophoresis.
3.3.5 Whole genome sequencing

Library preparation, Illumina sequencing, read alignments, and variant calling were performed by the Washington University Genome Technology Access Center (GTAC). Bead-cleaned SWGA product DNA (0.5-1.2 µg) was sheared, end-repaired, and adapter-ligated. PCR-based libraries were sequenced on an Illumina HiSeq 2500 to generate 101 bp paired end reads. Reads were aligned to the 3D7 reference (PlasmoDB v24) using Novoalign (V2.08.02) [16]. Duplicate reads were removed. SNPs were called using samtools (mpileup) [17], filtered (quality ≥20, read depth ≥5), and annotated using snpEff (3.3c, build 2013-06-28) [18].

*P. falciparum* reads are available through the NCBI BioProject database (PRJNA315887) and Sequence Read Archive (SRP072442).

For some analyses (comparison of SNPs in resistance genes, GO analysis), only sample pairs with both exomes showing sufficient coverage (≥60% covered at ≥5X, indicated in Supplementary Table 3) were used. Given a per-base error rate of 0.1-0.5% for Illumina sequencing [19,20], 5X coverage equates to >99.9% accuracy. Studies have indicated reasonable sensitivity and accuracy for our variant caller at this cutoff [21].

3.3.6 SNPhylo plot generation

Multi-sample VCFs were converted to hapmap format. SNPhylo [22] was run with the default settings, with an LD cutoff of 0.8.
3.3.7 Multiplicity of infection (MOI) determination

MOI was determined using the WHO-recommended PCR-based genotyping procedures at the *MSP1* (PF3D7_0930300), *MSP2* (PF3D7_0206800), and *GLURP* (PF3D7_1035300) loci [23].

3.3.8 Sequencing of the apicoplast 23S rRNA locus

The apicoplast 23S rRNA (PlasmoDB ID PCF10_API0010:rRNA) locus was amplified by PCR. Reactions contained 2 µM of each primer 23S_1 and 23S_6 (Supplementary Table 1), 4-5 µL of blood spot DNA, and CloneAmp HiFi PCR Premix (Clontech). Vendor-recommended cycling conditions were used (55 °C primer annealing). Amplicons were Sanger sequenced using primers 23S_1 – 23S_6 (Supplementary Table 1).

3.3.9 GO analysis

GO term enrichment of genes containing SNPs unique to post-treatment samples was determined using the PlasmoDB Gene Ontology Enrichment tool [24] (Bonferroni-corrected P-value <0.01).

3.4 RESULTS

3.4.10 Selective whole genome amplification of blood spot DNA generates parasite templates for whole genome sequencing

We evaluated blood spot DNA (mixed human and *P. falciparum*) from 12 patient samples with microscopic recrudescent infection (12 pre- and post-treatment pairs, 24 samples total). Pilot unmodified, low-input library preparation methods resulted in <5% reads mapping to the *P.*
P. falciparum genome. For this reason, we used SWGA to amplify P. falciparum DNA. This method uses the processive Φ29 DNA polymerase and genome-specific primers to selectively amplify a target genome from a mixed sample [25–27] and has been recently used to characterize chimpanzee Plasmodium genomes [15].

Enrichment of the P. falciparum genome was verified by PCR (Figure 1). We amplified P. falciparum genomes from a broad range of parasite densities (18-315,064 parasites/μL whole blood). Amplified samples were used to prepare libraries for Illumina sequencing.

The samples displayed varying degrees of sequencing success (Supplementary Table 3). An average (± S.E.M.) of 55.1% ± 2.9% of reads mapped to the P. falciparum genome, resulting in 48.1% ± 3.5% of the genome covered at ≥5X. This is comparable to previous studies using SWGA to sequence microbial genomes [27,28]. The genome coverage we obtain is consistent with very low proportions (<0.01%) of parasite DNA in the blood spot samples [15]. While more reads will increase genome coverage, including typically low-coverage intragenic regions, coverage is likely to be limited by incomplete genome representation in the sample, as very low proportions of parasite DNA are likely to result in incomplete or inefficient SWGA of some genome regions. Given the low coverage of AT-rich intragenic regions, our analyses focused on protein-coding regions of the P. falciparum genome. We observe an average of 69.0% ± 4.4% of the exome covered ≥5X, with as much as 90% of the exome covered ≥5X (Supplementary Table 3).
We observe an average of $35,451 \pm 1,220$ total genome SNPs in our sequenced samples, consistent with recent studies of African field isolates [29,30]. Of these, $15,862 \pm 534$ are exome SNPs. We find an average of $10,755 \pm 349$ non-synonymous SNPs in our samples with sufficient exome coverage ($\geq 60\%$ covered at $\geq 5X$).

3.4.11 SNP profiling confirms recrudescent infections

Maximum-likelihood phylogenetic tree construction from our exome SNP profiles from the 12 pre- and post-treatment sample pairs demonstrates clustering by patient-of-origin (Figure 2). These data support the conclusion that all infections analyzed indeed reflect failure to completely clear the original infection (recrudescence), as opposed to novel, independent infections. Pre-treatment and post-treatment samples share a majority ($66.0 \pm 3.4\%$) of non-synonymous SNPs, while independent (between-patient) pre-treatment infections are less related and share only $24.6 \pm 0.4\%$ of non-synonymous SNPs (strains with $\geq 60\%$ of the exome at $\geq 5X$, N=8). These data confirm the genotyping reported in the original study [11], which established recrudescence through WHO-recommended PCR-based genotyping at three *P. falciparum* loci: *MSP1*, *MSP2*, and *GLURP* [23].

3.4.12 Multiplicity-of-infection (MOI) pre- and post-treatment

In high transmission areas, patients may be simultaneously infected with multiple parasite strains. This genetic variation provides a potential reservoir for the development of resistance to antimalarials, both within a given patient and in the larger *P. falciparum* population. Multiplicity-of-infection (MOI) has been shown to decrease after recrudescence following chloroquine treatment [31], and, unsurprisingly, increased MOI is correlated with treatment
failure [32]. However, during any individual infection, selective pressures during infection (such as immune evasion) and drug treatment are expected to decrease genetic diversity. We therefore assessed MOI before and after recrudescence.

All infections were polyclonal, with a mean pre-treatment MOI of 3.9 ± 0.3. This is slightly higher than MOIs reported in other pediatric studies in Mozambique [33,34]. MOI decreased modestly in two-thirds of the samples after treatment and recrudescence, with a mean post-treatment MOI of 3.2 ± 0.2 (p=0.0065, paired t-test). This reduction was not a result of lower parasitemias in recrudescent infections, as average post-treatment parasite density was approximately equal to average pre-treatment parasite density (76,225 ± 33,944 vs. 70,977 ± 20,498 parasites/µL). Additionally, MOI and parasite density were not correlated (Pearson r = -0.041, p = 0.849).

Of note, loss-of-function mutations in the FSM-resistance gene PfHAD1 have been infrequently reported in the genomes of field P. falciparum isolates (PlasmoDB, accessed April 2016), suggesting that such alleles are rare [24]. Therefore, if resistance emerges during the course of infection in a given patient, recrudescence due to outgrowth of a rare resistant clone should result in a dramatic decrease in MOI. This was not observed in our population, suggesting that FSM resistance alone is unlikely to account for recrudescence of patients treated with FSM-CLN.
3.4.13 Genotyping of known resistance loci does not reveal a change in the resistance landscape between pre- and post-treatment samples

To evaluate whether FSM-CLN recrudescent parasites have genetic changes associated with drug resistance, we investigated the genomes of paired pre- and post-treatment parasites for genetic markers of antimalarial resistance. We evaluated loci previously implicated in resistance to FSM or CLN, as well as loci documented to contribute to resistance to other clinically available antimalarials.

FSM targets the isoprenoid synthesis enzyme 1-deoxy-D-xylulose 5-phosphate reductoisomerase (PfDXR, PF3D7_1467300). Following FSM-CLN treatment, recrudescent samples are not enriched in alternative PfDXR alleles (Figure 3, Supplementary Table 4). One sample (114) possesses the Q361R PfDXR allele, previously reported in African isolates [24]. The homologous residue (R276) in FSM-sensitive *Escherichia coli* DXR (NCBI reference NP_414715.1) is identical to the mutated residue (361R). Therefore, the DXR Q361R variant is unlikely to confer FSM resistance.

We have recently identified a metabolic regulator, PfHAD1, whose loss results in FSM resistance in cultured *P. falciparum* [13]. No PfHAD1 loss-of-function alleles were identified in our samples (Figure 3, Supplementary Table 4), consistent with a known lack of variation in this gene in field samples [24]. The N70S allele seen in a number of strains has been previously reported as a common wild-type variant [24], which does not confer FSM resistance [13]. Two pre-treatment samples (103 and 118) possessed additional polymorphisms in PfHAD1, S265L and E150K, respectively (Figure 3, Supplementary Table 4). These alleles are not predicted to
affection of PfHAD1 function [Polyphen-2 scores < 0.005 (benign)] [35] and are not enriched in the recrudescent samples. Altogether, our data indicate that known markers of FSM resistance in PfDXR and PfHAD1 do not emerge during the course of FSM treatment. Thus, known mechanisms of FSM resistance observed in the laboratory are unlikely to represent the underlying mechanism behind FSM-CLN failure. Since recrudescent parasites were not culture-adapted, we cannot distinguish whether these parasites were resistant or just failed to be cleared. Therefore, it is possible that resistance-causing mutations in PfDXR, PfHAD1, or other loci might occur if parasites were subjected to longer drug exposures.

Resistance to the FSM partner agent, CLN, has also been reported in P. falciparum [14]. In this previous study, mutation (A1875C) in the 23S rRNA locus (PFC10_API0010) is associated with an approximately 100-fold increase in clindamycin IC_{50} for clinical isolates [14]. Insufficient coverage of the plastid genome was obtained from whole genome sequencing; we therefore used targeted gene amplification and sequencing to interrogate for polymorphisms in PFC10_API10010. No samples possessed the A1875C variant. All sequenced samples matched the FSM- and CLN-sensitive 3D7 control strain (Supplementary Data 1, available online [36]). Notably, from our sequencing of laboratory 3D7 and HB3 strains, we identify variations not reported in the genome reference (T451A, A454T, and 401insG) (Supplementary Data 1, available online [36]) [24]. These changes may have emerged over time during culturing of laboratory isolates.

Finally, we interrogated the prevalence of other drug resistance markers before and after FSM-CLN treatment (Figure 3, Supplementary Table 4). Specifically, we evaluated variation in the
locus encoding the multi-drug transporter PfMDR1 (PF3D7_0523000), which modulates parasite sensitivity to hydrophobic antimalarials, such as mefloquine and halofantrine [37]. Strains possessing PfMDR1 mutations are sensitive to FSM and PfMDR1 mutations are not predicted to impact FSM effectiveness [3,9]. Three PfMDR1 variants (wild-type, Y184F, and N86Y) were identified in our analyzed strains, but PfMDR1 haplotype frequencies were not significantly different in pre- and post-treatment populations (p>0.5).

ACTs, introduced in Mozambique in 2004 [38], may have selected for parasites able to withstand drug treatment such as FSM-CLN. Alleles of PfK13 (Kelch13, PF3D7_1343700) have been implicated in artemisinin resistance. We identify two PfK13 variants in our population (K189T and A578S), neither of which has been implicated in laboratory artemisinin resistance. The K189T variant is common in African isolates [39,40] and is not believed to cause increased clearance times following ACT. While the A578S variant has been associated with increased clearance times [41], this mutation was only observed in a small fraction of our strains. Additionally, strains possessing known resistance mutations in PfK13 are FSM sensitive (Edwards RL et al., submitted) suggesting that selection for artemisinin resistance does not result in FSM resistance.

In our small study population, selection with FSM-CLN did not appear to alter frequency of PfMDR1 alleles or alleles of additional known genetic loci associated with antimalarial resistance, including PfCRT (chloroquine), PfATP4 (multiple drug classes), PfDHFR (antifolates), and PfDHPS (antifolates) (Figure 3, Supplementary Table 4) [42–46].
Our approach permitted an unbiased search for any novel non-synonymous SNPs that are associated with recrudescence following FSM-CLN treatment. To better understand SNPs that were unique to or enriched in recrudescent samples, we also subtracted pre-treatment non-synonymous SNPs from post-treatment non-synonymous SNPs. The eight recrudescent strains analyzed had an average of 3,448 ± 604 unique non-synonymous SNPs (approximately 33% of their total non-synonymous SNPs).

Sixty-eight SNPs were shared in ≥50% of the samples (Supplementary Table 5). However, no non-synonymous SNPs were shared between all 8 recrudescent samples, demonstrating that, in this small population, a genetic marker of recrudescence was not present.

Selective pressures in vivo are likely to be distinct from those described in vitro. We hypothesized that certain biological processes may be enriched for genetic variation in our post-treatment samples. We therefore performed gene ontology (GO) analysis on the genes with SNPs shared in ≥50% of recrudescent samples (Supplementary Table 5) to understand the functions of genes containing SNPs enriched upon recrudescence. Our analysis reveals enrichment for immune evasion and parasitism-related functions (Table 1). This result has been observed in other studies of variation in *P. falciparum* populations [14,30]. Because these genes are among the most variable in a population, they are likely to display changes in allele frequency following a population bottleneck, such as recrudescence. Notably, GO analysis did not reveal enrichment in pathways associated with drug resistance or with the mechanism of action of FSM (isoprenoid synthesis) or CLN (protein translation). While a novel genetic route to FSM or CLN resistance is
possible, we see no evidence for enrichment of new SNPs or pathways in our unbiased genome analysis.

### 3.5 DISCUSSION

Fosmidomycin (FSM) is an antimalarial with a novel, parasite-specific mechanism-of-action, a well-characterized target, and exceptional clinical safety. Despite this promise, FSM combination treatment of uncomplicated *P. falciparum* infection in children was unsuccessful due to unacceptably high rates of parasite recrudescence [11]. Parasite FSM resistance arises readily in culture and has been attributed to loss-of-function mutations in *PfHAD1* [13]. These in vitro studies suggested that selection for FSM resistance alleles during clinical infection and/or FSM treatment may represent a mechanism to explain clinical failures following FSM treatment.

To address this concern, we have surveyed the genetic diversity in Mozambican pediatric *P. falciparum* malaria infections before and after treatment failure with FSM-CLN. Overall, our data indicate that drug resistance does not account for treatment failures following FSM-CLN therapy. Our results confirm that treatment and recrudescence represent a population bottleneck, as MOI is decreased in recrudescent samples. Since resistance alleles are thought to represent only a miniscule proportion of the pre-treatment population, the modest decrease in MOI that we observe is inconsistent with the selection of resistant strains from the population.

Importantly, we do not find evidence of SNPs enriched in parasites following FSM-CLN treatment. Specifically, we do not identify enrichment for alleles already experimentally
implicated in FSM or CLN resistance. Our use of whole genome sequencing permits an unbiased screen for additional SNPs that may contribute to resistance, regardless of whether these alleles are directly responsible or otherwise associated with a recrudescent phenotype. While we are limited by the retrospective nature of our study and our inability to phenotype culture-adapted recrudescent parasites, both FSM and CLN have single, well-characterized targets and known SNPs underlying resistance [13,14]. We therefore conclude that neither FSM nor CLN resistance is responsible for clinical failure of FSM-CLN.

Our study was designed to evaluate the hypothesis that a simple coding mutation may underlie recrudescence in FSM-CLN-treated parasites. The results of our study cannot exclude other potential routes to resistance, such as non-coding mutations resulting in regulatory variation or non-genetic changes in gene expression or homeostatic responses. Further studies may address if and how these mechanisms contribute to resistance to FSM, CLN, and other antimalarials.

Our study highlights an important caution in applying the results of forward genetic screening in cultured parasites to clinical populations. As resistance alleles are identified in vitro, it is important to recognize that selective pressures during natural human infection (immune pressure, metabolic requirements) are likely to be distinct. Our data indicate that mutation in PfHAD1 is not readily achieved in clinical populations. Perhaps mutation of PfHAD1 comes at a fitness cost in P. falciparum, similar to what has been found for other resistance loci, such as PfCRT and PfATP4 [47,48]. However, PfHAD1 mutation appears to reduce fitness during human infection and not during culture, as loss of PfHAD1 is easily achieved in laboratory selections.
Finally, our study illustrates the utility of SWGA for the analysis of *P. falciparum* genomes from blood spot samples. This method has applications for future field studies, as blood spots are easier to acquire than whole blood. Further optimization will facilitate the extraction of more data from these types of samples. Furthermore, our data provide additional validation of PCR-based strategies to determine MOI and recrudescence. We find that the high rate of recrudescence following FSM-CLN treatment in children was not overestimated due to use of the standard 3-locus PCR genotyping protocol [11]. Eventual use of whole genome sequencing for genotyping field populations will provide more information regarding variation within a geographic region and within patients.

This work supports the current hypothesis that the disappointing clinical efficacy of FSM combinations is likely due to challenges of partner drug selection and formulation [12]. The short serum half-life of both FSM and CLN (1-3 h) limits parasite exposure and presumably reduces the selective pressure for resistance [8,49]. However, this limited serum exposure almost certainly contributes to decreased clinical efficacy. Currently, FSM is being evaluated in combination with the bisquinolone piperaquine in a Phase II clinical trial in Gabon (NCT02198807). Piperaquine has a notably long half-life (>20 days) [50], which holds promise to limit recrudescence when paired with FSM. Our findings support the continued development of antimalarials targeting PfDXR and the MEP pathway, as well as the development of alternative FSM combinations.
3.6 FIGURES

Figure 1. SWGA causes enrichment of *P. falciparum* DNA from mixed samples that can be detected by PCR.

Shown is an agarose gel of SWGA amplicons obtained using *P. falciparum*-specific (top) and human-specific (bottom) primers. Markers indicate DNA size in base pairs.
Figure 2. Samples cluster by patient, indicating recrudescent infections.

SNPhylo [22] was used to construct a maximum likelihood phylogenetic tree of SNP profiles from initial (pre-treatment) and recrudescent (post-treatment) infections. Bootstrap values were 100 for all pre- and post-treatment branch points. Samples are numbered by infection. Scale bar represents units of substitution.
Figure 3. Resistance landscape before and after recrudescence.

Only infection pairs with sufficient exome coverage were analyzed (≥60% at ≥5X, N=8). Shown are the percentages of pre- and post-treatment samples with indicated non-reference (3D7) alleles at resistance loci. Any codons not shown match the 3D7 reference in all samples.
## 3.7 TABLES

Table 1. Gene ontology (GO) analysis of genes containing SNPs unique to recrudescent samples.

<table>
<thead>
<tr>
<th>GO term ID</th>
<th>Description</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0051704</td>
<td>multi-organism process</td>
<td>1.5 x 10^{-4}</td>
</tr>
<tr>
<td></td>
<td>symbiosis, encompassing mutualism through parasitism</td>
<td></td>
</tr>
<tr>
<td>GO:0044403</td>
<td></td>
<td>1 x 10^{-3}</td>
</tr>
<tr>
<td>GO:0044419</td>
<td>interspecies interaction between organisms</td>
<td>1 x 10^{-3}</td>
</tr>
<tr>
<td>GO:0016337</td>
<td>cell-cell adhesion</td>
<td>3.6 x 10^{-3}</td>
</tr>
<tr>
<td>GO:0009405</td>
<td>pathogenesis</td>
<td>4.6 x 10^{-3}</td>
</tr>
</tbody>
</table>
## Supplementary Table 1. PCR primers used in this study.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer sequence (5’ -&gt; 3’)</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>23S_1</td>
<td>GATGTCTATAATTTTTATTAATAAGAAAGACG</td>
<td>Apicoplast 23S rRNA amplification, Sanger sequencing</td>
</tr>
<tr>
<td>23S_2</td>
<td>AGGTTAAATATATTTAATAATGAAGCCA</td>
<td>Apicoplast 23S rRNA Sanger sequencing</td>
</tr>
<tr>
<td>23S_3</td>
<td>AGAGTGCGTAAAAGCTCATTAATTTAATAA</td>
<td>Apicoplast 23S rRNA Sanger sequencing</td>
</tr>
<tr>
<td>23S_4</td>
<td>GCGACTGTTAACAAAACACAAATC</td>
<td>Apicoplast 23S rRNA Sanger sequencing</td>
</tr>
<tr>
<td>23S_5</td>
<td>CGGAGGAGTACAATGTTATATTTATT</td>
<td>Apicoplast 23S rRNA Sanger sequencing</td>
</tr>
<tr>
<td>23S_6</td>
<td>ATTCAACTTATTAGGAATTATACACTAATA</td>
<td>Apicoplast 23S rRNA amplification, Sanger sequencing</td>
</tr>
<tr>
<td>Pfal_F</td>
<td>GGTTTCAGAACAGAATTTGGAGAGT</td>
<td><em>P. falciparum</em>-specific amplification of SWGA products</td>
</tr>
<tr>
<td>Pfal_R</td>
<td>TCCTATCTATCACCTTGAACCTTT</td>
<td><em>P. falciparum</em>-specific amplification of SWGA products</td>
</tr>
<tr>
<td>Human_F</td>
<td>CTCCTGTAGTCCCAGCTACTTG</td>
<td>Human-specific amplification of SWGA products</td>
</tr>
<tr>
<td>Human_R</td>
<td>AATTCTGGAACAACCTCCCCCCCTCT</td>
<td>Human-specific amplification of SWGA products</td>
</tr>
</tbody>
</table>
Supplementary Table 2. Primer sets used for SWGA reactions.

