# Structural and functional studies of heme binding proteins toward the understanding of malaria

by
Allena Mistral Goren

B.S. Chemistry
University of Arizona, 2011

SUBMITTED TO THE DEPARTMENT OF CHEMISTRY IN PARTIAL FULFULLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN CHEMISTRY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2019

© 2019 Massachusetts Institute of Technology. All rights reserved.

Signature of Author: Signature redacted
Department of Chemistry
August 1st, 2019

Certified by: Signature redacted
Catherine L. Drennan
Professor of Biology and Chemistry
Howard Hughes Medical Institute Professor and Investigator
Thesis Supervisor

Accepted by: Signature redacted
Robert W. Field
Haslam and Dewey Professor of Chemistry
Chair, Departmental Committee on Graduate Students

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
OCT 03 2019
LIBRARIES
ARCHIVES

This doctoral thesis has been examined by a Committee of the Department of Chemistry as follows:

Signature redacted

Professor Elizabeth M. Nolan _____

Committee Chairwoman
Professor of Chemistry

Signature redacted

Professor Catherine L. Drennan _____

Research Supervisor
Professor of Biology and Chemistry
Howard Hughes Medical Institute Professor and Investigator

Signature redacted

Professor Jacquin C. Niles _____

Committee Member
Associate Professor of Biological Engineering
Director, Center for Environmental Health Sciences

# Structural and functional studies of heme binding proteins toward the understanding of malaria

by

Allena Mistral Goren

Submitted to the Department of Chemistry
on August 1st, 2019 in Partial Fulfillment of the
Requirements for Degree of Doctor of Philosophy
in Chemistry

## ABSTRACT

Malaria, a potentially lethal parasitic infection, causes approximately half a million deaths each year. Currently, nearly half of the world's population is at risk of being infected with *Plasmodium*, the parasite causing malaria. In this thesis, we use biochemical and biophysical techniques to describe the iron binding properties of two proteins relevant to the biological study of *Plasmodium falciparum*. The first is a previously uncharacterized protein, which we have named MFP, or malarial ferrous protein, that is found in the parasitophorous vacuole of *Plasmodium*. The second is mRuby2, a fluorescent probe often utilized in biological work to track the localization of proteins. Herein, we report on the novel heme-binding properties of both proteins.

We have identified a novel function of mRuby2, a commonly used fluorescent probe. Upon incubation with heme, the fluorescence of mRuby2 decreases, providing a potential use for the protein as a heme probe as well as a limitation of its utility for *in vivo* localization studies.

MFP is a putative lipocalin-like protein that we have identified as binding both an [Fe-S] cluster and a heme moiety. We show that the [4Fe-4S] cluster of MFP inhibits heme binding to the protein and have proposed a potential structural model to explain this finding. Taken together, our data show a complicated metal-binding protein with a yet unknown *in vivo* function. The identification and initial characterization of MFP, a conserved protein essential for *Plasmodium* viability during the blood stage of infection, has provided a new potential therapeutic target for the treatment of malaria.

Thesis supervisor: Catherine L. Drennan

Title: Professor of Biology and Chemistry
Howard Hughes Medical Investigator and Professor

# ACKNOWLEDGEMENTS

As I look back on the journey that this thesis marks the end of, I can't help but think of how lucky I am to have been surrounded by so much support over the past six years. To my advisor Cathy Drennan, thank you for all of the support in and out of lab. The environment that you have cultivated has taught me so much about both science and life and I will take these lessons forward with me. To my thesis chair Liz Nolan, thank you for being as excited about my cluster as I am. I'm so happy that I got to work with you as my thesis chair. Your enthusiasm for my science and your thoughtful suggestions went a long way toward encouraging me in my scientific growth. To Jacquin Niles, thank you for bringing me in on this project and always having riveting discussions about MFP. Your ability to provide levity within the rigorous discussions has been very appreciated.