Asterisks (*) indicate phosphorothioate bonds. Primers have been previously described [15].

<table>
<thead>
<tr>
<th>Primer set 6A</th>
<th>Primer set 8A</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primer name</strong></td>
<td><strong>Primer sequence</strong></td>
</tr>
<tr>
<td>SASPfSWGA6.1</td>
<td>TAAATAAAAAA<em>A</em>A</td>
</tr>
<tr>
<td>SASPfSWGA6.2</td>
<td>CATAAAAAAA<em>A</em>A</td>
</tr>
<tr>
<td>SASPfSWGA6.3</td>
<td>TAAATAATAA<em>T</em>A</td>
</tr>
<tr>
<td>SASPfSWGA6.4</td>
<td>ATCATATAA<em>A</em>T</td>
</tr>
<tr>
<td>SASPfSWGA6.5</td>
<td>TAAACAAAAAA<em>A</em>A</td>
</tr>
<tr>
<td>SASPfSWGA6.6</td>
<td>TAATAATAA<em>A</em>A</td>
</tr>
<tr>
<td>SASPfSWGA6.7</td>
<td>TAACATAGG<em>T</em>C</td>
</tr>
<tr>
<td>SASPfSWGA6.8</td>
<td>TAGTAGTAG<em>T</em>A</td>
</tr>
<tr>
<td>SASPfSWGA6.9</td>
<td>ATAATAATA<em>A</em>T</td>
</tr>
<tr>
<td>SASPfSWGA6.10</td>
<td>CATAATAATA<em>A</em>T</td>
</tr>
</tbody>
</table>
Supplementary Table 3. Whole genome sequencing statistics for samples sequenced in this study. Sample pairs with sufficient exome coverage (>60% at ≥5X) for further analysis are indicated in bold.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Sample (pre- or post-treatment)</th>
<th>Parasite density (parasites/µL)</th>
<th>Total reads</th>
<th>% Reads mapped to <em>P. falciparum</em> genome</th>
<th>% Genome &gt;5X</th>
<th>% Exome &gt;5X</th>
</tr>
</thead>
<tbody>
<tr>
<td>103</td>
<td>pre</td>
<td>3,605</td>
<td>126,792,578</td>
<td>46.67%</td>
<td>38.44%</td>
<td>60.46%</td>
</tr>
<tr>
<td>103</td>
<td>post</td>
<td>286,447</td>
<td>85,566,496</td>
<td>64.80%</td>
<td>47.85%</td>
<td>71.69%</td>
</tr>
<tr>
<td>105</td>
<td>pre</td>
<td>5,589</td>
<td>69,281,634</td>
<td>51.83%</td>
<td>22.29%</td>
<td>34.93%</td>
</tr>
<tr>
<td>105</td>
<td>post</td>
<td>38</td>
<td>89,675,366</td>
<td>90.43%</td>
<td>65.62%</td>
<td>88.48%</td>
</tr>
<tr>
<td>108</td>
<td>pre</td>
<td>77,220</td>
<td>92,607,064</td>
<td>56.89%</td>
<td>33.43%</td>
<td>51.91%</td>
</tr>
<tr>
<td>108</td>
<td>post</td>
<td>121,685</td>
<td>130,536,610</td>
<td>61.64%</td>
<td>49.01%</td>
<td>73.77%</td>
</tr>
<tr>
<td>113</td>
<td>pre</td>
<td>54,080</td>
<td>57,901,552</td>
<td>44.56%</td>
<td>33.59%</td>
<td>53.57%</td>
</tr>
<tr>
<td>113</td>
<td>post</td>
<td>18</td>
<td>115,118,156</td>
<td>32.06%</td>
<td>17.86%</td>
<td>25.97%</td>
</tr>
<tr>
<td>114</td>
<td>pre</td>
<td>39,520</td>
<td>142,997,152</td>
<td>52.06%</td>
<td>47.60%</td>
<td>73.39%</td>
</tr>
<tr>
<td>114</td>
<td>post</td>
<td>1,323</td>
<td>94,691,726</td>
<td>47.34%</td>
<td>40.73%</td>
<td>62.10%</td>
</tr>
<tr>
<td>118</td>
<td>pre</td>
<td>164,320</td>
<td>95,589,360</td>
<td>50.31%</td>
<td>45.07%</td>
<td>69.97%</td>
</tr>
<tr>
<td>118</td>
<td>post</td>
<td>315,064</td>
<td>176,525,898</td>
<td>68.09%</td>
<td>72.27%</td>
<td>92.42%</td>
</tr>
<tr>
<td>119</td>
<td>pre</td>
<td>199,680</td>
<td>137,251,874</td>
<td>68.77%</td>
<td>70.77%</td>
<td>92.03%</td>
</tr>
<tr>
<td>119</td>
<td>post</td>
<td>22,537</td>
<td>82,233,988</td>
<td>59.91%</td>
<td>54.58%</td>
<td>81.23%</td>
</tr>
<tr>
<td>121</td>
<td>pre</td>
<td>29,120</td>
<td>108,754,820</td>
<td>68.12%</td>
<td>66.78%</td>
<td>89.60%</td>
</tr>
<tr>
<td>121</td>
<td>post</td>
<td>39</td>
<td>158,120,464</td>
<td>44.13%</td>
<td>47.07%</td>
<td>69.18%</td>
</tr>
<tr>
<td>127</td>
<td>pre</td>
<td>44,720</td>
<td>122,622,608</td>
<td>61.89%</td>
<td>60.71%</td>
<td>86.20%</td>
</tr>
<tr>
<td>127</td>
<td>post</td>
<td>158,515</td>
<td>137,474,360</td>
<td>64.48%</td>
<td>65.01%</td>
<td>88.93%</td>
</tr>
<tr>
<td>128</td>
<td>pre</td>
<td>185,120</td>
<td>135,016,170</td>
<td>51.51%</td>
<td>46.59%</td>
<td>70.95%</td>
</tr>
<tr>
<td>128</td>
<td>post</td>
<td>6,818</td>
<td>151,147,456</td>
<td>60.57%</td>
<td>62.37%</td>
<td>87.13%</td>
</tr>
<tr>
<td>130</td>
<td>pre</td>
<td>13,312</td>
<td>63,915,724</td>
<td>49.03%</td>
<td>26.14%</td>
<td>41.67%</td>
</tr>
<tr>
<td>130</td>
<td>post</td>
<td>25</td>
<td>64,516,982</td>
<td>16.11%</td>
<td>14.30%</td>
<td>20.29%</td>
</tr>
<tr>
<td>132</td>
<td>pre</td>
<td>35,443</td>
<td>106,972,976</td>
<td>63.63%</td>
<td>72.02%</td>
<td>91.57%</td>
</tr>
<tr>
<td>132</td>
<td>post</td>
<td>2,187</td>
<td>117,596,246</td>
<td>47.00%</td>
<td>53.88%</td>
<td>78.52%</td>
</tr>
</tbody>
</table>
Supplementary Table 4. SNPs identified in known resistance loci.

<table>
<thead>
<tr>
<th>Sample</th>
<th>PICRT (PF3D7_070 9000)</th>
<th>PMDR1 (PF3D7_0 523000)</th>
<th>PFDHFR (PF3D7_041 7200)</th>
<th>PFDHPS (PF3D7_0 810800)</th>
<th>PIK13 (PF3D7_1 343700)</th>
<th>PIATP4 (PF3D7_121 1900)</th>
<th>PDXR (PF3D7_14 67300)</th>
<th>PHAD1 (PF3D7_1 033400)</th>
</tr>
</thead>
<tbody>
<tr>
<td>103-pre</td>
<td>WT</td>
<td>Y184F</td>
<td>N51I, C59R, S108N</td>
<td>K540E</td>
<td>A578S</td>
<td>G1128R, Q1081K</td>
<td>WT</td>
<td>S265L</td>
</tr>
<tr>
<td>103-post</td>
<td>WT</td>
<td>N86Y</td>
<td>N51I, C59R, S108N</td>
<td>K540E</td>
<td>A578S</td>
<td>G1128R, Q1081K</td>
<td>WT</td>
<td>WT</td>
</tr>
<tr>
<td>114-pre</td>
<td>M74I, N75E, K76T</td>
<td>Y184F</td>
<td>N51I, C59R, S108N</td>
<td>K540E</td>
<td>WT</td>
<td>G1128R, Q1081K</td>
<td>WT</td>
<td>Q361R</td>
</tr>
<tr>
<td>114-post</td>
<td>WT</td>
<td>Y184F</td>
<td>N51I, C59R, S108N</td>
<td>WT</td>
<td>WT</td>
<td>G223S</td>
<td>Q361R</td>
<td>WT</td>
</tr>
<tr>
<td>118-pre</td>
<td>WT</td>
<td>N86Y</td>
<td>N51I, C59R, S108N</td>
<td>K540E</td>
<td>WT</td>
<td>G1128R, Q1081K</td>
<td>WT</td>
<td>N70S</td>
</tr>
<tr>
<td>118-post</td>
<td>WT</td>
<td>N86Y</td>
<td>N51I, C59R, S108N</td>
<td>K540E</td>
<td>WT</td>
<td>G1128R, Q1081K</td>
<td>WT</td>
<td>N70S</td>
</tr>
<tr>
<td>119-pre</td>
<td>D24Y, M74I, N75E, K76T</td>
<td>N86Y</td>
<td>N51I, C59R, S108N</td>
<td>K540E</td>
<td>WT</td>
<td>G1128R, Q1081K</td>
<td>WT</td>
<td>N70S</td>
</tr>
<tr>
<td>119-post</td>
<td>D24Y, M74I, N75E, K76T</td>
<td>N86Y</td>
<td>N51I, C59R, S108N</td>
<td>K540E</td>
<td>WT</td>
<td>G1128R, Q1081K</td>
<td>WT</td>
<td>N70S</td>
</tr>
<tr>
<td>121-pre</td>
<td>WT</td>
<td>Y184F</td>
<td>N51I, C59R, S108N</td>
<td>K540E</td>
<td>WT</td>
<td>T558A</td>
<td>WT</td>
<td>N70S</td>
</tr>
<tr>
<td>121-post</td>
<td>WT</td>
<td>Y184F</td>
<td>N51I, C59R, S108N</td>
<td>K540E</td>
<td>K189T</td>
<td>N1045K</td>
<td>WT</td>
<td>WT</td>
</tr>
<tr>
<td>127-pre</td>
<td>WT</td>
<td>WT</td>
<td>N51I, C59R, S108N</td>
<td>G437A</td>
<td>WT</td>
<td>Q1081K, E30D</td>
<td>WT</td>
<td>N70S</td>
</tr>
<tr>
<td>127-post</td>
<td>WT</td>
<td>WT</td>
<td>N51I, C59R, S108N</td>
<td>G437A</td>
<td>WT</td>
<td>G1128R, Q1081K</td>
<td>WT</td>
<td>N70S</td>
</tr>
<tr>
<td>128-pre</td>
<td>M74I, N75E, K76T</td>
<td>WT</td>
<td>N51I, C59R, S108N</td>
<td>K540E</td>
<td>K189T</td>
<td>Q1081K</td>
<td>WT</td>
<td>WT</td>
</tr>
<tr>
<td>128-post</td>
<td>M74I, N75E, K76T</td>
<td>WT</td>
<td>N51I, C59R, S108N</td>
<td>K540E</td>
<td>K189T</td>
<td>G1128R, Q1081K</td>
<td>WT</td>
<td>WT</td>
</tr>
<tr>
<td>132-pre</td>
<td>D24Y, M74I, N75E, K76T</td>
<td>N86Y</td>
<td>N51I, C59R, S108N</td>
<td>K540E</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
</tr>
<tr>
<td>132-post</td>
<td>M74I, N75E, K76T</td>
<td>N86Y, Y184F</td>
<td>N51I, C59R, S108N</td>
<td>K540E</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
<td>WT</td>
</tr>
</tbody>
</table>
Supplementary Table 5. Unique, non-synonymous SNPs shared in ≥50% of recrudescent (post-treatment) samples.

<table>
<thead>
<tr>
<th># strains</th>
<th>Chrom</th>
<th>Position</th>
<th>Locus</th>
<th>Product Description</th>
<th>Ref</th>
<th>SNP</th>
<th>Old AA/ New AA</th>
<th>1 0</th>
<th>1 1</th>
<th>1 1</th>
<th>1 2</th>
<th>1 2</th>
<th>1 2</th>
<th>1 3</th>
<th>1 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1</td>
<td>158547</td>
<td>PF3D7_ 0103500</td>
<td>conserved Plasmodium protein, unknown function</td>
<td>A</td>
<td>T</td>
<td>K/N</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>145125</td>
<td>PF3D7_ 0203000</td>
<td>repetitive organellar protein, putative (ROPE)</td>
<td>C</td>
<td>G</td>
<td>E/Q</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>203816</td>
<td>PF3D7_ 0204600</td>
<td>conserved Plasmodium protein, unknown function</td>
<td>C</td>
<td>T</td>
<td>G/D</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>667545</td>
<td>PF3D7_ 0216100</td>
<td>conserved Plasmodium protein, unknown function</td>
<td>C</td>
<td>T</td>
<td>E/K</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>900068</td>
<td>PF3D7_ 0228800</td>
<td>stevor</td>
<td>T</td>
<td>G</td>
<td>C/G</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>43675</td>
<td>PF3D7_ 0300100</td>
<td>erythrocyte membrane protein 1, PfEMP1 (VAR)</td>
<td>T</td>
<td>G</td>
<td>N/K</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>550797</td>
<td>PF3D7_ 0313500</td>
<td>GTP-binding protein EnvA, putative</td>
<td>C</td>
<td>A</td>
<td>R/S</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>622763</td>
<td>PF3D7_ 0315200</td>
<td>circumsporozoite- and TRAP-related protein (CTRP)</td>
<td>G</td>
<td>C</td>
<td>E/Q</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>851156</td>
<td>PF3D7_ 0320400</td>
<td>oocyst capsule protein (Cap380)</td>
<td>C</td>
<td>T</td>
<td>A/T</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>853052</td>
<td>PF3D7_ 0320400</td>
<td>oocyst capsule protein (Cap380)</td>
<td>T</td>
<td>C</td>
<td>M/V</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>909278</td>
<td>PF3D7_ 0321700</td>
<td>microtubule and actin binding protein, putative</td>
<td>C</td>
<td>T</td>
<td>E/K</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>73407</td>
<td>PF3D7_ 0401000</td>
<td>rifin (RIF)</td>
<td>C</td>
<td>A</td>
<td>C/F</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>73479</td>
<td>PF3D7_ 0401000</td>
<td>rifin (RIF)</td>
<td>C</td>
<td>G</td>
<td>G/A</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>251022</td>
<td>PF3D7_ 0404600</td>
<td>conserved Plasmodium membrane protein, unknown function</td>
<td>T</td>
<td>G</td>
<td>F/V</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>262055</td>
<td>PF3D7_ 0404700</td>
<td>dipeptidyl aminopeptidase 3 (DAPAP3)</td>
<td>C</td>
<td>T</td>
<td>D/N</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>472117</td>
<td>PF3D7_ 0410300</td>
<td>protein phosphatase PPM1, putative (PPM1)</td>
<td>T</td>
<td>G</td>
<td>N/H</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>520965</td>
<td>PF3D7_ 0411800</td>
<td>conserved Plasmodium protein, unknown function</td>
<td>T</td>
<td>C</td>
<td>K/E</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>520967</td>
<td>PF3D7_ 0411800</td>
<td>conserved Plasmodium protein, unknown function</td>
<td>A</td>
<td>T</td>
<td>V/E</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>1102488</td>
<td>PF3D7_ 0424400</td>
<td>surface-associated interspersed protein 4.2 (SURFIN 4.2) (SURF4.2)</td>
<td>G</td>
<td>A</td>
<td>M/I</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>1137176</td>
<td>PF3D7_ 0425200</td>
<td>Plasmodium exported protein (hyp15), unknown function</td>
<td>G</td>
<td>C</td>
<td>D/H</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>1167005</td>
<td>PF3D7_0425800</td>
<td>erythrocyte membrane protein 1, PfEMP1 (VAR)</td>
<td>G</td>
<td>T</td>
<td>Q/H</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>633280</td>
<td>PF3D7_0515300</td>
<td>phosphatidylinositol 3-kinase (PI3K)</td>
<td>A</td>
<td>G</td>
<td>I/V</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>1004577</td>
<td>PF3D7_0524100</td>
<td>conserved Plasmodium protein, unknown function</td>
<td>G</td>
<td>A</td>
<td>R/K</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>524370</td>
<td>PF3D7_0612600</td>
<td>PP-loop family protein, putative</td>
<td>T</td>
<td>A</td>
<td>N/Y</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>665740</td>
<td>PF3D7_0615900</td>
<td>conserved Plasmodium protein, unknown function</td>
<td>C</td>
<td>T</td>
<td>D/N</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>1368413</td>
<td>PF3D7_0632600</td>
<td>rifin, pseudogene (RIF)</td>
<td>G</td>
<td>T</td>
<td>R/M</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>350194</td>
<td>PF3D7_0707500</td>
<td>conserved Plasmodium protein, unknown function</td>
<td>A</td>
<td>G</td>
<td>S/G</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>470161</td>
<td>PF3D7_0710200</td>
<td>conserved Plasmodium protein, unknown function</td>
<td>C</td>
<td>A</td>
<td>L/I</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>634749</td>
<td>PF3D7_0713900</td>
<td>conserved Plasmodium protein, unknown function</td>
<td>C</td>
<td>T</td>
<td>D/N</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>911077</td>
<td>PF3D7_0721000</td>
<td>conserved Plasmodium membrane protein, unknown function</td>
<td>G</td>
<td>C</td>
<td>M/I</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>911078</td>
<td>PF3D7_0721000</td>
<td>conserved Plasmodium membrane protein, unknown function</td>
<td>C</td>
<td>G</td>
<td>R/G</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>1100677</td>
<td>PF3D7_0726200</td>
<td>serine/threonine protein kinase, FIKK family (FIKK7.1)</td>
<td>G</td>
<td>A</td>
<td>E/K</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>1358889</td>
<td>PF3D7_0731500</td>
<td>erythrocyte binding antigen-175 (EBA175)</td>
<td>A</td>
<td>G</td>
<td>K/E</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>50206</td>
<td>PF3D7_0800300</td>
<td>erythrocyte membrane protein 1, PfEMP1 (VAR)</td>
<td>G</td>
<td>C</td>
<td>S/T</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>135161</td>
<td>PF3D7_0801900</td>
<td>conserved Plasmodium protein, unknown function</td>
<td>A</td>
<td>C</td>
<td>N/H</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>135171</td>
<td>PF3D7_0801900</td>
<td>conserved Plasmodium protein, unknown function</td>
<td>A</td>
<td>G</td>
<td>H/R</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>1325994</td>
<td>PF3D7_0831100</td>
<td>surface-associated interspersed protein 8.1 (SURFIN 8.1) (SURF8.1)</td>
<td>A</td>
<td>C</td>
<td>H/P</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>1400201</td>
<td>PF3D7_0832400</td>
<td>stevor</td>
<td>G</td>
<td>T</td>
<td>Q/K</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>27785</td>
<td>PF3D7_0900100</td>
<td>erythrocyte membrane protein 1, PfEMP1 (VAR)</td>
<td>C</td>
<td>G</td>
<td>T/S</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>453285</td>
<td>PF3D7_0910000</td>
<td>SET domain protein, putative (SET4)</td>
<td>A</td>
<td>G</td>
<td>V/A</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>1492960</td>
<td>PF3D7_0937700</td>
<td>rifin (RIF)</td>
<td>G</td>
<td>A</td>
<td>S/N</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>68850</td>
<td>PF3D7_1001100</td>
<td>acyl-CoA binding protein, isoform 1, ACBP1 (ACBP1)</td>
<td>A</td>
<td>T</td>
<td>L/I</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>721205</td>
<td>PF3D7</td>
<td>conserved Plasmodium</td>
<td>A</td>
<td>G</td>
<td>M/V</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein Name/Function</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein, unknown function</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>eukaryotic translation initiation factor subunit eIF2A, putative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>glutamate-rich protein (GLURP)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>conserved Plasmodium protein, unknown function</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dolichyl-diphosphooligosaccharide-protein glycosyltransferase subunit STT3, putative (STT3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>conserved Plasmodium protein, unknown function</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plasmodium exported protein (PHISTb), unknown function</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>non-SERCA-type Ca2+ -transporting P-ATPase (ATP4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>serpine receptor, putative (SR10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>asparagine and aspartate rich protein 1 (AARP1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>conserved Plasmodium protein, unknown function</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>conserved Plasmodium protein, unknown function</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>conserved Plasmodium protein, unknown function</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rifin (RIF)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-loop containing nucleoside triphosphate hydrolase, putative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rifin (RIF)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>conserved Plasmodium protein, unknown function</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>conserved Plasmodium protein, unknown function</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Accession Number</th>
<th>Start Position</th>
<th>End Position</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF3D7_1019000</td>
<td>758639</td>
<td>758644</td>
<td>eukaryotic translation initiation factor subunit eIF2A, putative</td>
</tr>
<tr>
<td>PF3D7_1019000</td>
<td>1399387</td>
<td>1399387</td>
<td>glutamate-rich protein (GLURP)</td>
</tr>
<tr>
<td>PF3D7_1110300</td>
<td>410321</td>
<td>410321</td>
<td>conserved Plasmodium protein, unknown function</td>
</tr>
<tr>
<td>PF3D7_11116600</td>
<td>627737</td>
<td>627737</td>
<td>dolichyl-diphosphooligosaccharide-protein glycosyltransferase subunit STT3, putative (STT3)</td>
</tr>
<tr>
<td>PF3D7_1117200</td>
<td>658260</td>
<td>658260</td>
<td>conserved Plasmodium protein, unknown function</td>
</tr>
<tr>
<td>PF3D7_1200600</td>
<td>51465</td>
<td>51465</td>
<td>erythocyte membrane protein 1, PiEMP1 (VAR2CSA)</td>
</tr>
<tr>
<td>PF3D7_1201000</td>
<td>77787</td>
<td>77787</td>
<td>Plasmodium exported protein (PHISTb), unknown function</td>
</tr>
<tr>
<td>PF3D7_1201000</td>
<td>77847</td>
<td>77847</td>
<td>Plasmodium exported protein (PHISTb), unknown function</td>
</tr>
<tr>
<td>PF3D7_1211900</td>
<td>529418</td>
<td>529418</td>
<td>non-SERCA-type Ca2+ -transporting P-ATPase (ATP4)</td>
</tr>
<tr>
<td>PF3D7_1215900</td>
<td>642447</td>
<td>642447</td>
<td>serpine receptor, putative (SR10)</td>
</tr>
<tr>
<td>PF3D7_1233600</td>
<td>1390639</td>
<td>1390639</td>
<td>asparagine and aspartate rich protein 1 (AARP1)</td>
</tr>
<tr>
<td>PF3D7_1237100</td>
<td>1543212</td>
<td>1543212</td>
<td>conserved Plasmodium protein, unknown function</td>
</tr>
<tr>
<td>PF3D7_1240600</td>
<td>1721650</td>
<td>1721650</td>
<td>erythocyte membrane protein 1, PiEMP1 (VAR)</td>
</tr>
<tr>
<td>PF3D7_1241200</td>
<td>1749231</td>
<td>1749231</td>
<td>conserved Plasmodium protein, unknown function</td>
</tr>
<tr>
<td>PF3D7_1300600</td>
<td>54740</td>
<td>54740</td>
<td>rifin (RIF)</td>
</tr>
<tr>
<td>PF3D7_1352700</td>
<td>2103626</td>
<td>2103626</td>
<td>P-loop containing nucleoside triphosphate hydrolase, putative</td>
</tr>
<tr>
<td>PF3D7_1400200</td>
<td>8038</td>
<td>8038</td>
<td>rifin (RIF)</td>
</tr>
<tr>
<td>PF3D7_1400500</td>
<td>17604</td>
<td>17604</td>
<td>rifin (RIF)</td>
</tr>
<tr>
<td>PF3D7_1436700</td>
<td>1496415</td>
<td>1496415</td>
<td>conserved Plasmodium protein, unknown function</td>
</tr>
<tr>
<td>PF3D7_1449600</td>
<td>2035328</td>
<td>2035328</td>
<td>conserved Plasmodium protein, unknown function</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>function</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>5</td>
<td>14</td>
<td>2764134</td>
<td>PF3D7_1467600</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>2821537</td>
<td>PF3D7_1468600</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>3235381</td>
<td>PF3D7_1478600</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>3235382</td>
<td>PF3D7_1478600</td>
</tr>
</tbody>
</table>
3.8 REFERENCES