To the Drennan Lab, **thank you**. I have learned something from every single one of you. I'd especially like to thank a few people: Paty Feliciano for helping keep my expectations real and helping me learn all about [Fe-S] cluster biology in tropical diseases, Mary Andorfer for spending way too many hours with me in the sub-basement of 18 running an EPR instrument, Manny Ortega for the light that you bring to the room and the way you have shared your brilliance with all of us, Dante Avalos and Hannah Grupe for being as excited about these projects as I am and for keeping my spirits up during the typically rough time of writing a thesis, Francesca Vaccaro for taking over pretty much all of my non-lab tasks when I needed more time, Rohan Jonnalagadda for hysterical Monday morning discussions every fall, and Tsehai Grell for your kindness and help throughout the years. And of course, I could not finish my thank yous to the lab without thanking the Bae Bay for providing endless laughs as well as occasional scientific help. Somehow, we manage to be the loudest bay and yet still super productive.

To all of my friends outside of the Drennan Lab, whether I met you at MIT or through some other means, you have helped me. My friends have brought me joy, kept me grounded, mourned the failures, and celebrated the successes right along with me. I am so grateful to each and every one of you.

Lastly, my family. Mom and Dad, you both raised me to be a strong, independent woman and you are the reason I am where I am today. Thank you for always believing in me. Your support and encouragement have always been appreciated. Mom, thank you for showing me that women can be Doctors (of Philosophy) from a very early age. Both yours and dad's commitment to education was never lost on me. To my grandparents, thank you for being so loving. Especially to my Grandma Gladys, for taking care of me through most of my childhood. To Alex, my love, thank you for everything. I think you know this, but your constant cooking and taking care of me fueled this thesis and it definitely wouldn't have been complete without your support, so this is sort of like your second PhD. And of course, I can't end my thank yous without thanking the most important creature in my life, Daisy. Daisy, you can't read this, but I love you and the joy and friendships that you have brought me throughout this journey are unmatched.

*"Most importantly, you must always have faith in yourself." – Elle Woods*

# TABLE OF CONTENTS

Abstract ..... 3
Acknowledgements ..... 4
List of Figures ..... 10
List of Tables ..... 13

CHAPTER I: UTILIZATION OF HEME BY *PLASMODIUM* AND ITS UTILITY TOWARD NOVEL THERAPEUTICS... 14

Chapter Summary ..... 14
Malaria and its societal burden ..... 15
Malaria is caused by the parasite *Plasmodium* ..... 15
Emerging drug resistance ..... 15
*Plasmodium* protection from the erythrocyte ..... 17
Heme acquisition and utilization by *Plasmodium* ..... 17
The parasitophorous vacuole (PV) as a first line of defense ..... 18
Conclusion and thesis overview ..... 18
Figures ..... 19
References ..... 23

CHAPTER II: BIOINFORMATICS OF MFP, A PUTATIVE DRUG TARGET FOR THE TREATMENT OF MALARIA.. 25

Chapter Summary ..... 25
Contributions ..... 25
Introduction ..... 26
Results ..... 26
*Mfp* is a highly expressed, conserved gene across *Plasmodium* species expressing a 24.7 kDa protein ..... 26
*Mfp* encodes a putative lipocalin-like protein ..... 27
Homologs of *mfp* are found in other well-studied apicomplexa ..... 27
Similar proteins are putatively present in other infective organisms ..... 27
Discussion ..... 28

Conclusion ..... 29
Experimental Procedures ..... 29
Alignments and analyses of sequences ..... 29
Generation of MFP homology model ..... 30
Identification of MFP homologs ..... 30
Figures and tables ..... 31
References ..... 44