Chapter 4

Fosmidomycin resistance reveals coordinated metabolic control in malaria parasites
PREFACE

The following work was performed by me, Aakash Y. Gandhi, Emilio F. Merino, Maria B. Cassera, Samuel J. Erlinger, and Audrey R. Odom. I quantified FSM resistance, generated and genotyped the cloned strains, prepared samples for genome sequencing and variant discovery, performed parasite transfections and genetic complementation, prepared samples for mass spectrometry, measured strain growth rates, performed live fluorescent microscopy, and wrote the manuscript. A.Y.G. performed the enzymatic characterization and lysate activity assays. E.F.M. and M.B.C. performed mass spectrometry to measure AMP levels. S.J.E. cloned the HAD2 expression construct. A.R.O. designed, supervised, and analyzed the studies and wrote the manuscript.

We thank the Proteomics and Mass Spectrometry Facility at the Donald Danforth Plant Science Center, St. Louis, MO for assistance with LC-MS/MS. We thank Daniel Goldberg (Washington University, St. Louis) for supplying the pTEOE110 plasmid. Finally, we thank the Malaria Research Reference and Reagent Resource (MR4) for providing us with reagents contributed by DJ Carucci (MRA-102) and John Adams (MRA-912).

This material is based upon work supported by the National Science Foundation under grant number DBI-0521250 for acquisition of the QTRAP LC-MS/MS.
4.1 ABSTRACT

In the malaria parasite *Plasmodium falciparum*, the synthesis of isoprenoids from glycolytic intermediates is essential for parasite survival. The antibiotic fosmidomycin (FSM) is a well-validated inhibitor of isoprenoid synthesis and an effective antimalarial. In FSM-resistant *P. falciparum*, we identify a loss-of-function mutation in *HAD2* (PF3D7_1226300) as causative for resistance. Enzymatic characterization shows that HAD2, a member of the haloacid dehalogenase-like hydrolase (HAD) superfamily, functions as a purine phosphatase. Harnessing a growth defect in *HAD2*-mutant parasites, we select for suppression of HAD2-mediated FSM resistance and uncover hypomorphic suppressor mutations in the locus encoding the glycolytic enzyme PFK9 (PF3D7_0915400). Metabolic profiling demonstrates that FSM resistance is achieved via increased steady-state levels of MEP pathway and glycolytic intermediates and confirms reduced PFK9 function in the suppressed strains. Together, these data point to HAD2 as a novel and essential regulator of parasite metabolism and drug sensitivity and expand our understanding of this large and important family of regulators.

4.2 INTRODUCTION

Malaria remains a persistent global health burden, causing hundreds of millions of infections and nearly a half million deaths per year [1]. The most severe malaria cases are caused by the apicomplexan parasite *Plasmodium falciparum*. One of the greatest challenges in malaria eradication is the continued development of resistance to antimalarial drugs. Delayed parasite clearance times have been observed after use of artemisinin combination therapies (ACTs), the
current frontline treatment for malaria in endemic areas [1,2]. New antimalarials are urgently needed. As an obligate intracellular parasite of human erythrocytes, the parasite has a defined and unique metabolism that may be exploited for the discovery of new drug targets and the development of new therapies.

In the red blood cell niche, Plasmodium parasites are highly dependent on glucose metabolism. Infection with Plasmodium spp. results in nearly 100-fold increase in glucose import in red blood cells [3–5]. Despite these energy requirements, the parasite demonstrates little aerobic respiration via the TCA cycle and instead relies on anaerobic glycolysis to produce ATP [6–9]. This is evidenced by studies showing that most parasite glucose is excreted as lactate [10,11].

Besides ATP production, glucose also has a number of anabolic fates in *P. falciparum*. One such fate is the synthesis of isoprenoids. Isoprenoids are a large class of hydrocarbons with extensive structural and functional diversity [12]. They are present and essential in all organisms, including *P. falciparum* [13,14] and a number of related microbes [15,16]. In the malaria parasite, isoprenoids perform a number of essential functions, such as protein prenylation and dolichylation and synthesis of GPI anchors [17–23]. In other organisms, isoprenoids also function in essential cellular processes such as cell wall synthesis, aerobic respiration, and photosynthesis [12].

Despite their diversity, all isoprenoids are synthesized from the five-carbon building block, isopentyl pyrophosphate (IPP). Evolution has produced two distinct pathways for IPP synthesis: the mevalonate pathway, which is found in archaea, fungi, animals, and the cytoplasm of plants;
and the methyly erythritol phosphate (MEP) pathway, which is found in most bacteria, plant chloroplasts, and apicomplexan parasites such as *P. falciparum* [24]. Because of its absence in the human host and its essential role in the parasite, the MEP pathway is an excellent target for antimalarial development. The antibiotic and antimalarial fosmidomycin (FSM) is a competitive inhibitor of the first committed enzymatic step of the MEP pathway, catalyzed by 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR, E.C. 1.1.1.267) [25–27]. FSM has been validated as a specific inhibitor of the MEP pathway in *P. falciparum* [14, 28–31] and is an excellent chemical tool to study MEP pathway biology.

The utilization of glucose by the parasite for the production of energy and secondary metabolites such as isoprenoids is likely subject to regulation. We previously used a screen for FSM resistance to identify HAD1, a negative metabolic regulator whose loss results in increased levels of MEP pathway intermediates and resistance to MEP pathway inhibition. HAD1 is a cytoplasmic sugar phosphatase that may hydrolyze glycolytic intermediates upstream of the MEP pathway [32]. HAD1 belongs to the haloacid dehalogenase-like hydrolase (HAD) enzyme superfamily (Interpro families IPR023214 and IPR000150) [33]. While HADs are found in all kingdoms of life, HAD1 is most related to bacterial members of this superfamily [32,34], which have been implicated in metabolic regulation, stress response, and phosphate homeostasis [35–39]. However, many members of this superfamily remain uncharacterized. The study of HAD proteins in evolutionarily and metabolically distinct organisms such as *P. falciparum* will greatly facilitate our understanding of this large and important class of enzymes.
In many systems, metabolic regulation of a pathway is often centered on the first committed enzymatic step [40]. In glycolysis, this step is catalyzed by phosphofructokinase (PFK, E.C. 2.7.1.11) [41]. PFKs in other organisms have been found to be subject to allosteric regulation, including inhibition by ATP and citrate and activation by AMP, ADP, and fructose 2,6-bisphosphate [42,43]. PFKs of some apicomplexans have been studied and are found to be most related to pyrophosphate (PPi)-utilizing PFKs from plants [44–47]. A previous study described *P. falciparum* PFK9, one of two PFKs encoded in the *P. falciparum* genome [44]. The second PFK (PFK11) is expressed mainly in the gametocyte stages of the parasite lifecycle and is not believed to be catalytically active. Mony et al. [44] describe PFK9 as a single polypeptide encoding two subunits (α and β). While full-length recombinant enzyme could not be purified, PFK from cell lysate and the recombinant β subunit were found to be active with ATP as the phosphoryl donor. In these assays, PFK9 was found to be insensitive to most tested allosteric regulators [44]. Further studies are required to understand how the parasite utilizes this canonically important step of glycolysis for metabolic regulation.

In this study, we describe the role of HAD2, another HAD family member in *P. falciparum*. We find that HAD2 functions as a negative metabolic regulator whose loss results in increased MEP pathway and glycolytic metabolites and resistance to FSM. In vitro, HAD2 dephosphorylates purine monophosphates, such as AMP and GMP. Selection for suppression of FSM resistance identifies mutations in *PFK9*. Through our forward genetics approach, we further define the role of the HAD protein family and uncover novel mechanisms of regulation of essential parasite metabolism and drug sensitivity.
4.3 METHODS

4.3.1 Reagents

All reagents were purchased from Sigma-Aldrich (St. Louis, MO) unless otherwise specified.

4.3.2 Parasite culture maintenance

All strains are derived from the genome reference strain 3D7 [48] (MRA-102, MR4, ATCC, Manassas, VA). Unless otherwise indicated, parasites were maintained in a 2% suspension of human erythrocytes in RPMI medium (Sigma Aldrich, SKU R4130) modified with 2.25 g/L NaHCO$_3$, 2 g/L glucose, 1.19 g/L HEPES, 2.5 mg/L thymidine, 0.22 g/L sodium pyruvate, 50 mg/L hypoxanthine, 10 µg/mL gentamycin, and 5 g/L Albumax (Life Technologies, Carlsbad, CA). Parasites were maintained at 37 °C in 5% O$_2$/5% CO$_2$/90% N$_2$.

4.3.3 Generation of FSM$^R$ strain E2 and E2 clones

FSM$^R$ strain E2 was generated as previously described [32]. Briefly, a clone of genome reference strain 3D7 (MRA-102, MR4, ATCC, Manassas, VA) was used as a parent strain. Culture was at 4% parasitemia at the start of selection with 0.5 µM FSM. Selection was scaled to 3 µM FSM in 0.5-1 uM increments. Clones of strain E2 were isolated by limiting dilution.

4.3.4 Quantification of FSM resistance

Asynchronous parasite cultures were diluted to 0.5-1.0% parasitemia. Assays were performed in opaque 96-well plates in 100 µL volumes with FSM concentrations ranging from 0.63 nM to 1 mM. After 3 days, media was removed and parasitemia was measured by DNA content using
Picogreen fluorescence (Life Technologies, Carlsbad, CA) as previously described [49]. Half maximal inhibitory concentration (IC$_{50}$) values were calculated by nonlinear regression analysis using Graphpad Prism software.

### 4.3.5 Generation of recombinant HAD2

The predicted coding sequence of *HAD2* (PlasmoDB ID PF3D7_1226300) was amplified from *P. falciparum* genomic DNA by PCR using the following primers: 5’-

CTCACCACCACCACCACCATATGGCTTCTAGTAACGATGTACA-3’ and 5’-

ATCCTATCTTACTCATTATTTTTTTTCAAGTCAAATACTTTTTTTAA-3’. Ligation-independent cloning was used to clone the PCR product into vector BG1861 [50], which introduces an N-terminal 6xHis fusion to the expressed protein. Sanger sequencing was used to verify the cloned construct. BG1861:6xHis-HAD2 construct was transformed into One Shot BL21(DE3)pLysS *Escherichia coli* cells (Life Technologies, Carlsbad, CA) according to supplier instructions. Protein expression was induced with isopropyl-β-D-thiogalactoside at mid-log growth (OD$_{600}$ 0.4 – 0.5). After 3 hours, cells were collected by centrifugation and stored at -20°C.

Induced cell pellets were resuspended in lysis buffer containing 1 mg/mL lysozyme, 20 mM imidazole, 1 mM dithiothreitol, 10 mM Tris HCl (pH 7.5), 30 U benzonase (EMD Millipore, Darmstadt, Germany), and Complete Mini EDTA-free protease inhibitor tablets (Roche Applied Science, Penzberg, Germany). 6xHis-HAD2 was bound to nickel agarose beads (Gold Biotechnology, St. Louis, MO) and eluted in 300 mM imidazole, 20mM Tris HCl (pH 7.5), and 150 mM NaCl. This eluate was further purified by size-exclusion gel chromatography using a
HiLoad 16/600 Superdex 200 pg column (GE Healthcare) equilibrated in 50 mM Tris HCl (pH 7.5), 1 mM dithiothreitol, and 1 mM MgCl$_2$. The elution fractions containing HAD2 were pooled into a centrifugal filter and concentrated to approximately 3 mg/ml. Glycerol was added to a final concentration of 10% (w/v), and the solution was immediately flash frozen in liquid nitrogen and stored at -80°C.

The purification of recombinant HAD1 has been previously described [32].

4.3.6 Antisera generation

Anti-HAD1 antisera has been previously described (MRA-1256, MR4, ATCC, Manassas, VA) [32]. Anti-HAD2 polyclonal antisera was raised against 6xHis-HAD2 in rabbits by Cocalico Biologicals (Reamston, PA) using their standard protocol. Titermax was used as an adjuvant. Antisera specificity was confirmed by immunoblot of purified recombinant 6xHis-HAD2 and 6xHis-HAD1 (Supplemental Figure 10) as well as for *P. falciparum* lysate lacking HAD2 (Supplemental Figure 9B, strain S1). Anti-HAD2 and anti-HAD1 antisera were found to be specific for their respective antigens.

4.3.7 Immunoblotting

Lysates from saponin-lysed parasites were separated on a polyacrylamide gel and transferred to a polyvinylidene difluoride membrane. Membranes were blocked in 5% non-fat dry milk in PBS, 0.1% Tween-20. HAD2 was detected in using a 1:2,000-5,000 dilution (against *P. falciparum* lysates) or 1:15,000 (against recombinant protein) of anti-HAD2 rabbit polyclonal antisera (Cocalico Biologicals, Reamstown, PA). Anti-HAD1 antisera was used at a 1:20,000 dilution.
Rabbit polyclonal anti-heat shock protein 70 (Hsp70) was detected using a 1:5,000 dilution (AS08 371, Agrisera Antibodies, Vännäs, Sweden). All blots used an HRP-conjugated goat anti-rabbit IgG secondary antibody at 1:20,000 (ThermoFisher Scientific 65-6120). When necessary, blots were stripped by washing with 200 mM glycine, 0.1% SDS, 1% Tween-20, pH 2.2 before re-probing.

4.3.8 Quantitative PCR of DXS and DXR

Parasite pellets from saponin-lysed infected erythrocytes were stored at -80 °C. Total RNA was extracted using the Ambion Purelink RNA Mini kit (Life Technologies, Carlsbad, CA) according to kit instructions. After the first wash step, an on-column DNase (Qiagen, Venlo, Netherlands) treatment was performed. RNA was stored at -80 °C in RNase-free water until use. To synthesize cDNA, one µg of total RNA was used in reverse transcriptase reactions using the Quantitect Reverse Transcription kit (Qiagen, Venlo, Netherlands).

Thirty ng of cDNA was used as a template in quantitative PCR reactions containing Fast SYBR Green PCR Master Mix (Applied Biosystems, Waltham, MA) and 300 nM each primer. Primers used were as follows:

**DXS** (PF3D7_1337200): 5’- AACGTGGATAAAAGTACACATTGC-3’ and 5’- TGATATACCTACGGGATTTTGG-3’

**DXR** (PF3D7_1467300): 5’- ACATGGCCTGATAATAGCTACAGTT-3’ and 5’- TTCATTTGACGCATTTAGTCAGTT-3’

Beta tubulin and 18S rRNA were used as control genes:
Beta tubulin (PF3D7_1008700): 5’- ATCCCATCCCACGTTTACATT-3’ and 5’-
TCCTTTGTGGACATTCTTCCTC-3’
18s rRNA (PF3D7_0112300): 5’- GAACGAGGAATGCCTAGTAAGCA-3’ and 5’-
TTCATCATATCTTTCAATCGGTAGGA-3’.

Thermocycling was performed using an Applied Biosystems 7500 Fast RT PCR System with the following parameters: 95 °C for 30 seconds and 40 cycles of 95 °C for 3 seconds, 60 °C for 30 seconds. Controls lacking reverse transcriptase and template both produced no significant signal. Melt curve analysis was used to verify that all primer sets produce single products.

Ct values were calculated using Applied Biosystems 7500 Fast System software. Relative expression levels of DXS and DXR were calculated using the ΔΔCt method. ΔCt represents the Ct value of the reference gene (beta tubulin and 18s rRNA combined) subtracted from the target. ΔΔCt represents the ΔCt value of the parental sample subtracted from the sample of interest. Fold changes are calculated as $2^{-\Delta\Delta C_t}$.

4.3.9 Whole genome sequencing and SNP analysis

Parasite genomic DNA was prepared for sequencing as previously described [32]. One microgram of genomic DNA was sheared, end repaired, and adapter ligated. Libraries were sequenced on an Illumina HiSeq 2500 in Rapid Run mode to generate 101 bp paired end reads. Reads were aligned to the P. falciparum 3D7 reference genome (PlasmoDB v7.2) using Novoalign (V2.08.02). Duplicate reads were removed. SNPs were called using samtools, using a
quality score cutoff of 20 and a read depth cutoff of 5. SNPs were annotated using snpEff. Parental SNPs were removed by comparing to previous sequenced parental genomes [32].

Greater than 75% of the exome was covered >5X for all samples. Whole genome sequencing data from this study is available through the NCBI BioProject database (PRJNA222697) and Sequence Read Archive (SRP038937). Library preparation, Illumina sequencing, read alignment, and variant analysis were performed by the Genome Technology Access Center (GTAC) at Washington University.

4.3.10 Sanger sequencing

The E2 A469T (R157X) HAD2 SNP was verified by amplifying *P. falciparum* genomic DNA using the following gene-specific primers: 5’-GGATATACTTTATTAGATGAGAC-3’ and 5’-ATTATATGTGTTGAAATATGGTCAATT-3’. The *PFK9* locus was amplified using the following primers, which also include sites for ligation-independent cloning: 5’-CTCACACCACCACCACCATATGGATACCAAGAGTGAGATAAAA-3’ and 5’-ATCCTATCTTACTCACTTAGTTCATTCTTTTTCTCTGTTTTTC-3’. Amplicons were sequenced using the following primers: 5’-CACAGGTAATGAATTCCCAGC-3’, 5’-ATCATCGGCATTCTGACATAAC-3’, 5’-GATCTATGCGTTTTGAACAATTAG-3’, 5’-CTGTGAAGCCCAGTTTCAATAAC-3’, 5’-CCCAGTTGATTCTCTATCCATTAA-3’, 5’-GGTCCATTTGATGCTTCGAAAC-3’, 5’-CGTTGTTATACGTAAAGACTCCA-3’, and 5’-CTAGTCCATTATTGGAAATAAGAA-3’. Amplicons sequenced by the Washington University Protein and Nucleic Acid Laboratory using BigDye Terminator v3.1 Cycle
Sequencing Reagents (Life Technologies, Carlsbad, CA). Chromatogram files were viewed using DNASTar SeqMan software.

4.3.11 HAD2 activity assays

Assay reactions were performed at 37°C in clear 96-well half-area plates and measured using a FLUOstar Omega microplate reader (BMG Labtech, Ortenberg, Germany). Assay data was analyzed and reaction rates determined using GraphPad Prism software.

Phosphatase activity of recombinant enzyme was initially evaluated against the substrate pNPP (New England Biolabs, Ipswich, MA). Reactions contained 10 mM pNPP, 50 mM Tris HCl (pH 7.5), 5 mM MgCl₂, 0.5 mM MnCl₂, and 0.25–4 µg purified recombinant enzyme. Para-nitrophenyl production was quantified by absorbance at 405 nm.

Specific enzyme activity against phosphorylated sugar substrates was measured using the EnzChek Phosphate Assay Kit (Life Technologies, Carlsbad, CA) according to supplier instructions. Substrates were purchased from Sigma-Aldrich (St. Louis, MO), except for DOXP and MEP (Echelon Biosciences, Salt Lake City, UT) and fructose 1-phosphate (Santa Cruz Biotechnology, Dallas, TX). Reactions were performed in 50 µL volumes containing 1 mM substrate and 150–750 ng recombinant purified enzyme.

Kinetic parameters were measured using 150 ng recombinant purified HAD2 enzyme, with continuous measurement of product formation in response to 0.06 – 10 mM substrate. Michaelis–Menten parameters $K_m$ and $V_{max}$ were determined by non-linear regression (GraphPad
Prism). The reaction rate \((k_{\text{cat}})\) and catalytic efficiency \((k_{\text{cat}}/K_{m})\) were also determined from these parameters.

4.3.12 Measurement of AMP levels

Trophozoite pellets were prepared as described above. Briefly, synchronized early trophozoites were lysed with saponin, washed with cold 2 g/L glucose in PBS, and frozen at -80 °C. Samples were extracted and prepared for ultra-high performance liquid chromatography tandem mass spectrometry (IP-RP–UPLC–MS/MS) as previously described [51].

4.3.13 P. falciparum growth assays

Asynchronous parasite cultures were seeded at ~1% parasitemia. Samples were taken every 1-2 days for 12 days and fixed in 4% paraformaldehyde, 0.05% glutaraldehyde in PBS. Media was exchanged daily. Parasites were subcultured 1:2 or 1:4 throughout the assay to avoid growth rate perturbations due to high parasite density. Fixed samples were washed with PBS and stained with 0.01 mg/ml acridine orange (Invitrogen) in PBS. Parasitemia was determined by measuring fluorescence on a BD Biosciences LSRII flow cytometer. Parasitemias were corrected for subculturing and data was plotted as % growth from day 0.

4.3.14 Preparation of soluble P. falciparum whole-cell lysate

Cultures were synchronized with 5% sorbitol. Synchronized trophozoites were isolated 32 hours post-invasion by treatment of infected erythrocytes with 0.1% saponin. Isolated trophozoites were washed three times in 100 μL buffer containing 100 mM Tris-HCl (pH 7.5), 1 mM MgCl₂, 1 mM DTT, 10% glycerol (w:w), and supplemented according to manufacturer instructions with
Complete Mini EDTA-free protease inhibitor tablets (Roche). Lysates were prepared by three sonication steps at 4°C (Fischer Scientific Model 550 Sonic Dismembrator at an amplitude setting of 3.5), followed by centrifugation at 4°C (10,000 x g, 10 min). Supernatant was isolated and total protein concentration of lysate was determined using a bicinchoninic acid (BCA)-linked assay kit and performed according to manufacturer’s instruction (Thermo Scientific Pierce). Lysates were prepared fresh and kept on ice for all assays.

4.3.15 Assay of PFK9 activity from lysate

The activity of PFK9 in lysate was monitored by linking it to the oxidation of NADH via linking enzymes, as adapted from Mony et al [44] and as originally described by Beutler [52]. Reactions were performed in 150 µL volumes with assay buffer containing 100 mM Tris-HCl (pH 7.5), 1 mM MgCl₂, 1 mM DTT, 0.25 mM NADH, 1 mM ATP, 3 mM fructose 6-phosphate, as well as excess of linking enzymes aldolase (7.5 U), triose-phosphate isomerase (3.8 U), and glycerol 3-phosphate dehydrogenase (3.8 U). After adding fresh cell lysate (15 µg total protein content), the absorbance at 340 nm was measured at 37°C for 40 min. Activity (µmol/min/mg total protein) was determined from a linear regression taken between 25 min and 35 min of the absorbance curve. For some experiments, a range of substrate (ATP or fructose 6-phosphate) or lysate concentrations were used as indicated. In some cases, Bacillus stearothermophilus PFK (BsPFK) was used as a positive control.

4.3.16 LC-MS/MS sample preparation

All cultures were grown to >5% parasitemia at 4% hematocrit in 25 mL volumes. Cultures were synchronized with 5% sorbitol to obtain ring-stage parasites. At approximately 20 hours post-
invasion (early trophozoites), parasites were isolated by saponin lysis, washed with cold 2 g/L glucose in PBS, and frozen at -80 °C until extraction. Samples were extracted in 600 µL of ice-cold extraction solvent [chloroform, methanol, and acetonitrile (2:1:1, v/v/v)] using two liquid-nitrogen cooled 3.2 mm stainless steel beads and homogenization in Tissue-Lyser II instrument (Qiagen, Venlo, Netherlands) at 20 Hz for 5 minutes in a cold sample rack. Ice-cold water was added and samples were homogenized for 5 more minutes at 20 Hz. Samples were centrifuged at 14,000 rcf at 4°C for 5 min. The polar phase was lyophilized and re-dissolved in 100 µL water and analyzed by LC-MS/MS.

4.3.17 LC-MS/MS analysis of MEP pathway and glycolytic intermediates

LC-MS/MS was performed on a 4000QTRAP system (AB Sciex, Framingham, MA) in multiple-reaction monitoring mode (MRM) using negative ionization. The specific parameters used for analysis of MEP pathway metabolites have been previously described [28]. Liquid chromatography separation was done using ion pair reverse-phase chromatography [53]. The ion pair reagent used was 10 mM tributylammonium acetate (pH 5.1-5.5). The referenced method [53] was modified: (1) RP-hydro 100 mm × 2.0 mm, 2.5 µm high performance liquid chromatography column (Phenomenex, Torrance, CA), (2) flow rate of 0.14 mL/min, (3) solvent A of 10 mM tributylammonium acetate in 5% methanol, (4) binary LC gradient (20% solvent B (100% methanol) from 0 to 2.5 min, 30% B for 12.5 min, 80% B for 5 min, and column equilibration at for 5 minutes), and (5) 20 µL autosampler injection volume.
4.3.18 pTEOE110:HAD2 plasmid construction

The pTEOE110:HAD2-GFP construct contains the heat shock protein 110 (PF3D7_0708800) 5’ UTR and a C-terminal GFP tag. Human dihydrofolate reductase (hDHFR) is present as a selectable marker. Inverted terminal repeats are included for integration into the genome by a co-transfected piggyBac transposase (pHTH, MRA912, MR4, ATCC, Manassas, VA).

HAD2 was amplified with the following primers: 5’-
GATCCTCGAGATGGCTTCTAGTAACGATGTACATT-3’ and 5’-
GATCCCTAGGTTTTTTTTTCAAGTCAAATACTTTTTTTATAAG-3’ and cloned into AvrII and XhoI sites in the pTEOE110 plasmid. Insertion and sequence was verified by colony PCR and Sanger sequencing.

4.3.19 Parasite transfections

Transfections were performed as previously described [32]. Briefly, 50-100 µg of plasmid DNA was precipitated and resuspended in Cytomix solution (25 mM HEPES pH 7.6, 120 mM KCl, 0.15 mM CaCl2, 2mM EGTA, 5 mM MgCl2, 10 mM K2HPO4).

A 4 mL, 2% hematocrit, 5% ring-stage P. falciparum culture was washed with Cytomix and resuspended in the DNA/Cytomix solution. The culture was electroporated at 950 µF and 0.31 kV, washed with media, and returned to normal culture conditions. Parasites expressing the construct were selected by continuous treatment with 5 nM WR92210 (Jacobus Pharmaceuticals, Princeton, NJ). Transfectants were cloned by limiting dilution and presence of the HAD2-GFP construct was verified by PCR using gene-specific and GFP-specific primers (5’-
AGGATATACTTTATAGATGAGAC-3' and 5'-CCGTATGTTGCATCACCTTC-3'). The presence of the endogenous HAD2 and PFK9 mutations were verified by Sanger sequencing.

### 4.3.20 Live fluorescent microscopy

Erythrocytes infected with E2 Hsp110:HAD2-GFP parasites were stained with 10 ng/µL Hoescht 33258 (Life Technologies, Carlsbad, CA) and mounted under coverslips on polysine adhesion slides. Cells were viewed on an Olympus BH8 microscope (Tokyo, Japan). Minimal adjustments to contrast and brightness were applied equally to all microscopy images using GIMP software.

### 4.3.21 Electron microscopy

*P. falciparum*-infected erythrocytes were fixed in 4% paraformaldehyde/0.05% glutaraldehyde (Polysciences Inc., Warrington, PA) in 100 mM PIPES/0.5 mM MgCl$_2$, pH 7.2 for 1 hour at 4°C. Samples were then embedded in 10% gelatin and infiltrated overnight with 2.3 M sucrose/20% polyvinyl pyrrolidone in PIPES/MgCl$_2$ at 4°C. Samples were trimmed, frozen in liquid nitrogen, and sectioned with a Leica Ultracut UCT cryo-ultramicrotome (Leica Microsystems Inc., Bannockburn, IL). Fifty nm sections were blocked with 5% fetal bovine serum/5% normal goat serum for 30 minutes and subsequently incubated with rabbit anti-GFP (ab6556) (Abcam, Cambridge, MA), goat anti-rabbit IgG (H+L) conjugated to 18 nm colloidal gold (Jackson ImmunoResearch Laboratories, Inc., West Grove, PA). Sections were washed in PIPES buffer followed by a water rinse, and stained with 0.3% uranyl acetate/2% methylcellulose. Samples were viewed with a JEOL 1200EX transmission electron microscope (JEOL USA Inc., Peabody, MA) equipped with an AMT 8 megapixel digital camera (Advanced Microscopy Techniques, Woburn, MA). All labeling experiments were conducted in parallel with controls omitting the
primary antibody. These controls were consistently negative at the concentration of colloidal gold-conjugated secondary antibodies used in these studies.

4.4 RESULTS

4.4.1 FSM\textsuperscript{R} strain E2 possesses a nonsense allele of HAD2, a homolog of the MEP pathway regulator HAD1

The MEP pathway is responsible for the synthesis of the essential isoprenoid precursors IPP and DMAPP. This pathway is specifically inhibited by the antibiotic fosmidomycin (FSM) [14,28,54]. We previously generated 20 \textit{P. falciparum} strains resistant to FSM [32]. While loss of a negative regulator, HAD1, is responsible for the resistance phenotype in 17 of these strains, 3 strains remain uncharacterized, including strain E2. We report a FSM IC\textsubscript{50} for strain E2 of 4.8 ± 1.2 µM (parental IC\textsubscript{50} ~1 µM), similar to our previously reported value for this strain [32]. This resistance phenotype is not due to changes in the \textit{HAD1} locus or changes in \textit{DXS}, \textit{DXR}, or \textit{HAD1} expression (Supplementary Figure 1, Supplementary Figure 2). We performed whole genome sequencing on this strain to identify genetic changes that may result in FSM resistance. The sequences of previously described parental and control strains [32] were used to filter SNPs also found in these strains. SNPs found in strain E2 are listed in Supplementary Data 1. We prioritized changes found in coding regions and eliminated changes found in highly variable loci such as rifins and var genes [55,56]. Among the changes identified in E2 is an A469T mutation in PF3D7_1226300 (PlasmoDB[57] ID), hereafter referred to as \textit{HAD2}. Sanger sequencing of the \textit{HAD2} locus in strain E2 confirms the presence of the A469T allele. This allele encodes a truncated R157X variant and therefore we expect HAD2 function is lost in strain E2.
Interestingly, HAD2 is a close homolog of the known MEP pathway regulator and sugar phosphatase HAD1 [32]. Sequence homology places both proteins in the haloacid dehalogenase-like hydrolase (HAD) superfamily and within this family, the Cof-like hydrolase subfamily (Interpro IPR000150) [33]. HAD2 and HAD1 protein sequences share ~29% sequence identity and ~53% sequence similarity (Supplementary Figure 3). HAD2 possesses the four conserved sequence motifs found in HAD proteins (Supplementary Figure 3).

While no structural information exists for HAD2, the structure of the *Plasmodium vivax* HAD2 (PVX_123945, PvHAD2) has been solved (PDB ID 2B30). This protein shares 93% sequence identity and 98% sequence similarity to with HAD2. PvHAD2 shows strong structural similarity to HAD1 and other HADs, with a core and cap domain structure (Supplementary Figure 4). We hypothesized that HAD2, like HAD1, regulates metabolism in *P. falciparum* malaria parasites.

### 4.4.2 HAD2 is a purine monophosphatase

Based on sequence homology to HAD1 and other HAD proteins, we predicted that HAD2 would function enzymatically as a phosphatase. The *in vitro* phosphatase activity of HAD2 was confirmed by reacting purified recombinant enzyme against *para*-nitrophenyl phosphate (*p*NPP), a promiscuous chromogenic phosphorylated substrate [32,34,58]. HAD2 was found to yield colored *p*-nitrophenyl product at $1.05 \pm 0.22 \mu$mol/min/mg enzyme.

Recent work on the HAD2 homolog from *P. vivax* (PvHAD2) has demonstrated a varied substrate profile for this enzyme that includes sugar phosphates and nucleotide phosphates [59].
We tested HAD2 against a wide range of phosphorylated substrates. Surprisingly, we find that HAD2 shows markedly lower activity against sugar phosphate substrates than HAD1 and its PvHAD2 homolog (Figure 1) [59]. In contrast, we find that HAD2 is active against a number of phosphorylated nucleotide substrates (Figure 1). In particular, HAD2 appears to prefer the purine monophosphates adenosine 5’-monophosphate (AMP) and guanosine 5’-monophosphate (GMP), while HAD1 is minimally active against nucleotides (Figure 1). Kinetic parameters for HAD2 against its top 5 substrates can be found in Supplementary Table 1.

We hypothesized that loss of HAD2-mediated purine nucleotidase function in FSMR strain E2 may result in changes in cellular AMP levels. We evaluated AMP levels using a recently described liquid chromatography- mass spectrometry (LC-MS) method [51]. We do not find that AMP levels are increased in strains lacking functional HAD2 (Supplementary Figure 5).

4.4.3 HAD2 is expressed throughout the parasite intraerythrocytic lifecycle

To determine if HAD2 was functional during a specific stage of the parasite lifecycle, we evaluated HAD2 expression throughout intraerythrocytic development. We also compared the expression profile of HAD2 to HAD1. Polyclonal anti-HAD2 antisera was generated in rabbits against recombinant HAD2. The specificities of anti-HAD1 and anti-HAD2 antisera for their respective HAD proteins were verified (see Methods).

Sorbitol-synchronized parasites were sampled at various timepoints and parasite lysates were probed by immunoblot with anti-HAD1 and anti-HAD2 antisera (Figure 2). Samples were also evaluated by microscopy to verify the developmental stage (Figure 2A). HAD1 and HAD2 are
both expressed throughout the asexual cycle, and protein expression appears to peak during the trophozoite and schizont stages (Figure 2B), consistent with RNA expression profiling of these genes [60,61].

4.4.4 FSM$^R$ had2$^{R157X}$ parasites have a growth defect that drives loss of FSM resistance

FSM$^R$ had2$^{R157X}$ parasites have impaired growth compared to the parental strain (Figure 3, purple vs. black lines). Upon long term in vitro culture (>1 month) in the absence of FSM, we observe that the FSM$^R$ had2$^{R157X}$ strain increases in growth rate. Importantly, we find that improved growth corresponds with loss of FSM resistance. We hypothesize that in FSM$^R$ had2$^{R157X}$ parasites, FSM resistance via loss of HAD2 results in a fitness cost. Strain E2 may represent a mixed population containing slow growing, FSM$^R$ parasites and fast growing, FSM-sensitive (FSM$^S$) parasites. Over time, the FSM$^S$ population thus predominates. To test this hypothesis, we cloned the E2 strain via limiting dilution and randomly selected five clones. Consistent with the presence of a mixed parasite population, three of the six clones are FSM$^R$ (designated clones R1, R2, and R3), and two are FSM$^S$ (designated S1 and S2) (Figure 4, Figure 5A). Additionally, the FSM$^S$ clones grow noticeably faster than the FSM$^R$ clones (Figure 3, teal vs. purple line). To test whether the FSM$^S$ strains had lost FSM resistance by reverting their had2$^{R157X}$ mutation, we genotyped all E2 clones at the HAD2 locus. All five E2 clones maintained loss of HAD2 via the had2$^{R157X}$ truncation mutation.
4.4.5 Long-term culture selects for suppressor mutations in PFK9

We predicted that the FSM\(^S\) E2 clones, driven by a fitness advantage, had acquired suppressor mutation(s) at an additional locus, resulting in loss of FSM resistance and an increase in growth rate. We performed whole genome sequencing on the five E2 clones to identify any genetic changes that segregated with the FSM\(^R\) and FSM\(^S\) (suppressed) phenotypes. Indeed, we find that a mutation of the locus encoding phosphofructokinase-9 (PFK9, PF3D7_0915400) is present in all the suppressed (FSM\(^S\)) E2 clones and none of the FSM\(^R\) E2 clones (Figure 4, Supplementary Figure 6, Supplementary Table 2). PFK9 is the only mutated locus that segregates with the change in FSM tolerance.

Of note, the R157X mutation in HAD2 was one of only two SNPs common to the FSM\(^R\) clones, increasing our confidence that had2\(^R_{R157X}\) is the causative change driving FSM resistance in these strains. A second common SNP was found in PF3D7_0107600, which encodes a putative protein kinase whose P. berghei homolog has been implicated in control of latency of sporozoites [62] and is not expressed in asexual stage parasites [57].

Phosphofructokinase (PFK) catalyzes the first committed and canonically rate-limiting step of glycolysis, the conversion of fructose 6-phosphate to fructose 1,6-bisphosphate. PFK9 is a single polypeptide expressed in asexual stage parasites and is comprised of two domains, designated alpha and beta [44]. In other systems (plants, yeast, humans) the alpha domain is typically regulatory, and the beta domain is catalytic [63–66]. Although we and others have been unsuccessful in purifying full-length recombinant PFK9, previous work suggests that the N-terminal beta domain has more independent catalytic activity than the C-terminal alpha domain.
The previous study describes PFK9 as being similar to PPi-utilizing PFKs from plants, although it was demonstrated to have weak ATP-utilizing activity [44]. The T1206I variant found in our E2 clones maps to the alpha domain (Supplementary Figure 6, Supplementary Table 2).

### 4.4.6 Suppressor alleles encode hypomorphic variants of PFK9

In silico analysis of the $pfk9^{T1206I}$ suppressor allele using Polyphen-2 suggests that it negatively impacts PFK9 function (Table 1) [67]. As we were unable to obtain full-length recombinant protein, we tested the effects of the suppressor mutation on PFK activity in *P. falciparum* lysate. A linked enzyme assay to measure PFK activity was implemented as previously described [44,52]. The assay is linear with respect to total protein content and is specific for parasite PFK activity (Supplementary Figure 7).

We find that strains possessing the $pfk9^{T1206I}$ mutation have significantly reduced PFK activity when compared to both the parental strain and the E2-RX clones ($PFK9^{WT}$) (Figure 5B). Interestingly, the FSM$^R$ E2 clones also show reduced PFK activity despite possessing a wild-type PFK allele. This may be explained by changes in PFK9 expression in FSM$^R$ clones.

### 4.4.7 HAD2 regulates glycolysis at PFK9

We identify hypomorphic variants in the glycolytic enzyme PFK9 that suppress FSM resistance in strains lacking HAD2. Thus, we hypothesized that HAD2 acts as a negative glycolytic regulator, possibly at the step catalyzed by PFK9. Loss of HAD2 would result in increased levels
of glycolytic and MEP pathway metabolites and FSM resistance. Reduced activity of PFK9 would negate this increase, restoring normal metabolite levels and FSM tolerance.

To test this model, we performed metabolic profiling on the parental parasite strain as well as E2 clones R1-R3 and S1-S2 (Figure 5C). Levels of the MEP pathway intermediate DOXP are significantly increased in FSM\textsuperscript{R} strains lacking HAD2 (Figure 5C, \(p < 0.05\), one-way analysis of variance (ANOVA)). While not statistically significant for all strains, we also observe increases in steady-state MEcPP levels in our FSM\textsuperscript{R} strains. As DOXP is the substrate for the FSM target enzyme DXR, our data are consistent with FSM\textsuperscript{R} strain E2 achieving FSM resistance via increased levels of DOXP counteracting competitive inhibition of DXR by FSM, similar to the mechanism previously proposed for HAD1 [32].

In order to better understand the role of PFK9 in suppressing FSM resistance in strains lacking functional HAD2, we measured the steady-state levels of a number of glycolytic intermediates, including fru6P and FBP, the substrate and product of PFK9 (Figure 5C). We observe that FSM\textsuperscript{R} resistant E2 clones possessing the \(had2^{R157X}\) allele have increased levels of FBP. This is consistent with HAD2-mediated negative regulation acting at the PFK step.

Furthermore, strains possessing the \(pfk9^{T1206I}\) suppressor allele display nearly parental levels of glycolytic and MEP pathway intermediates (Figure 5C). This is consistent with the hypothesis that hypomorphic alleles of \(PFK9\) are able to suppress increased metabolite levels caused by loss of HAD2.
Additionally, FBP levels are highly correlated with the levels of MEP pathway intermediates DOXP and MEcPP (Pearson r = 0.93 and 0.99, respectively; p < 0.001). This is not the case for the upstream metabolites glu6/fru6P, which shows little correlation with DOXP or MEcPP (Pearson r < 0.7, p > 0.05). DOXP, MEP, and FBP levels are all significantly correlated with FSM IC\textsubscript{50} (Pearson r > 0.8, p < 0.02).

4.4.8 Suppression and complementation support model of HAD2- and PFK9-mediated metabolic regulation

Our model of PFK9-mediated suppression of FSM resistance in strains with the $had2^{R157X}$ allele predicts that a fitness defect selects for acquisition of suppressor mutations in PFK9. To further test this hypothesis, we grew the three FSM\textsuperscript{R}E2 clones ($had2^{R157X}$, $PFK9^{WT}$) without FSM for >1 month (Figure 4). As previously observed, these strains also lost their FSM resistance phenotype (Supplementary Table 2) and exhibited increased growth rates (Figure 3). We sequenced $HAD2$ and $PFK9$ in these three suppressed strains. Indeed, all maintained the $had2^{R157X}$ mutation and had acquired new, independent $PFK9$ mutations (Figure 4, Supplementary Figure 5, Supplementary Table 2), which correlated with increased growth rate and FSM sensitivity (Supplementary Table 2). As predicted, these strains also demonstrate reduced PFK activity in our lysate assay (Supplementary Figure 8).

Our model also predicts that restoration of functional HAD2 in FSM\textsuperscript{R} strain E2 will restore FSM sensitivity. To test this hypothesis, we used the piggyBac transposon system to express GFP-
tagged HAD2, driven by the heat shock protein 110 (Hsp110) promoter in strains with the had2\textsuperscript{R157X} allele [32,68,69].

We were unable to obtain successful transfectants from the FSM\textsuperscript{R} clones (had2\textsuperscript{R157X}, PFK9), perhaps because of these strains possess decreased overall fitness. However, we were able to rescue loss of HAD2 in a suppressed FSM\textsuperscript{S} E2 clone (had2\textsuperscript{R157X}, pfk9\textsuperscript{T1206I}) (Supplementary Figure 9A). The transformed strain was confirmed to maintain its endogenous HAD2 and PFK9 mutations, as well as successful expression of exogenous HAD2-GFP (Figure 3, Supplementary Figure 9B). Consistent with our model of HAD2 as a negative regulator of glycolysis and the MEP pathway, expression of HAD2-GFP in had2\textsuperscript{R157X}, pfk9\textsuperscript{T1206I} parasites results in increased sensitivity to FSM compared to both the suppressed and parental parasite strains (Figure 5D, Supplementary Figure 9C).

The creation of HAD2-GFP parasites allowed us to assess the in vivo substrates available to HAD2 via live fluorescence microscopy. We observe that HAD2-GFP is localized throughout the cytoplasm in asexual *P. falciparum* trophozoites and schizonts (Figure 6). HAD2 appears to be excluded from the food vacuole. This is consistent with the lack of a predicted signal sequence for HAD2 using SignalP, PlasmoAP, and PlasMit algorithms [70–72]. Like HAD1 [32], HAD2 does not appear to primarily concentrate in the apicoplast, the site of isoprenoid synthesis by the MEP pathway.
4.5 DISCUSSION

Using resistance to the inhibitor FSM, we identify the purine phosphatase HAD2 as a novel regulator of parasite metabolism. Parasites lacking HAD2 exhibit increased steady-state levels of glycolytic and MEP pathway intermediates that allow them to overcome FSM inhibition. HAD2 appears to be necessary for parasite fitness, and we select for mutations that suppress the HAD2-mediated FSM resistance and growth defect. We find that mutations in phosphofructokinase \((PFK9)\) restore wild-type growth rates and FSM sensitivity, connecting HAD protein function to regulation of essential central carbon metabolism in the parasite.

We observe strong correlation between levels of MEP pathway metabolites DOXP and MEcPP and FBP, the product of PFK9 (Pearson \(r > 0.9, p < 0.001\)). This is consistent with the lack of a rate-limiting enzymatic step between PFK and the MEP pathway enzyme DXR. Similar to what has been described in \(E.\ coli\) [73], FBP appears to be able to function as an indicator of metabolic flux, which is subject to perturbations from drug treatment or genetic mutation. A recent study described FBP-centered metabolic regulation in the related apicomplexan \(Toxoplasma gondii\), which constitutively expresses and regulates the gluconeogenic enzyme fructose 1,6-bisphosphatase (FBPase) to fine-tune its glucose metabolism [74]. While \(P. falciparum\) does not appear to possess an FBPase necessary for gluconeogenesis, it is possible that the parasite possesses FBP-sensing mechanisms for tuning its metabolism, perhaps via regulators such as HAD1 and HAD2.

HAD2 is a member of the HAD superfamily and a homolog of the previously described
metabolic regulator HAD1 [32]. Both enzymes belong to the Cof-like hydrolase subfamily (IPR000150) [33], which mainly comprises proteins from bacteria and plastid-containing organisms (plants and apicomplexans). These enzymes display wide and varied substrate preference [34,37,75–79], although the function of most members remains unknown. Along with our previous studies on HAD1 [32,80], we define the role of these proteins in *P. falciparum* and contribute to the greater understanding of this evolutionarily conserved and important protein family.

Our approach demonstrates the power of a forward genetics approach to uncover novel biology in a clinically relevant, non-model organism. We employ a previously described screen for FSM resistance [32] to uncover a novel resistance mutation and employ a second selection for parasite fitness to identify changes that suppress our resistance phenotype. While fitness costs associated with antimalarial resistance mutations have long been observed (summarized in [81,82]), our study represents, to our knowledge, the first to utilize this fitness defect to identify suppressor mutations in a non-target locus. Fitness assessment of resistance mutations may allow for suppressor screening for other antimalarials and other target pathways.

Resistance screens in malaria are often used for inhibitor target identification [83–88]. However, we utilize relatively low drug selection concentrations (≤10X IC₅₀) and a high sample number to uncover mutations in non-target loci that regulate essential biochemical pathways in the parasite. Indeed, our screen did not uncover any genetic changes in the FSM target, DXR. This underscores the importance of interpreting the results of a resistance selection with caution and the necessity of genetically validating resistance mutations.
The novel points of metabolic regulation we uncover in this study may serve as targets for future antimalarial development. We describe a novel resistance mutation in HAD2 that results in a significant decrease in parasite fitness. This is consistent with the recent observation that loss of the HAD2 homolog in Plasmodium berghei results in a significant growth defect during infection [89]. Thus, HAD2 mutations are unlikely to be biomarkers of clinical FSM resistance. However, HAD2 is an excellent target for development of future antimalarials. HAD2 lacks close homologs in mammals. Recombinant HAD2 is easily expressed and assayed and its use of common cellular metabolites (purines) lends itself to informed inhibitor screening and optimization.

In addition to HAD2, our work highlights the glycolytic enzyme PFK9 as an important point of regulation in essential parasite glucose metabolism. While the parasite is able to tolerate mutation to the PFK9 locus in a had2R157X background, these mutations likely come with a fitness defect in wild-type parasites. A kinetic model of parasite glycolysis shows that PFK has a high flux control coefficient, is sensitive to competitive inhibition, and is effective at reducing flux through glycolysis [90,91], making it a leading candidate for targeting essential glucose metabolism in the parasite. Like HAD2, PFK9 is plant-like and divergent from its mammalian homologs [44]. These differences may be exploited for PFK inhibitor design.
4.6 FIGURES

Figure 1. HAD2 is a purine monophosphatase with a unique substrate specificity profile.

Displayed are the means +/- standard error of the mean (S.E.M.) of the specific enzyme activity for HAD1 (A) and HAD2 (B) from at least three independent experiments. HAD1 activity against sugar phosphates was previously described [32]. Sugar phosphates (white bars) are ordered from left to right by increasing number of carbon atoms (3 – 15). Nucleotides (black bars) are ordered from left to right by increasing degree of phosphorylation. Abbreviations: Glc2P, glycerol 2-phosphate; Glc1P, glycerol 1-phosphate; Gly3P, glyceraldehyde 3-phosphate; DHAP, dihydroxyacetone phosphate; Glu1P, glucose 1-phosphate; PEP, phosphoenolpyruvate; 2-PGA, 2-phosphoglyceric acid; Ery4P, erythrose 4-phosphate; Ribu5P, ribulose 5-phosphate;
DOXP, deoxyxylulose 5-phosphate; Rib5P, ribose 5-phosphate; 2drib5P, 2-deoxyribose 5-phosphate; MEP, methylerithritol phosphate; Fru1P, fructose 1-phosphate; Man6P, mannose 6-phosphate; Glu6P, glucose 6-phosphate; Fru6P, fructose 6-phosphate; Fru1,6bisP, fructose 1,6-bisphosphate; Gal1P, galactose 1-phosphate; Sorb6P, sorbitol 6-phosphate; Man1P, mannose 1-phosphate; Sedo7P, sedoheptulose 7-phosphate; Tre6P, trehalose 6-phosphate; Pyr5P, pyridoxal 5’-monophosphate; GMP, guanosine 5’-monophosphate; AMP, adenosine 5’-monophosphate; XMP, xanthosine 5’-monophosphate; IMP, inosine 5’-monophosphate; CMP, cytidine 5’-monophosphate; dCMP, 2’-deoxycytidine 5’-monophosphate; dGMP, 2’-deoxyguanosine 5’-monophosphate; TMP, thymidine 5’-monophosphate; UMP, uridine 5’-monophosphate; dUMP 2’-deoxyuridine 5’-monophosphate; ADP, adenosine 5’-diphosphate; TDP, thymidine 5’-diphosphate; UDP, uridine 5’-diphosphate; CDP, cytidine 5’-diphosphate; GDP, guanosine 5’-diphosphate; CTP, cytidine 5’-triphosphate; ATP, adenosine 5’-triphosphate.
Figure 2. HAD2 and HAD1 are expressed throughout the intraerythrocytic parasite lifecycle.

(A) Synchronized 3D7 parental parasites were sampled at various timepoints (0-48 hours) throughout a single asexual lifecycle. (B) Samples were probed for HAD1 and HAD2 expression using polyclonal antisera. Expression peaks during the trophozoite and schizont stages. Data are representative of at least three independent experiments.
Figure 3. FSM resistance results in a fitness cost.

A representative FSM\(^R\) clone with the had\(^{2}\text{R}^{157X}\) allele (R1, purple line) has a reduced growth rate compared to the wild-type parental strain (black). The growth defect is rescued in two representative clones with mutations in \(PFK9\) (S1 and S3, teal lines). Error bars represent S.E.M. from at least three independent experiments.
Figure 4. Model of parasite populations and genetic changes that modulate FSM sensitivity.

Parasites are colored by their FSM tolerance phenotype (teal for FSM$^S$, purple for FSM$^R$).

Cloned strains are named by FSM tolerance phenotype (E2-SX, sensitive; E2-RX, resistant). A FSM$^S$ parental strain was selected under FSM pressure to enrich for FSM$^R$ strain E2 ($had2^{R157X}$).

After relief of FSM pressure, a fitness advantage selects for suppressor mutations in PFK9 ($pfk9^{mut}$, yellow star) that result in FSM sensitivity. Upon cloning, FSM$^R$ clones are grown without FSM pressure. As before, fitness advantage selects for suppressor mutations in PFK9 that result in increased growth rate and loss of FSM resistance.
Figure 5. HAD2 and PFK9 alleles alter FSM resistance, PFK activity, and metabolite levels in *P. falciparum*.

(A) FSM IC$_{50}$s for parental (par) strain and E2 clones. Colors represent the log$_2$ of the fold change (FC) over the parental strain, as indicated in the legend. Cloned strains are named by FSM tolerance phenotype (SX, sensitive; RX, resistant). Data are representative of at least three independent experiments. (B) Measurement of PFK activity of *P. falciparum* lysate indicates that E2-SX clones with PFK9 suppressor mutations have reduced PFK activity. Error bars represent
S.E.M. from at least three independent experiments. P-values were determined using a one-way analysis of variance (ANOVA; Tukey’s post-test). * = P < 0.05, *** = P < 0.001, **** = P < 0.0001. (C) LC-MS/MS metabolic profiling of glycolytic and MEP pathway intermediates of parental and E2 clone strains confirms decreased PFK activity (FBP levels) in E2-SX strains with PFK9 suppressor mutations. FSMR strains achieve resistance via increased levels of MEP pathway intermediates, which correlate with FBP. Colors represent the log2 of the fold change (FC) over the parental strain, as indicated in the legend. Data are representative of at least two independent biological replicates each composed of at least two technical replicates. Isomer pairs that cannot be confidently distinguished are indicated (glu6P/fru6P and DHAP/gly3P). (D) When loss of HAD2 is rescued in had2R157X, pfk9T1206I parasites, the resulting strain is over-sensitive to FSM, due to having functional HAD2 and a hypomorphic PFK9 allele.
Figure 6. HAD2 is localized to the cytoplasm.

Shown is live microscopy of E2 HAD2-GFP parasites, stained with Hoescht 33258 nuclear stain.

Scale bars, 2 µm.
Supplementary Figure 1. FSM\textsuperscript{R} strain E2 does not have increased levels of \textit{DXS} or \textit{DXR} transcripts.

Shown are relative expression levels calculated from duplicate independent experiments comprised of three technical replicates. Error bars represent S.E.M. N.s. = not significant, $P > 0.05$, by an unpaired Student’s t-test.
Supplementary Figure 2. HAD1 expression is unchanged in FSM\textsuperscript{R} strain E2.

Shown are immunoblots of the parent strain and FSM\textsuperscript{R} strain E2. Marker units are kilodaltons (kDa). Top panel was probed with anti-HAD1 antisera. Bottom panel was probed with anti-heat shock protein 70 antisera as a loading control. Expected protein masses: HAD1, 33 kDa; Hsp70, 74 kDa. Blot is representative of at least three independent experiments.
Supplementary Figure 3. HAD2 is a homolog of the MEP pathway regulator HAD1. Alignment was produced using Clustal Omega [92] and BoxShade. Residues highlighted in black are identical. Residues highlighted in grey are similar. HAD1 and HAD2 share ~29% sequence identity and ~53% sequence similarity. Like HAD1, HAD2 possesses the four sequence motifs found in HAD proteins (blue boxes).
Supplementary Figure 4. *P. vivax* HAD2 is structurally similar to *P. falciparum* HAD1.

PDB structures were compared using rigid-body alignment (TM-align) in Lasergene Protean 3D software. *P. vivax* HAD2 (PDB ID 2B30) is shown in ochre and *P. falciparum* HAD1 (4QJB) is shown in blue. Ions (Mg$^{2+}$, Ca$^{2+}$, Cl$^{-}$) are shown in green. Cap domains are indicated by lighter coloring.
Supplementary Figure 5. AMP levels are unchanged in E2 FSMR parasites.

Data shown are the result of three independent experiments. Error bars represent S.E.M. N.s. = not significant (p=0.95, unpaired Student’s t-test). Y-axis units are arbitrary, computed from the area under the mass spectrometer signal and normalized to an internal standard and cell number.
Supplementary Figure 6. Map of suppressor mutations in *PFK9*.

Mutations are designated by black lines and labeled with the resulting amino acid change. The 1418 amino acid protein is composed of two domains, designated beta and alpha, which are connected by a short linker.
Supplementary Figure 7. Assay of PFK activity from *P. falciparum* lysate is linear and specific.

(A) Schematic of linked enzyme assay of PFK activity. PFK catalyzes the phosphorylation of fructose 6-phosphate (fru6P) to fructose 1,6-bisphosphate (FBP). Excess linking enzymes aldolase (ALD), triose-phosphate isomerase (TPI), and glycerol 3-phosphate dehydrogenase (GDH) produce glycerol 3-phosphate. NADH utilization is monitored spectrophotometrically at 340 nm. (B) The assay is linear with respect to total lysate protein content. Error bars represent S.E.M. of two technical replicates. (C) The assay is specific for *P. falciparum* lysate-dependent PFK activity. Error bars represent S.E.M. of three independent experiments.
Supplementary Figure 8. Suppressor mutations in PFK9 result in reduced activity.

Measurement of PFK activity of *P. falciparum* lysate indicates that the second set of E2-SX clones with PFK9 suppressor mutations have reduced PFK activity, compared to parental (par) and FSMR strains (RX, as shown in Figure 4B). Error bars represent S.E.M. from at least two independent experiments. P-values were determined using a one-way analysis of variance (ANOVA; Tukey’s post-test). * = P < 0.05, ** = P < 0.01, **** = P < 0.0001.
Supplementary Figure 9. Rescue of HAD2 confirms role as a negative regulator.

(A) Successful transfection of pTEOE110:HAD2-GFP in strain S1 (had2<sup>R157X</sup>, pfk9<sup>T1206I</sup>) was confirmed by PCR using construct-specific primers. Marker units are in base pairs (bp).

Expected size of amplicon is 738 bp. (B) Immunoblot confirming successful expression of HAD2-GFP in strain S1. Marker units are kilodaltons (kDa). The top blot was probed with anti-HAD2 antisera. The bottom blot was probed with anti-heat shock protein 70 antisera as a loading control. (C) Representative IC<sub>50</sub> graph showing expression of HAD2-GFP in strain S1 (had2<sup>R157X</sup>, pfk9<sup>T1206I</sup>) results in oversensitivity to FSM. The parental strain has a FSM IC<sub>50</sub> of 0.8 ± 0.1 µM, and S1 has an IC<sub>50</sub> of 0.5 ± 0.08 µM. When loss of HAD2 is rescued in strain S1, the resulting strain has an IC<sub>50</sub> of 0.3 ± 0.1 µM. Data are means and S.E.M.s of at least three independent experiments.
Supplementary Figure 10. Anti-HAD1 and anti-HAD2 polyclonal antisera are specific for their target HAD antigens.

Recombinant 6xHis-HAD1 and 6xHis-HAD2 probed with (A) rabbit anti-HAD1 polyclonal antisera [32] and (B) rabbit anti-HAD2 polyclonal antisera.
**4.7 TABLES**

**Supplementary Table 1. Kinetic parameters for HAD2.**

Shown are the means ± standard error of the mean (S.E.M.) of at least three independent experiments. Abbreviations used: AMP, adenosine 5’-monophosphate; GMP, guanosine 5’-monophosphate; 2dGMP, 2’-deoxyguanosine 5’-monophosphate; XMP, xanthosine 5’-monophosphate; IMP, inosine 5’-monophosphate.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>$K_m$ (mM)</th>
<th>$K_{cat}$ (sec$^{-1}$)</th>
<th>$k_{cat}/K_m$ (M$^{-1}$ sec$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMP</td>
<td>1.0 ± 0.2</td>
<td>1.7 ± 0.3</td>
<td>$1.7 \times 10^3$</td>
</tr>
<tr>
<td>GMP</td>
<td>0.30 ± 0.07</td>
<td>2.0 ± 0.1</td>
<td>$6.9 \times 10^3$</td>
</tr>
<tr>
<td>2dGMP</td>
<td>0.30 ± 0.01</td>
<td>3.2 ± 0.5</td>
<td>$11.0 \times 10^3$</td>
</tr>
<tr>
<td>XMP</td>
<td>0.73 ± 0.09</td>
<td>3.6 ± 0.1</td>
<td>$5.0 \times 10^3$</td>
</tr>
<tr>
<td>IMP</td>
<td>8.4 ± 0.6</td>
<td>2.5 ± 0.1</td>
<td>$0.30 \times 10^3$</td>
</tr>
</tbody>
</table>
Supplementary Table 2. Genotypes of strains used to identify \textit{HAD2} and \textit{PFK9} mutations influencing FSM sensitivity.

E2 represents the uncloned strain that resulted from FSM selection. E2 clones are indicated by the E2-XX format, with resistant strains designated E2-RX and sensitive strains designated as E2-SX. Genotypes are indicated by their resulting amino acid change. Predicted effect of amino acid changes were determined using Polyphen-2 [67].

<table>
<thead>
<tr>
<th>Strain</th>
<th>\textit{HAD2} variant</th>
<th>\textit{PFK9} variant</th>
<th>\textit{PFK9} mutation Polyphen-2 score</th>
<th>FSM IC\textsubscript{50} (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>parent</td>
<td>WT</td>
<td>WT</td>
<td>NA</td>
<td>0.8 ± 0.1</td>
</tr>
<tr>
<td>E2</td>
<td>R157X</td>
<td>WT</td>
<td>NA</td>
<td>2.2 ± 0.4</td>
</tr>
<tr>
<td>E2-R1</td>
<td>R157X</td>
<td>WT</td>
<td>NA</td>
<td>4.3 ± 1.2</td>
</tr>
<tr>
<td>E2-R2</td>
<td>R157X</td>
<td>WT</td>
<td>NA</td>
<td>4.0 ± 0.2</td>
</tr>
<tr>
<td>E2-R3</td>
<td>R157X</td>
<td>WT</td>
<td>NA</td>
<td>5.5 ± 0.7</td>
</tr>
<tr>
<td>E2-S1</td>
<td>R157X</td>
<td>T1206I</td>
<td>0.915 (possibly damaging)</td>
<td>0.5 ± 0.08</td>
</tr>
<tr>
<td>E2-S2</td>
<td>R157X</td>
<td>T1206I</td>
<td>0.915 (possibly damaging)</td>
<td>0.6 ± 0.08</td>
</tr>
<tr>
<td>E2-S3</td>
<td>R157X</td>
<td>S335L</td>
<td>0.998 (probably damaging)</td>
<td>0.7 ± 0.4</td>
</tr>
<tr>
<td>E2-S4</td>
<td>R157X</td>
<td>N1359Y</td>
<td>0.974 (probably damaging)</td>
<td>0.2 ± 0.06</td>
</tr>
<tr>
<td>E2-S5</td>
<td>R157X</td>
<td>S1267L</td>
<td>0.663 (possibly damaging)</td>
<td>0.2 ± 0.05</td>
</tr>
</tbody>
</table>
4.8 REFERENCES


27. Steinbacher S, Kaiser J, Eisenreich W, Huber R, Bacher A, Rohdich F. Structural basis of


39. Kang Y, Weber KD, Qiu Y, Kiley PJ, Blattner FR. Genome-wide expression analysis indicates that FNR of Escherichia coli K-12 regulates a large number of genes of

188


42. Kemp RG, Gunasekera D. Evolution of the allosteric ligand sites of mammalian phosphofructo-1-kinase. 2002;


Chapter 5

Conclusion and future directions
5.1 SUMMARY

The human malaria parasite, *Plasmodium falciparum*, requires a constant supply of glucose for synthesis of ATP and essential secondary metabolites, including isoprenoids. We utilized resistance to the isoprenoid synthesis inhibitor fosmidomycin (FSM) to uncover novel mechanisms of drug resistance and metabolic regulation in the parasite. In this work, we present the discovery and characterization of PfHADs as a novel class of parasite metabolic regulators. The PfHADs are members of the haloacid dehalogenase-like hydrolase (HAD) superfamily, a large and diverse protein class found in all kingdoms of life. In *P. falciparum*, PfHADs function as negative regulators, whose loss results in upregulation of essential metabolism and increased drug tolerance. We find that PfHADs are phosphatases that hydrolyze intracellular phosphometabolites, including sugar phosphates and nucleotides, and define the genetic, enzymatic, structural, and metabolic basis for resistance and regulation by these proteins. Additionally, we evaluate the genetic changes following FSM treatment failure (recrudescence) in human patients and find that PfHADs are not biomarkers of recrudescence.

5.2 SUBSTRATE SPECIFICITY OF THE HAD SUPERFAMILY

We define the function of two HAD family members in the malaria parasite, *P. falciparum*. Both HAD1 and HAD2 belong to the Cof-like hydrolase subfamily (Interpro IPR000150) [1], composed mostly of bacterial proteins. While the cellular function of most members remains unknown, several bacterial HADs have been implicated in metabolic regulation [2–6].
Enzymatic and structural studies of HADs have found them to have wide and varied substrate profiles [7–9], leading to the hypothesis that HADs are platforms for enzyme evolution [10,11]. This flexibility in substrate specificity may be harnessed to modify or create enzymes to produce desirable metabolites. Bacterial homologs of PfHADs have been found to increase production of metabolites such as lycopene, ribose, and amino acids [2–4]. However, few efforts have been made to alter the substrate specificities of these enzymes to further increase their utility.

Our findings support the notion that, even outside of bacteria, Cof-like HAD family members have wide and varied substrate profiles. Thus, given the structural and functional information generated from our studies and others [12–14], Plasmodium HADs are an effective platform to study determinants of substrate specificity in this enzyme family. Findings may be applied to homologous proteins in biological systems applicable to bioengineering, such as bacteria or algae.

HAD1 and HAD2 share ~29% sequence identity and ~53% sequence similarity (Chapter 4, Supplementary Figure 3). Interestingly, the P. falciparum genome also encodes a third Cof-like HAD protein, designated as HAD3 (PlasmoDB ID PF3D7_1226100) [15]. HAD3 shares >20% identity and >40% similarity with both HAD1 and HAD2. HAD3 is primarily expressed in the gametocyte (sexual) stages of the parasite lifecycle and the in vivo function has not yet been investigated for this protein [15,16]. However, studies from our laboratory find that recombinant HAD3 is a phosphatase capable of utilizing both sugar phosphates and nucleotides as substrates (data not shown). The overlapping substrate profiles of HAD1, HAD2, and HAD3 and their close sequence similarity are ideal for a comparative approach to enzyme engineering. This
approach would be aided by the previous structural characterization of HAD1, including the description of several substrate-bound structures (Appendix A) [12,13], as well as *Plasmodium vivax* HAD2 [14].

### 5.3 THE CELLULAR FUNCTION OF HAD1

We find that HAD1 functions as a sugar phosphatase (Chapter 2, Figure 6) with access to substrates upstream of the MEP pathway. Electron microscopy confirms the primarily cytosolic localization of HAD1 observed by fluorescent microscopy of live and immunostained parasites (Chapter 2, Figure 7 and Supplementary Figure 4). However, we also observe a small fraction of HAD1 in the lumen and membrane of the apicoplast, the site of isoprenoid synthesis via the MEP pathway (Appendix B, Figure B2). Membrane-associated HAD1 is not detectable via cellular fractionation (Appendix B, Figure B3).

Given the possible membrane localization of HAD1, we tested the ability of the recombinant protein to bind membrane lipids. In a lipid-binding assay, HAD1 does not appear capable of binding the major lipid species of the apicoplast membrane [17], including phosphatidylcholine (PC), phosphatidylethanolamine (PE), and phosphatidylinositol (PI). However, we find that HAD1 is capable of binding phosphatidylinositol phosphates (PIPs) (Appendix B, Figure B4). Interestingly, HAD1 does not appear to utilize PIPs or their cognate inositol phosphates as substrates in vitro (Appendix B, Figure B5). Profiling of PIPs from parental and FSMR strains will illuminate whether HAD1 has any effect on PIP levels in vivo. PIPs are well-established second messengers in other systems [18], and in *P. falciparum*, PIP-binding proteins have been
implicated in parasite development, drug resistance, and trafficking of hemoglobin to the food vacuole [19–22]. Further experiments are required to determine if the small fraction of apicoplast-localized HAD1 performs a meaningful biological function and if and how this function may be related to regulation of the MEP pathway or central carbon metabolism. It may be possible to create HAD1 mutants that are incapable of binding PIPs but maintain sugar phosphatase activity. These HAD1 mutants could be assayed for their ability to restore FSM sensitivity and metabolite levels in FSM\textsuperscript{R} strains lacking HAD1.

Given HAD1’s broad substrate specificity and primarily cytosolic localization, specific function or in vivo substrate specificity of HAD1 may be conferred by its interaction with other proteins. Immunoprecipitation of HAD1 may identify interactors that point to regulation of a specific pathway or metabolite by HAD1. Genetic manipulation of these interactors may reveal HAD1-related phenotypes, such as increased metabolite levels and FSM resistance.

Our model predicts that HAD1 dephosphorylates intermediates of glycolysis, upstream of the MEP pathway (Chapter 2, Figure 9). A number of metabolic profiling approaches can be used to further test this model. LC-MS/MS measurement of glycolytic intermediates in FSM\textsuperscript{R} parasite strains with loss of HAD1 function and FSM\textsuperscript{R} strains rescued with Hsp110:HAD1-GFP may identify changes in glycolytic intermediates attributable to loss of HAD1. However, this approach is limited to the measurement of steady-state levels of a focused set of metabolites. Unbiased profiling of a broader class of phosphorylated metabolites may identify other unexpected changes due to loss of HAD1 and HAD2. Additionally, metabolic flux analysis using isotope-labeled sugars may provide insight on reaction rates affected by HAD1 or HAD2.
5.4 HAD2 FUNCTION AND METABOLIC REGULATION IN P. FALCIPARUM

While we define HAD1 and HAD2 as metabolic regulators, there is still much to learn regarding the mechanism of this regulation in the parasite. In particular, it is not yet clear how HAD2 is able to influence levels of glycolytic intermediates, namely the PFK9 product fructose 1,6-bisphosphate (FBP). As a purine phosphatase, HAD2 may modulate levels of canonical PFK regulators, such as AMP. In a previous study using recombinant PFK9 beta domain, AMP was found to have a modest, albeit variable, activating effect on catalytic activity [23]. Further experiments will be required to understand if AMP exerts this effect on full-length enzyme, which would include the potentially regulatory alpha domain. Additionally, it would be informative to test PFK9 activity in the presence of other HAD2 substrates, such as GMP, XMP, IMP, and dGMP (Chapter 4, Figure 1), as these compounds may function as non-canonical allosteric regulators whose levels are modulated by HAD2. These efforts are currently limited by our inability to successfully express and purify full-length recombinant PFK9.

We find that mutations in PFK9 are able to suppress FSMR resistance and metabolite changes caused by loss of HAD2. We find that these PFK9 variants have reduced catalytic activity in assays using P. falciparum lysate and result in reduced levels of the PFK product, FBP (Chapter 4, Figure 5). The exact mechanism of this suppression is unknown, but may be informative to HAD2 function. While PFK9 mutants may be achieving suppression via gross reduction in PFK activity, it is also possible that these mutations affect a specific function of PFK, such as
activation or binding by an allosteric regulator. This model predicts that mutations that successfully suppress the FSM\textsuperscript{R} phenotype will be clustered, perhaps near an allosteric binding site. Generation of additional \textit{PFK9} suppressor mutations and structural studies of these mutations can be used to test this hypothesis.

Our work identifies the \textit{PFK9} product, FBP, as highly correlated with MEP pathway metabolites (Chapter 4, Figure 5). This finding is consistent with a recent study identifying \textit{PFK9} as having a high glycolytic flux control coefficient, meaning that perturbations to this enzymatic step have the greatest impact on pathway flux [24,25]. As FBP appears to be an indicator of parasite glycolytic and MEP pathway metabolite levels, it is possible that the parasite is utilizing this indicator to monitor and respond to changes in its metabolism. FBP-sensing mechanisms have been described in \textit{E. coli} [26], which possesses an FBP-responsive transcription factor, Cra (FruR), that drives expression of a number of genes related to central carbon metabolism [26–28]. The Apicomplexan \textit{Toxoplasma gondii} modulates expression of a fructose 1,6-bisphosphatase (FBPase) for regulation of its glucose metabolism [29].

Further experiments are required to understand the existence and role FBP-responsive factors in \textit{P. falciparum}. Transcriptional profiling of HAD1 and HAD2 mutant parasites may reveal genes that are differentially expressed upon relief of negative glycolytic regulation by HADs. However, \textit{P. falciparum} transcriptional control has been shown to be driven by a limited suite of transcription factors in a robotic fashion correlating with protein function during development. It has been hypothesized that post-transcriptional regulation may instead play a more dominant role in controlling parasite gene expression [30–32].
Perhaps more likely is the possibility that FBP may function in enzymatic regulation. FBP is known to activate the downstream glycolytic enzyme pyruvate kinase in other systems [33]. However, *P. falciparum* pyruvate kinase was found to be insensitive to activation by FBP [34]. Further studies are required to understand if FBP regulates other glycolytic enzymes. Instead, HAD1 or HAD2 may be sensitive to regulation by FBP, which may be easily tested using in vitro enzyme assays. Immunoprecipitation of protein interactors HAD2 and PFK9 may reveal enzymes or pathways subject to regulation by FBP.

### 5.5 TARGETING HAD1, HAD2, AND PFK9

The screen for FSM resistance yielded 17 parasite strains possessing 13 deleterious *HAD1* alleles that result in loss of protein function. We hypothesized that loss of HAD1 has a negligible effect on parasite fitness in culture and that *HAD1* mutations are thus easy to acquire. Indeed, we find that loss of HAD1 does not result in a significant growth defect compared to the parental strain (Appendix B, Figure B1). This observation is consistent with in vivo growth assays of genetic mutants in *Plasmodium berghei* infections [35]. Thus, it appears that HAD1 is dispensable for asexual stage growth. However, little variation is seen in this gene in clinical isolates, even after FSM treatment [15, 23] (Chapter 3, Figure 3). It is possible that loss of HAD1 has deleterious effects on parasite fitness during human or insect infection.

In contrast, our screen produced only one strain with a deleterious mutation in *HAD2*. *HAD2* mutants appear to have an appreciable fitness defect, as HAD2 mutants grow considerably
slower than the parental strain (Chapter 4, Figure 3). This growth defect has also been observed for *HAD2* mutants in *P. berghei* infections [35]. Further and more detailed profiling of the lifecycle of HAD2 mutants is necessary to illuminate stage-specific growth effects of HAD2 loss.

Given its necessity for parasite fitness in asexual stage growth, HAD2 is a candidate target for antimalarial development. Like the MEP pathway enzymes, HAD2 does not have a close homolog in humans, reducing concerns regarding the host toxicity of HAD2 inhibitors. Screening efforts would be greatly aided by the ease with which recombinant HAD2 is expressed and assayed. Presumably, parasites may acquire resistance to HAD2 inhibition via mutation of *PFK9*. In vivo fitness assessments of *had2 pfk9* parasites are necessary to determine if this is likely to occur during infection.

PFK9 is also an attractive target for antimalarial development. Given the necessity of glycolysis in the parasite, PFK9 is likely essential. Using kinetic models, PFK9 has been identified as an effective step to reduce glycolytic flux [24,25]. PFK9 possesses low sequence similarity to the human enzyme [23]. Due to its similarity to plant PPi-PFKs and the parasite’s unique metabolism, PFK9 may also be subject to regulatory mechanisms distinct from the host enzyme which may be exploited for inhibition. Further work is required to develop an assay of PFK9 that is efficient for inhibitor screening, as recombinant PFK9 has not been successfully purified [23].
5.6 FOSMIDOMYCIN AS AN ANTIMALARIAL

Fosmidomycin’s (FSM) promise as an antimalarial has been dampened by high rates of recrudescence [37,38]. However, FSM possesses a number of valuable characteristics as an antimalarial: it is well-validated, highly specific [39,40], and non-toxic in humans [41,42]. Previous clinical trials (clinicaltrials.gov identifiers NCT02198807, NCT01361269, NCT01002183, NCT00214643, and NCT00217451) used the antibiotic clindamycin (CLN) as a partner agent for FSM, due to early findings of synergy between the two drugs [41,43,44]. Both FSM and CLN exhibit short serum half-lives (1-3 hours) [45,46]. The short exposure time to effective levels of both drugs likely contributes to the reduced treatment efficacy observed in these trials. We do not find a genetic signature of recrudescence following FSM treatment failure (Chapter 3). This finding supports the hypothesis that FSM treatment failure is not due to resistance-causing mutations and instead is primarily due to partner drug selection.

FSM is currently being tested in a clinical trial (NCT02198807), where it is partnered with piperaquine, a bisquinolone with a long serum half-life (>20 days) [47] that is also being evaluated as part of an artemisinin combination therapy (ACT) [48]. Results of this trial will show whether FSM treatment is more effective when partnered with a drug with a long half-life. Additionally, a number of groups are pursuing chemical modification of FSM to improves its potency and half-life [49–51]. We encourage further studies on the efficacy of FSM and the development of other MEP pathway inhibitors.
5.7 FINAL THOUGHTS

The causative agent of human malaria, *P. falciparum*, remains a threat to global health. Developing resistance to current therapies remains an ever-present challenge to malaria eradication. An understanding of drug resistance mechanisms as well as novel target pathways is necessary for the development of future antimalarials. In this work, we have described the mechanism of resistance to fosmidomycin, an inhibitor of essential isoprenoid synthesis in the parasite. We uncovered the PfHADs, a novel class of parasite metabolic regulators. Additional studies will continue to define the role of these proteins in parasite biology.
5.8 REFERENCES


26. Kochanowski K, Volkmer B, Gerosa L, Haeverkorn van Rijsewijk BR, Schmidt A,


Appendix A

Cap-domain closure enables diverse substrate recognition by the C2-type haloacid dehalogenase-like sugar phosphatase *Plasmodium falciparum* HAD1
PREFACE

The following work was performed by Jooyoung Park, myself, Audrey R. Odom, and Niraj H. Tolia. J.P. performed protein expression, crystallization, data collection, and structure solution, enzyme assays, and wrote the manuscript. I generated the catalytically active PfHAD1 mutant, performed enzyme assays, and wrote the manuscript. A.R.O. and N.H.T. designed, supervised, and analyzed the studies and wrote the manuscript.


We thank J. Nix and ALS Beamline 4.2.2 (contract DE-AC02-05CH11231) for assistance with X-ray data collection. This work was supported by the Children’s Discovery Institute of Washington University and St. Louis Children’s Hospital (MD-LI-2011-171, to AO and NT), NIH/NIAID R01AI103280 (to AO), a March of Dimes Basil O’Connor Starter Scholar Research Award (to AO) and a Doris Duke Charitable Foundation Clinical Scientist Development award (to AO). J.P. is supported by a Sondra Schlesinger Graduate Student Fellowship and Stephen I. Morse Graduate Student Fellowship from Washington University. A.M.G. is supported by an NIGMS Training grant (5T32GM007067) and a Monsanto Excellence Fund Graduate Fellowship.
A.1 ABSTRACT

Haloacid dehalogenases (HADs) are a large enzyme superfamily of more than 500,000 members with roles in numerous metabolic pathways. *Plasmodium falciparum* HAD1 (PfHAD1) is a sugar phosphatase that regulates the methylenedioxyethyl phosphate (MEP) pathway for isoprenoid synthesis in malaria parasites. However, the structural determinants for diverse substrate recognition by HADs are unknown. Here, crystal structures were determined of PfHAD1 in complex with three sugar phosphates selected from a panel of diverse substrates that it utilizes. Cap-open and cap-closed conformations are observed, with cap closure facilitating substrate binding and ordering. These structural changes define the role of cap movement within the major subcategory of C2 HAD enzymes. The structures of an HAD bound to multiple substrates identifies binding and specificity-determining residues that define the structural basis for substrate recognition and catalysis within the HAD superfamily. While the substrate-binding region of the cap domain is flexible in the open conformations, this region becomes ordered and makes direct interactions with the substrate in the closed conformations. These studies further inform the structural and biochemical basis for catalysis within a large superfamily of HAD enzymes with diverse functions.

A.2 INTRODUCTION

The HAD superfamily, the name of which derives from bacterial haloacid dehalogenases, is one of the largest classes of enzymes and is represented throughout all kingdoms of life [1]. This superfamily is comprised of more than 500,000 members (InterPro IPR023214) that have
evolved to serve diverse biological functions [2]. HADs have been shown to play roles in primary [3,4] and secondary metabolism [5–8], regulation of metabolic pools [9], cell housekeeping [10–13], and nutrient uptake [14]. HADs are promiscuous enzymes, displaying catalytic activity towards a broad range of substrates [15]. The majority of HADs are involved in variations of phosphoryl-transfer reactions [16] and include phosphoesterases, ATPases, phosphonatases and sugar phosphomutases [17,18]. HADs contain four highly conserved sequence motifs, which coordinate a magnesium ion and bind the phosphoryl group of substrate compounds. Together, these HAD-motif residues comprise a self-contained scaffold for catalysis through a phosphoaspartyl intermediate [17,19,20]. Further biochemical and structural studies, combined with biological function studies, are needed to understand how substrate recognition and catalysis are achieved by these promiscuous enzymes in order to carry out their diverse metabolic regulatory functions.

We recently identified and reported the crystal structure of the first HAD protein from *Plasmodium falciparum*, which we called PfHAD1 [21]. PfHAD1 is a sugar phosphatase that regulates substrate availability to the methylerithritol phosphate (MEP) pathway for isoprenoid precursor biosynthesis (Figure A1). Understanding the regulation of the MEP pathway is important, as the pathway is an attractive target for drug development. While isoprenoids are essential for all organisms, humans employ a completely different isoprenoid-biosynthesis pathway: the mevalonate pathway. The MEP pathway has been chemically and genetically validated to be essential in a number of important pathogens, including *P. falciparum* [22–24], *Toxoplasma gondii* [25], *Mycobacterium tuberculosis* [26], and *Escherichia coli* [27]. Furthermore, isoprenoid production is of interest for a wide range of commercially important
natural products such as pharmaceuticals and microbial biofuels. Thus, structural insight into the regulation of the MEP pathway is important for its implications in drug development and for metabolic engineering to enhance isoprenoid production for commercial interests.

PfHAD1 serves as an excellent candidate for structural studies of substrate binding and catalysis within the HAD superfamily because its cellular function has been defined. We found that changes in PfHAD1 confer malaria-parasite resistance to the antimalarial fosmidomycin by sequencing *P. falciparum* parasites lines that are fosmidomycin-resistant [21]. Fosmidomycin is a small-molecule inhibitor of the MEP pathway. Fosmidomycin-resistant parasite strains are highly enriched in genetic changes in the *PfHAD1* locus that result in loss of PfHAD1 activity. These genetic changes correlate with a metabolic effect of increased cellular levels of MEP pathway intermediates. Complementation of a deleterious *PfHAD1* allele restores sensitivity to fosmidomycin. While the metabolic effects and a biological phenotype of PfHAD1 have been described, the substrate specificity and mechanism of catalysis for PfHAD1, or HADs generally, have not been well defined.

The essential structural element of HADs is a Rossmannoid fold characterized by a three-layered α/β sandwich comprised of repeating β-α units [16,19]. The HAD Rossmannoid fold is distinguished from other Rossmann folds by two motifs called the 'squiggle' and ‘flap' elements [16,17]. The squiggle element occurs immediately following the first β-strand of the Rossmannoid fold and assumes a nearly complete single-helical turn. The flap element is a β-hairpin downstream of the squiggle. While the core Rossmannoid folds of HAD enzymes have undergone minimal modifications, the introduction of a cap-insertion module has added a
sophisticated means of substrate recognition and diversification of enzyme function by presenting new surfaces for interaction with substrates [15]. HAD proteins can be classified into three general categories based on the topology and location of their cap insertions: C0, C1 or C2. The C0 element consists of short loops or β-strands that do not fold into its own structural unit, whereas the C1 element folds into an independently folded cap domain that is distinct from the catalytic core domain. The insertion in C0 and C1 HADs occurs in the middle of the β-hairpin of the flap motif. The C2 element also folds into its own distinct domain, but the cap insertion occurs in the linker following the third β-strand of the Rossmannoid fold.

In addition to providing substrate-specificity determinants, the cap module plays a role in solvent access and exclusion from the active site, which is a necessary aspect of catalysis via the two-step phosphoaspartyl-transferase mechanism [19,28]. The C0 HADs, which lack a cap domain, primarily use large macromolecule substrates [28,29] or oligomerization [6,8,30] as a means of substrate binding and solvent exclusion from the catalytic site. The C1 HADs employ extensive movement of the cap domain, which is dependent on winding and unwinding of the squiggle element, to form open and closed conformations [17,31,32]. The role of cap movement in C2 HADs, such as PfHAD1, has not been extensively explored. It has been postulated that the squiggle–flap elements in C2 HADs also exhibit drastic movements to exclude solvent from the substrate-binding site [16,19]. However, to our knowledge evidence for this movement in C2 HADs has not been established.

Detailed structural analysis of the determinants for substrate ambiguity and specificity in HAD enzymes has not been investigated. To investigate the structural basis for substrate recognition
and catalysis in *P. falciparum* PfHAD1, we constructed, purified and characterized a catalytically inactive PfHAD1-D27A mutant. We determined three substrate-bound crystal structures: those of PfHAD1-D27A complexed with mannose 6-phosphate (Man6P), glucose 6-phosphate (Glu6P) or glyceraldehyde 3-phosphate (Gly3P). Our crystallographic and enzymatic activity data provide insights into the structural determinants for substrate recognition and the role of C2 cap-domain movement in substrate binding, ordering and catalysis. We find that a large cap-domain movement results in sealing of the substrate-binding cavity. This movement is important for the enzymatic mechanism, as it positions cap residues in the correct orientation to bind and order substrates prior to catalysis. These results alter the paradigm that substrate specificity is imparted by the squiggle–flap motifs and is instead driven by cap closure. The structures of multiple substrate-bound complexes identified key specificity-determining residues within the cap. These specificity-determining residues are conserved in unique patterns within the HAD superfamily that aid in functional characterization. Understanding the PfHAD1 mechanism informs catalysis of the entire HAD superfamily, which regulates important metabolic pathways.

**A.3 METHODS**

**A.3.1 Reagents**

All reagents were purchased from Sigma–Aldrich unless otherwise indicated.

**A.3.2 Site-directed mutagenesis for PfHAD1 mutants**

Wild-type PfHAD1 (WT-PfHAD1) was cloned as described previously [21]. Site-directed
mutagenesis to create the D27A allele was carried out using standard PCR-based techniques using the following primers: 5′-TGTTCCATC-TAAAGC-CGTAAG-3′ and 5′-CTTTAC-GGCTTTAG-ATGGAACA-3′. Sites for ligation-independent cloning were added using the following primers: 5′-CTCACCACCACCACCACCATTGCACGAAATTGTA-GATAAGAA-3′ and 5′-ATCCTATCTTACTCATTATA-TGTCACAGAATGTCTTCA-3′. The resulting PCR product was inserted into the BG1861 expression vector [33] to produce 6×His-PfHAD1. The construct sequence was verified by Sanger sequencing. Site-directed mutagenesis was performed to generate the V151A, V151L, E152A, L173A, E205A and E205W mutants.

A.3.3 Expression and purification

*E. coli* BL21(DE3) pLysS competent cells (Life Technologies) harboring the WT-PfHAD1 or mutant expression plasmids were grown at 37°C in LB medium containing final concentrations of 0.1 mg ml⁻¹ ampicillin and 0.3 mg/ml chloramphenicol. Once the cells reached an OD₆₀₀ of 0.6, they were induced with 1 mM IPTG for 3–4 h. After this period, the cells were harvested by centrifugation at 4000 rpm for 10 min at 4°C. The cell pellets were suspended in 10 ml buffer A (50 mM Tris pH 8.0, 100 mM NaCl, 10 mM imidazole pH 8.0, 1 mM PMSF, 5 mM β-mercaptoethanol) per liter of LB medium and were stored at −80°C.

The cells were thawed in the presence of 0.25 mg ml⁻¹ lysozyme and disrupted using sonication on ice for 60 s. The cell extract was obtained by centrifugation at 13,000 rpm for 30 min at 4°C and was applied onto Nickel Rapid Run agarose beads (Goldbio) equilibrated with buffer B (50 mM Tris pH 8.0, 150 mM NaCl, 20 mM imidazole pH 8.0, 5 mM β-mercaptoethanol). Buffer B
was used to wash the nickel column three times with five column volumes. After washing, the protein was eluted with five column volumes of buffer C (buffer B with 300 mM imidazole). The PHAD1 was further purified by gel chromatography using a HiLoad 16/600 Superdex 200 pg column (GE Healthcare) equilibrated with buffer D [10 mM Tris pH 8.0, 150 mM NaCl, 5 mM dithiothreitol (DTT)]. The fractions containing PfHAD1 were pooled and concentrated using a 10K molecular-weight cutoff Amicon centrifugal filter (Millipore).

A.3.4 Enzyme assays

Enzyme assays were performed as described previously [21]. Briefly, general phosphatase activity was measured by monitoring the hydrolysis of $p$-nitrophenylphosphate ($p$NPP) to $p$-nitrophenyl. Reactions consisted of 10 mM $p$NPP (New England Biolabs), 50 mM Tris–HCl pH 7.5, 5 mM MgCl$_2$, 0.5 mM MnCl$_2$ and 2 µg enzyme. Phosphate cleavage of phosphorylated sugar substrates was monitored using the EnzChek Phosphatase Assay Kit (Life Technologies). Reactions contained 1 mM substrate and 400 ng enzyme. All substrates were purchased from Sigma–Aldrich. For all assays, reaction rates were calculated using GraphPad Prism software.

A.3.5 Crystallization

For crystallization, PfHAD1-D27A was concentrated to 20 mg/ml and incubated for 30 min on ice with 5 mM substrate for co-crystallization studies. Crystals of apo PfHAD1-D27A or PfHAD1-D27A grown in 5 mM substrate were obtained by vapor diffusion using hanging drops equilibrated at 18°C against 500 µl of a reservoir consisting of 0.1 M HEPES pH 7.5, 20–25% PEG 8000. PfHAD1-D27A crystals were cryoprotected with 0.1 M HEPES pH 7.5, 30% PEG 8000 and either no substrate for apo PfHAD1-D27A crystals or 5 mM substrate for co-crystals
before flash-cooling under liquid nitrogen.

A.3.6 Data collection and processing

X-ray data were collected from a single crystal using a wavelength of 1 Å on beamline 4.2.2 of the Advanced Light Source, Berkeley, California, USA. Data were collected with the CMOS detector and were processed with XDS [34]. All crystals had the same crystal packing and dimensions as the previously reported WT-PfHAD1 crystal [21]. R and R_free flags were imported from the WT-PfHAD1 mtz file using UNIQUEIFY within the CCP4 package [35].

A.3.7 Structure determination and refinement

Structure solution was performed within PHENIX [36] by refinement with the previously solved WT-PfHAD1 crystal structure (PDB entry 4qjb). Subsequent iterated manual building/rebuilding and refinement of models were performed using Coot [37] and PHENIX [36], respectively. The MolProbity structure-validation server [38] was used to monitor the refinement of the models. All final refined models have favorable crystallographic refinement statistics, as provided in Table A1. A stereo image of a representative region of the electron-density map is shown for each structure in Supplementary Figure A1. Figures were generated and rendered in PyMOL (v.0.99rc6; Schrödinger).

A.3.8 Accession codes for NCBI BLAST search

The PfHAD1 protein sequence was used as the query sequence to search for all nonredundant protein sequences in the NCBI database using the BLASTP 2.2.31+ (protein–protein BLAST) algorithm. Plasmodium (Taxid 5820) was excluded from the search. Sequence similarity was

### A.4 RESULTS

#### A.4.1 Generation of inactive PfHAD1-D27A for substrate-binding studies

Based on the sequence and structural analysis of WT-PfHAD1 [21], Asp27 was predicted to perform a nucleophilic attack on the phosphate group of PfHAD1 substrates. The PfHAD1-D27A mutant protein was generated and tested for catalytic activity towards a non-specific phosphatase substrate, *p*-nitrophenylphosphate (*p*NPP), and a representative panel of sugar phosphates, Man6P, Glu6P and Gly3P. These sugar phosphates represent six-carbon (6C; Man6P and Glu6P) and three-carbon (3C; Gly3P) sugar phosphates with different stereochemical properties that have biological relevance. Gly3P is one of the starting substrates for the MEP pathway, and Glu6P and Man6P are upstream precursors of Gly3P (Figure A1). PfHAD1-D27A lacks phosphatase activity towards all compounds tested (Supplementary Figure A2) and was employed for co-crystallization studies to obtain substrate-bound structures.
A.4.2 Overall structures of PfHAD1-D27A bound to substrates

Crystal structures of PfHAD1-D27A in complex with Man6P, Glu6P or Gly3P were determined. All structures are of high resolution (1.8–2.0 Å) sufficient to define small-molecule binding, and a complete summary of crystallographic refinement statistics is given in Table A1. The PfHAD1-D27A mutant has the same overall structure as WT-PfHAD1 (Figure A2A), with the active site found at the interface between the core and cap domains. In each structure, two monomers were present in the asymmetric unit. Overlay of the two monomers by superposition of the core domains shows that PfHAD1 adopts a ‘closed' and an ‘open' conformation (Figure A2B).

Domain-movement analysis using DynDom [39,40] reveals an 18° rotation between the two conformations about the two hinge loops (residues 104–105 and residues 210–214) connecting the core and cap domains.

For each of the substrate-bound structures, clear electron density corresponding to the substrate could be observed in the substrate-binding site of the closed conformation. The validity of the models was confirmed by inspecting the Fo − Fc density prior to the addition of the substrate models (Figure A3A) and inspecting the 2Fo − Fc electron-density maps after refinement with the substrate (Figure A3B). In the open conformation, electron density corresponding to the phosphate group could be visualized, but the electron density for the orientation of the sugar moiety was unclear (Figure A3C). Attempts to model in the substrate molecule were unsuccessful. Therefore, only the phosphate group was modeled in the open conformation (Figure A3D).
A.4.3 PfHAD1 substrate-binding site

The active-site interactions in the co-crystal structures were examined more closely in order to investigate the structural basis for substrate binding and catalysis. HADs contain four highly conserved sequence motifs. These residues in PfHAD1 are as follows: motif I (Asp27 and Asp29), motif II (Thr61), motif III (Lys215) and motif IV (Asp238 and Asp242) (Figure A4A and A4B). In the initial reaction, the Asp27 nucleophile (which is absent in our PfHAD1-D27A mutant) is positioned to attack the phosphoryl group (Figure A4A and A4B). Asp29, which is positioned two residues away from Asp27, serves as a general acid/base residue to protonate the leaving sugar group. In the second reaction, Asp29 deprotonates a water molecule so that it can perform nucleophilic attack on the phosphoaspartyl intermediate, thus releasing free phosphate and restoring the enzyme to its native state (Figure A1).

In addition to the conserved molecular interactions between the catalytic residues of the core domain and the phosphoryl group, residues in the cap domain make molecular interactions with the sugar moiety of the substrates in the closed conformation (Figure 4C, 4D, 4E). In this conformation, the substrate-binding site is closed off and free exchange of water molecules with the bulk solvent is prevented, thus favoring the aspartate-based nucleophilic reaction. The open conformation, on the other hand, exposes the active site to bulk solvent, allowing substrate release. Owing to the shift of the cap domain away from the core domain in this open conformation, no molecular interactions are made between the sugar moiety of the substrates and the substrate-binding residues of the cap domain.

A comparison of the substrate-binding site in each of the three crystal structures reveals that
Man6P, Glu6P and Gly3P all make contacts with cap residues Glu152, Thr201, Phe202 and Tyr205 (Figure A4C, A4D, A4E). In addition, the larger 6C sugar phosphates (Man6P and Glu6P) make further contacts with cap residues Val151, Leu173, Glu207 and Arg63 from the catalytic domain (Figure A4C, A4D, A4E). Owing to their different stereochemistry, slightly different interactions are made by Man6P and Glu6P. These contacts are consistent with the increased enzyme activities observed toward Man6P and Glu6P over Gly3P (Supplementary Figure A2). Thus, distinct interactions for 6C and 3C sugar phosphates form the basis for substrate binding and specificity.

### A.4.4 The cap domain contains a key substrate-recognition element

Although movement in the squiggle and flap elements has previously been postulated to be important for substrate binding and solvent exclusion in C2 HADs [16], no movement in these elements was observed between the open and closed conformations of PfHAD1. In our structures, the flap element does not form a complete β-hairpin, but rather a simple β-turn. The $B$-factor values for the two monomers were examined in order to identify whether these regions demonstrated high flexibility and disorder in our structures. We observed average $B$-factor values for the squiggle and flap elements in all three structures, indicating that these elements are ordered in our structures (Figure A5A and Supplementary Figure A3). However, we observed abnormally high $B$-factor values in an α-helical region of the cap domain that is important for substrate binding. Glu152 of this region has the highest $B$-factor value of $\sim$110 Å (Figure A4A and Supplementary Figure A3A). In contrast, this region becomes ordered in the closed conformation, where the cap residues make direct interactions with the sugar moieties (Figure A5B and Supplementary Figure A3B). Glu152 makes a conserved hydrogen-bond interaction
with an active-site water molecule in all of the substrate-bound structures (Figure A4C, A4D, and A4E). This water molecule hydrogen-bonds to sugar moieties of the substrates and helps to orient them in the active site. To test the relevance of Glu152 in catalysis, we generated a PfHAD1-E152A variant protein. Loss of Glu152 significantly lowers the enzyme activities equally for all substrates (Figure A5C), demonstrating its important role in substrate recognition.

**A.4.5 Substrate specificity informs the HAD superfamily**

Val151, Glu152, Leu173, Thr201, Phe202, Tyr205 and Glu207 are cap-domain residues that bind substrates and enable specificity for PfHAD1. Having identified the structural determinants for substrate binding and specificity, we asked whether these determinants define the large superfamily of HAD enzymes. We examined the conservation of these substrate-binding residues in an unbiased sampling of the ten closest PfHAD1 homologs identified solely by sequence similarity.

Sequence alignment of these PfHAD1 homologs demonstrates that they are highly conserved in their catalytic core domains, but that their sequences diverge in their cap domains (Figure A6A and Supplementary Figure A4A). This is expected as our structural analysis indicates that the cap domains impart substrate specificity. Within these sequences, Glu207 is invariant and Leu173 is almost invariant, with substitutions to Met, Ile or Val in certain HADs (Figure A6B). These two residues therefore play a conserved role in substrate binding across HADs, but a minor role in specificity. Strikingly, analysis of the remaining residues suggests that HAD enzymes can be subdivided into two groups. In the first subdivision (Figure A6B, blue boxes) the hydroxyl group of Thr201 and the aromatic side chain of Tyr205 are well conserved, while Val151, Glu152 and
Phe202 are variable. The defining features of the second subdivision (Figure A6B, red boxes) include an invariant Tyr at residue 151, Ala/Pro at residue 201, Val/Ile at residue 202 and Met/Asn at residue 205. Therefore, while the overall sequence conservation in the cap domains is lower than the catalytic core domains, patterns of sequence conservation and divergence are observed in cap residues that are important for substrate recognition. These patterns are only evident through the structural definition of substrate recognition presented in this study.

The analysis above used an unbiased sampling of proteins similar to PfHAD1. As PfHAD1 is a regulator of the MEP pathway in *P. falciparum* [21], we examined PfHAD1 homologs in other important pathogens and model organisms that also employ the MEP pathway: *E. coli, M. tuberculosis, A. thaliana* and *C. reinhardtii* (Supplementary Figure A4B). All of these species employ the MEP pathway, as determined by the presence of DXR, the first committed enzyme of the pathway. Again, similar patterns of conservation and variance emerge within these selected PfHAD1 homologs (Figure A6C). Val151 and Leu173 both provide hydrophobic contacts to the substrates and are important for sealing the substrate-binding cavity. However, Leu173 is strongly conserved while Val151 is variant (Figure A5B and A5C and above). We mutated each of these residues to alanine to compare the effects of nonconserved and conserved residues on catalysis (Figure A6D). Mutation of the conserved Leu173 resulted in an enzyme that was essentially inactive to all three substrates. In contrast, mutation of the nonconserved Val151 retained ∼75% of the wild-type activity for all three substrates. These results are consistent with the plasticity at Val151 and other variant residues to allow altered substrate specificity of HAD proteins.
None of the homologs from the analyses above have known substrate preferences except for *E. coli* YidA. The patterns of conserved and substituted binding residues identified by this structural work provide a framework for the future determination of substrate preferences for related HAD enzymes. To evaluate the validity of this framework, we examined the C2-type *E. coli* HAD enzymes YbiV, YidA, YbhA, YbjI, YigL, OtsB, Cof and NagD, as the substrate preferences of each of these enzyme is well defined [12]; C0 and C1 HAD enzymes were omitted from this analysis as they have different domain architectures. YbiV and YidA both catalyze dephosphorylation of 3–6-carbon sugar phosphates similar to PfHAD1, with YbiV having greatest activity against fructose 1-phosphate and YidA preferring erythrose 4-phosphate. As predicted, we find that YbiV and YidA have similar patterns of conservation/divergence as PfHAD1: variable sequences at positions 151/152, Leu/Met at position 173, Tyr/Ser at position 201, Tyr/Phe at position 205 and Asp/Glu at position 207 (Supplementary Figure A4C). In contrast, YbhA, YbjI, YigL, OtsB, Cof and NagD utilize structurally different phosphorylated metabolites. YbhA, YigL and Cof have greatest activity against pyridoxal 5-phosphate, YbjI is most active against flavin mononucleotides, OtsB has a preference for trehalose 6-phosphate and NagD is most active against adenosine triphosphate. These *E. coli* homologs diverge dramatically from PfHAD1 at the substrate-binding residues (Supplementary Figure A4C), further validating the correlation between structure and substrate preference.

### A.5 DISCUSSION

In this study, we have solved co-crystal structures of the sugar phosphatase PfHAD1 bound to diverse sugar-phosphate substrates in order to investigate the structural determinants for
substrate recognition and specificity. By investigating molecular interactions within PfHAD1 bound to multiple different substrates, our study is the first structural study to examine how the HAD enzyme family is able to remain promiscuous with substrate utilization: a common feature of the HAD superfamily members, which is otherwise unusual among enzymes that recognize metabolites. Our results demonstrate that PfHAD1 utilizes conserved residues in the catalytic domain to bind the phosphoryl group of sugar-phosphate substrates. Substrate specificity is determined by residues in the cap domain that make interactions with the sugar moiety, thereby defining which substrates PfHAD1 can utilize.

Prior to this study, the role of cap movement in C2 HADs was unclear. It was speculated that C2 HADs undergo structural changes in the squiggle and flap elements rather than the cap domain [16,17] to allow catalysis to occur. However, no direct biochemical or structural evidence for this squiggle–flap movement has been reported. The only other structural clue was reported for a cyanobacterial sucrose-phosphatase, for which an open and closed conformation of the cap domain in two different crystal forms was observed [41].

Here, we have observed open and closed conformations within the same crystal structures, and we find that these changes coordinate substrate binding and ordering. The open and closed conformations are likely to represent physiologically relevant states of PfHAD1 that are essential for substrate binding and catalysis (Figure A7). The open conformation of the enzyme allows a substrate molecule to bind the catalytic residues of the core domain with its phosphoryl group. In this open state, we see clear electron density for the phosphoryl group but not the sugar moiety that is facing bulk solvent. Upon inspection of the $B$-factor values in these structures, we observe
that a substrate-binding element in the cap domain is highly flexible and disordered. We have also captured the subsequent closed conformation of PfHAD1 in our structures. In this closed state, a substrate-binding cavity is formed at the interface between the core and cap domains that prevents the free exchange of water molecules with bulk solvent. The cap domain orders the substrate through polar and nonpolar interactions with the sugar moiety of the substrate, and low B factors are observed for the entire protein. This closed state provides an environment that favors an aspartate-based nucleophilic attack to occur. Subsequent opening of the cap would facilitate product release and allow the enzyme to be restored to its native state.

We identified distinct residues that mediate substrate binding and specificity for PfHAD1. Analysis of these residues in HAD homologs identified patterns of conservation and variance that subdivide the superfamily and may predict their substrate specificities. This analysis suggests that the sub-divisions of HAD homologs may represent similar substrate-utilization profiles and/or biological functions \textit{in vivo}. While mutation of the conserved cap residues are detrimental to enzyme function in these HADs, the variable cap residues are more suitable for changes to alter substrate specificity or for metabolic engineering. Understanding which residues to manipulate enables rational protein-engineering approaches to alter the substrate-utilization profiles of these enzymes. Future biochemical and structural studies of these PfHAD1 homologs will further our understanding of the substrate specificity and biological functions of these additional HAD enzymes.

HAD superfamily members are widespread in biology, but few subfamilies have experimentally defined biological roles. We previously found that PfHAD1 regulates substrate availability to the
MEP pathway for isoprenoid precursor biosynthesis and that loss of PfHAD1 function has dramatic effects on the levels of MEP pathway intermediates [21]. PfHAD1 homologs in other organisms are restricted to phyla that also employ the MEP pathway, suggesting that PfHAD1-like HADs are also regulators of the MEP pathway in these other organisms. Thus, our studies are an important first step in the development of chemical tools to disrupt or alter HAD enzyme functions. These tools will allow us to probe the biological and metabolic functions of HAD enzymes in living cells. As the MEP pathway is essential in parasites and pathogenic bacteria, defining the regulation of this pathway by HAD enzymes may lead to novel avenues for therapeutic intervention.
A.6 FIGURES

Figure A1. Model of PfHAD1 regulation.

PfHAD1 regulates substrate availability to the MEP pathway and is capable of catalyzing the phosphorolysis of small, 3–6-carbon sugar monophosphates upstream of the MEP pathway, including mannose 6-phosphate (Man6P), glucose 6-phosphate (Glu6P) and glyceraldehyde 3-phosphate (Gly3P) [21].
Figure A2. PfHAD1 binds and orders substrates in a closed conformation.

(a) The overall PfHAD1-D27A + Man6P structure is shown here, with Man6P bound in the active site found at the interface between the core and cap domains. (b) Superimposition of the core domains of chain A and chain B reveals a shift in the cap domain relative to the core domain (gray), resulting in a closed (green) and an open (orange) conformation.
Figure A3. Substrates bind to the active site of PfHAD1-D27A.

(a) The $F_o - F_c$ map (contoured at 2.5$\sigma$) of the substrate-binding site of the closed conformations prior to modeling of the substrate (from left to right: Man6P, Glu6P, Gly3P). Clear electron density can be observed for the entire substrate. (b) The $2F_o - F_c$ map (contoured at 1.0$\sigma$) of the closed substrate-binding site after refinement with the substrate model. (c) The $F_o - F_c$ map (contoured at 2.5$\sigma$) of the open substrate-binding site prior to modeling of the substrate. Electron density for the entire substrate is ambiguous. (d) The $2F_o - F_c$ map (contoured at 1.0$\sigma$) of the open substrate-binding site after refinement with a phosphate-group model.
Figure A4. Distinct interactions for six-carbon and three-carbon sugar phosphates form the basis for substrate binding and specificity.

(a) The conserved HAD motifs I–IV make interactions with a magnesium ion, a conserved water molecule and the phosphoryl group. (b) An overlay of active-site residues from the WT-PfHAD1 structure (colored light orange; PDB entry 4qjb) shows that the substrate-binding mode is conserved in the PfHAD1-D27A mutant. (c–e) Detailed van der Waals (colored green) and hydrogen-bond (dashed lines) interactions are shown between cap residues and the sugar moiety.
of each substrate: (c) Man6P, (d) Glu6P, (e) Gly3P. For simplicity, hydrogen-bond interactions between the conserved HAD motifs I–IV and the substrate are not shown here.
Figure A5. The cap domain contains a flexible substrate-recognition element that is ordered upon substrate binding and cap closure.

(a) In the open conformation for Man6P, the region of the cap domain that is important for substrate recognition has high $B$-factor values, indicating disorder and flexibility in this region. (b) In the closed conformation for Man6P, this region becomes ordered, as reflected by the low $B$-factor values. (c) Enzyme activities for WT-PfHAD1 and PfHAD1-E152A are shown, normalized as a percentage of the WT-PfHAD1 activity for each substrate. Displayed are the means ± standard error of the mean of enzyme activity from at least three independent experiments.
Figure A6. Sequence alignment of substrate-binding cap-domain residues in PfHAD1 and PfHAD1 homologs from other organisms.

(a) PfHAD1 homologs in other organisms are highly conserved in their catalytic core domains (blue) but diverge in their cap-domain sequences (green). The black lines denote the five segments comprising the seven residues of the cap domain in PfHAD1 that are important for substrate binding. (b) The substrate-binding cap residues in the ten closest PfHAD1 homologs are shown. The blue and red boxes denote two subdivisions of PfHAD1 homologs. (c) The substrate-binding cap residues in PfHAD1 homologs from model organisms are shown. (d) Enzyme activities for WT-PfHAD1, PfHAD1-V151A and PfHAD1-L173A are shown, normalized as a percentage of the WT-PfHAD1 activity for each substrate. Displayed are the means ± standard error of the mean of enzyme activity from at least three independent experiments.
Figure A7. Model for catalysis in PfHAD1.

The open conformation of PfHAD1 allows the substrate to access the active site and bind the conserved catalytic residues with the phosphoryl group. Subsequent cap closure prevents the free exchange of water molecules with bulk solvent and allows ordering of the substrate for catalysis. Opening of the cap allows solvent access to restore PfHAD1 to its native state.
Supplementary Figure A1. Stereo image of the electron density maps for representative regions of the substrate-bound PfHAD1-D27A structures.

The 2fo-fc electron density map contoured at 1.0 σ is colored blue for (a) man6p, (b) glu6p, and (c) gly3p.
Supplementary Figure A2. PfHAD1-D27A is inactive.

Mutation of the Asp-27 nucleophile to an alanine residue renders recombinant PfHAD1 enzyme inactive against all substrates tested: mannose-6-phosphate (man6P), glucose-6-phosphate (glu6P), glyceraldehyde 3-phosphate (gly3P), and para-nitrophenylphosphate (pNPP).
Supplementary Figure A3. The cap domain contains a flexible substrate recognition element that is ordered upon substrate binding and cap closure.

(a) glu6p open conformation. (b) glu6p closed conformation. (c) gly3p open conformation. (d) gly3p closed conformation. In the open conformations (a,c), the region of the cap domain that is important for substrate recognition has high B-factor values, indicating disorder and flexibility in this region. In the closed conformations (b,d), this region becomes ordered, as reflected by the low B-factor values.
Supplementary Figure A4. Sequence alignment of PfHAD1 homologs from other organisms.

Sequence alignments are shown for (a) an unbiased sampling of the ten closest PfHAD1 homologs, (b) PfHAD1 homolog sequences from selected model organisms, (c) sequences of PfHAD1 homologs from E. coli with known substrate preferences. The cap domain is boxed in green, and substrate-binding core and cap residues in PfHAD1 are annotated with blue and green dots, respectively. The blue and red boxes denote two subdivisions of PfHAD1 homologs.
### Table A1. Data collection and refinement statistics.

<table>
<thead>
<tr>
<th></th>
<th>PfHAD1-D27A + man6P</th>
<th>PfHAD1-D27A + glu6P</th>
<th>PfHAD1-D27A + gly3P</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P2₁</td>
<td>P2₁</td>
<td>P2₁</td>
</tr>
<tr>
<td>a, b, c, (Å)</td>
<td>77.60, 44.50, 84.50</td>
<td>77.70, 44.50, 84.80</td>
<td>77.60, 44.60, 84.10</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>90.00, 101.30, 90.00</td>
<td>90.00, 101.40, 90.00</td>
<td>90.00, 101.30, 90.00</td>
</tr>
<tr>
<td>Resolution (Å)*</td>
<td>20-1.80 (1.90-1.80)*</td>
<td>20-1.90 (2.00-1.90)*</td>
<td>20-2.0 (2.10-2.00)*</td>
</tr>
<tr>
<td>R&lt;sub&gt;meas&lt;/sub&gt; (%)</td>
<td>6.5% (65.2%)</td>
<td>8.4% (66.5%)</td>
<td>11.1% (74.0%)</td>
</tr>
<tr>
<td>I/σI</td>
<td>17.04 (2.45)</td>
<td>16.10 (2.42)</td>
<td>13.78 (2.20)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.3% (99.6%)</td>
<td>98.9% (98.3%)</td>
<td>98.4% (97.7%)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>3.68 (3.71)</td>
<td>3.77 (3.80)</td>
<td>3.79 (3.82)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>20-1.80</td>
<td>20-1.90</td>
<td>20-2.00</td>
</tr>
<tr>
<td>No. of reflections</td>
<td>52,536</td>
<td>44,805</td>
<td>38,024</td>
</tr>
<tr>
<td>R&lt;sub&gt;work&lt;/sub&gt;/&lt;R&lt;sub&gt;free&lt;/sub&gt;</td>
<td>18.83/21.91</td>
<td>17.89/21.95</td>
<td>17.32/22.10</td>
</tr>
<tr>
<td>No. of atoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>4,527</td>
<td>4,512</td>
<td>4,504</td>
</tr>
<tr>
<td>Mg</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Ligand</td>
<td>21</td>
<td>21</td>
<td>15</td>
</tr>
<tr>
<td>Water</td>
<td>364</td>
<td>288</td>
<td>296</td>
</tr>
<tr>
<td>B-factors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>29.90</td>
<td>29.80</td>
<td>34.30</td>
</tr>
<tr>
<td>Ligands</td>
<td>59.80</td>
<td>42.40</td>
<td>52.10</td>
</tr>
<tr>
<td>Water</td>
<td>30.20</td>
<td>28.80</td>
<td>32.50</td>
</tr>
<tr>
<td>Ramachandran</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favored (%)</td>
<td>97.72</td>
<td>97.70</td>
<td>97.16</td>
</tr>
<tr>
<td>Allowed (%)</td>
<td>2.28</td>
<td>2.30</td>
<td>2.84</td>
</tr>
<tr>
<td>Outliers (%)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>R.m.s. deviations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.008</td>
<td>0.008</td>
<td>0.007</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.198</td>
<td>1.187</td>
<td>0.969</td>
</tr>
<tr>
<td>PDB ID</td>
<td>4ZEV</td>
<td>4ZEW</td>
<td>4ZEX</td>
</tr>
</tbody>
</table>

* Highest resolution shell is shown in parenthesis.
A.8 REFERENCES


13. Titz B, Häuser R, Engelbrecher A, et al. The *Escherichia coli* protein YjjG is a house-

    phosphatase (AphA) of *Escherichia coli* MG1655. Biochim. Biophys. Acta - Proteins
    Proteomics 2006; 1764:13–19.

15. Pandya C, Farelli JD, Dunaway-Mariano D, Allen KN. Enzyme promiscuity: engine of

16. Burroughs AM, Allen KN, Dunaway-Mariano D, Aravind L. Evolutionary genomics of
    the HAD superfamily: understanding the structural adaptations and catalytic diversity in a

17. Allen KN, Dunaway-Mariano D. Markers of fitness in a successful enzyme superfamily.

18. Allen KN, Dunaway-Mariano D. Phosphoryl group transfer: evolution of a catalytic

19. Seifried A, Schultz J, Gohla A. Human HAD phosphatases: structure, mechanism, and

20. Collet J-F, Stroobant V, Van Schaftingen E. Evidence for phosphotransferases
    phosphorylated on aspartate residue in N-terminal DXDX(T/V) motif. Methods Enzymol.

    5:4467.

22. Cassera MB, Gozzo FC, D’Alexandri FL, et al. The methylerythritol phosphate pathway is
    functionally active in all intraerythrocytic stages of *Plasmodium falciparum*. J. Biol.

23. Odom AR, Van Voorhis WC. Functional genetic analysis of the *Plasmodium falciparum*
    170:108–11.

    agent fosmidomycin revealed by cellular metabolic profiling. Biochemistry 2011;


Appendix B

Additional insights on the cellular function of PfHAD1
PREFACE

We are thankful to Wandy Beatty (Molecular Microbiology Imaging Facility at Washington University School of Medicine) for performing electron microscopy. We thank the Malaria Research Reference and Reagent Resource (MR4) for providing us with a reagent contributed by JH Adams (MRA-1). We are grateful to the Sibley lab for the use of their ultracentrifuge.
B.1 METHODS

B.1.1 Growth assays

Asynchronous parasites were cultured starting at 0.5-1% parasitemia. Samples were taken at day 0 and every 2 days for 12 days. Samples were fixed (4% paraformaldehyde, 0.05% glutaraldehyde in PBS) and stored at 4 °C. Culture media was replaced daily and cultures were subcultured into fresh RBCs throughout the assay to ensure growth was not stunted due to high parasitemia. Fixed cells were washed with PBS, followed by staining with 0.01 mg/ml acridine orange (Invitrogen). Infected cells were counted using a BD LSRII flow cytometer. Data reflects corrections due to subculturing.

B.1.2 PfHAD1 expression

*P. falciparum* cultures (strain 3D7) were synchronized by treatment with 5% sorbitol. Parasites were allowed to complete one lifecycle (~48 hours) before beginning the experiment. Samples were taken at various timepoints for microscopy and immunoblot. Parasite morphology was evaluated using Giemsa-stained blood smears. PfHAD1 expression was evaluated by immunoblotting using anti-PfHAD1 antisera (1:20,000) [1].

B.1.3 Electron microscopy

*P. falciparum*-infected erythrocytes were fixed in 4% paraformaldehyde/0.05% glutaraldehyde (Polysciences Inc., Warrington, PA) in 100 mM PIPES/0.5 mM MgCl$_2$, pH 7.2 for 1 hour at 4°C. Samples were then embedded in 10% gelatin and infiltrated overnight with 2.3 M sucrose/20%
polyvinyl pyrrolidone in PIPES/MgCl\(_2\) at 4°C. Samples were trimmed, frozen in liquid nitrogen, and sectioned with a Leica Ultracut UCT cryo-ultramicrotome (Leica Microsystems Inc., Bannockburn, IL). Fifty nm sections were blocked with 5% fetal bovine serum/5% normal goat serum for 30 minutes and subsequently incubated with rabbit anti-PfHAD1 or rabbit anti-GFP (ab6556) (Abcam, Cambridge, MA), goat anti-rabbit IgG (H+L) conjugated to 12 or 18 nm colloidal gold (Jackson ImmunoResearch Laboratories, Inc., West Grove, PA). Sections were washed in PIPES buffer followed by a water rinse, and stained with 0.3% uranyl acetate/2% methylcellulose. Samples were viewed with a JEOL 1200EX transmission electron microscope (JEOL USA Inc., Peabody, MA) equipped with an AMT 8 megapixel digital camera (Advanced Microscopy Techniques, Woburn, MA). All labeling experiments were conducted in parallel with controls omitting the primary antibody. These controls were consistently negative at the concentration of colloidal gold-conjugated secondary antibodies used in these studies.

**B.1.4 PIP-binding assays**

All steps were performed at room temperature. PIP strips from Echelon Biosciences (#P-6001) were blocked for 1 hr in blocking buffer (3% bovine serum albumin in PBS-T). Strips were incubated with 1-5 µg/mL recombinant 6xHis-PfHAD1 in blocking buffer for 1 hr. Strips were washed with PBS-T and incubated with 1:20,000 rabbit anti-PfHAD1 antisera in blocking buffer. Strips were washed and binding was detected using the colorimetric K-TMBP reagent (Echelon #K-TMBP).

**B.1.5 Enzyme assays**

PfHAD1 enzyme assays using inositol phosphates and phosphatidylinositols (diC8 lipids) were
performed as described for other PfHAD1 substrates [1] in Chapter 2, section 2.3.13, using 200 ng recombinant enzyme and 1 mM substrate. PIPs and inositol phosphates were purchased from Echelon Biosciences and Sigma-Aldrich.

**B.1.6 Cellular fractionation**

Approximately 1 x 10⁹ parasites were isolated by saponin lysis. Parasites were lysed by resuspension in hypotonic buffer (20 mM HEPES, 2 mM EGTA, Roche mini EDTA-free protease inhibitor tablets) and incubated on ice for 1-2 hours. All centrifugation steps were performed at 4 °C. Unlysed cells were removed by centrifugation at 5000 x g for 5 minutes. Lysate was ultracentrifuged at 100,000 x g for 1 hr to separate soluble and insoluble (membrane-bound) fractions. Subsequent centrifugations were performed after washing and incubating with high salt (hypotonic buffer + 0.5 M sodium chloride) and carbonate (0.1 M sodium carbonate, pH 11) to release peripheral membrane proteins. Fractions were assayed by immunoblot using rabbit polyclonal anti-PfHAD1 antisera (1:20,000) [1], anti-aldolase (1:5,000, Abcam ab38905), and anti-PfERD2 (1:5,000, MRA-1 from MR4 repository) [2]. When necessary, HRP-conjugated goat anti-rabbit antibody was used as a secondary (1:20,000, ThermoFisher Scientific 65-6120).
B.2 PFHAD1 MUTANTS DO NOT HAVE A GROWTH DEFECT

Figure B1. FSM⁰ P. falciparum with PfHAD1 mutations do not have a growth defect.

Shown is growth over time for the parental strain (black), two FSM⁰ strains (AM1, deep red and D6, deep blue). FSM⁰ strains rescued with functional PfHAD1 (Hsp110:PfHAD1-GFP) are shown in light red and blue. Strains D6, AM1, and AM1 Hsp110:PfHAD1-GFP are described in [1] and Chapter 2. Strain D6 Hsp110:PfHAD1-GFP was generated in the same way as AM1 Hsp110:PfHAD1. Error bars represent S.E.M. from at least three independent experiments.
B.3 PFHAD1 LOCALIZATION VIA ELECTRON MICROSCOPY

Figure B2. Localization of PfHAD1 by immunolabeling and electron microscopy.

All images are of ACP\textsubscript{L}-GFP (apicoplast-localized GFP) parasites [3] stained with anti-PfHAD1 polyclonal antiserum and goat anti-rabbit conjugated with 18 nm colloidal gold. Food vacuoles (FV) and red blood cell cytoplasm (RBC) are indicated. (A and B) PfHAD1 localization is
generally cytosolic, with some localization to membrane structures (endoplasmic reticulum) in the cytoplasm. (C) A small fraction of PfHAD1 localizes to apicoplast lumens or membranes. ACP$_L$-GFP is stained with anti-GFP and goat anti-rabbit conjugated with 12 nm colloidal gold.
B.4 PFHAD1 DOES NOT BEHAVE LIKE A MEMBRANE PROTEIN

Figure B3. PfHAD1 does not behave like a membrane-associated protein.

Ultracentrifugation was used to fractionate *P. falciparum* cell lysate. Fractions were immunoblotted to probe for aldolase (PfALD), PfERD2, and PfHAD1. PfALD is a control for both cytoplasmic and peripheral membrane localization [5]. As expected, a fraction of PfALD is insoluble (membrane-localized), but this peripheral-membrane localization is sensitive to high salt concentrations. PfERD2 is an integral membrane protein in the endoplasmic reticulum [2]. PfERD2 is entirely insoluble and is resistant to both high salt and carbonate washes. PfHAD1 behaves like a soluble protein (not membrane-localized).
## B.5 PfHAD1 Can Bind to Membrane Lipids

### Abbreviations

- LPA, Lysophosphatidic acid; LC, Lysophosphocholine; PI, Phosphatidylinositol; PI(3)P,
  Phosphatidylinositol 3-phosphate; PI(4)P, Phosphatidylinositol 4-phosphate; PI(5)P,
  Phosphatidylinositol 5-phosphate; PE, Phosphatidylethanolamine; PC, Phosphatidylcholine; S1P,
  Sphingosine 1-phosphate; PI(3,4)P₂, Phosphatidylinositol 3,4-bisphosphate; PI(3,5)P₂,
  Phosphatidylinositol 3,5-bisphosphate; PI(4,5)P₂, Phosphatidylinositol 4,5-bisphosphate; PI(3,4,5)P₃,
  Phosphatidylinositol 3,4,5-trisphosphate; PA, Phosphatidic acid; PS,
  Phosphatidylserine. Membranes contain 100 pmol of lipid per spot. PIPs used are diC₁₆ lipids.

### Figure B4. PfHAD1 Can Bind Membrane Lipids

<table>
<thead>
<tr>
<th></th>
<th>WT</th>
<th>G30E</th>
<th>D27A</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPA</td>
<td>SIP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LPC</td>
<td>PI(3,4)P₂</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PI</td>
<td>PI(3,5)P₂</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PI(3)P</td>
<td>PI(4,5)P₂</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PI(4)P</td>
<td>PI(3,4,5)P₃</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PI(5)P</td>
<td>PA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PE</td>
<td>PS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC</td>
<td>Blank</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(A) Lipids spotted on phosphatidyl inositol phosphate (PIP) strip membrane.
(B-D). PfHAD1 is capable of binding a variety of PIPs. The G30E PfHAD1 variant found in FSM\(^R\) strain D6 is also capable of binding PIPs, while the D27A catalytic mutant (described in [4] and Appendix A) is unable to bind. Data are representative of at least three independent experiments.
B.6 PFHAD1 IS UNABLE TO UTILIZE INOSITOL PHOSPHATES OR PIPS AS SUBSTRATES

![Figure B5. PfHAD1 activity against various phosphatidylinositols (diC8 lipids) and soluble inositol phosphates.](image)

Data shown are means and S.E.M.s of at least two independent experiments. Abbreviations used:
B.7 REFERENCES


3. Waller RF, Reed MB, Cowman AF, McFadden GI. Protein trafficking to the plastid of Plasmodium falciparum is via the secretory pathway. EMBO J. 2000; 19:1794–802.