CHAPTER III: BIOCHEMICAL AND BIOPHYSICAL CHARACTERIZATION OF *PLASMODIUM* PROTEIN MFP ..... 46

Chapter Summary ..... 46
Contributions ..... 46
Introduction ..... 47
Results ..... 48
Recombinant MFP is a soluble, 24.7 kDa protein ..... 48
MFP binds an oxygen-sensitive iron-sulfur cluster ..... 48
MFP binds a heme moiety ..... 49
The presence of a [4Fe-4S] cluster limits heme binding by MFP ..... 50
Substitution of cysteines to alanine alters heme-binding properties under aerobic conditions ..... 50
Individual cysteine variants display altered aerobic heme binding ..... 51
MFP contains a disulfide bond between Cys70 and Cys216 ..... 51
Discussion ..... 52
Conclusion ..... 55
Experimental Procedures ..... 56
Expression of PfMFP variants ..... 56
Aerobic purification of His-MFP ..... 56
Anaerobic purification of His-MFP ..... 57
Aerobic purification of Strep-tagged MFP ..... 58
Anaerobic purification of Strep-tagged MFP ..... 58
Oligomeric analysis of MFP ..... 59

Reconstitution of a [4Fe-4S] cluster in PfMFP and metal analysis ..... 59
Electron paramagnetic resonance spectroscopy ..... 60
Heme binding assays ..... 60
Figures and tables ..... 62
References ..... 85

CHAPTER IV: ATTEMPTS TO STRUCTURALLY CHARACTERIZE MFP ..... 88

Chapter Summary ..... 88
Contributions ..... 88
Introduction ..... 89
Results ..... 89
His-MFP crystallizes in high alcohol solutions ..... 89
The crystallized form of apo-MFP contains a disulfide bond ..... 90
Heavy atom soaks of apo-MFP crystals have not yielded phase information ..... 91
[Fe-S] cluster-bound MFP has not yielded a structure ..... 92
Heme bound MFP has also not yielded a structure ..... 92
Discussion ..... 93
Conclusion ..... 95
Experimental Procedures ..... 95
Purification of proteins for crystallography ..... 95
Protein preparation for crystallization ..... 97
Protein crystallization ..... 98
Attempts at structure solution ..... 102
Identification of models for molecular replacement ..... 102
Mass spectrometry of MFP crystals ..... 103
Figures and tables ..... 104
References ..... 144

CHAPTER V: CHARACTERIZATION OF MFP HOMOLOGS ..... 145

Chapter Summary ..... 145

Contributions ..... 145
Introduction ..... 146
Results ..... 147
MFP homologs express well in *E. coli* ..... 147
PbMFP contains an iron-sulfur cluster ..... 148
PbMFP binds a heme moiety ..... 148
Attempts at crystallization of PbMFP ..... 148
Discussion ..... 149
Conclusion ..... 151
Experimental Procedures ..... 151
Expression of PbMFP ..... 151
Expression of non-*Plasmodium* MFP homologs ..... 152
Aerobic purification of MFP homologs ..... 152
Anaerobic purification of PbMFP ..... 152
Oligomeric analysis of MFP homologs ..... 153
Reconstitution of a [4Fe-4S] cluster in PbMFP and metal analysis ..... 153
Electron paramagnetic resonance (EPR) spectroscopy of PbMFP ..... 153
Heme binding assays ..... 154
Crystallization of PbMFP ..... 155
Figures ..... 156
References ..... 163

CHAPTER VI: STRUCTURAL STUDIES OF MRUBY2 TOWARD THE DESIGN OF A BIOLOGICALLY ENGINEERED HEME SENSOR ..... 164

Chapter Summary ..... 164
Contributions ..... 164
Introduction ..... 165
Results ..... 165
Binding of heme to mRuby2 quenches fluorescence ..... 165
mRuby2 fluoresces *in crystallo* ..... 166
mRuby and mRuby2 are similar in structure ..... 166

Attempts to structurally characterize heme-bound mRuby2 are on-going ..... 167
Discussion ..... 167
Conclusion ..... 169
Experimental Procedures ..... 169
Expression of mRuby2 ..... 169
Purification of mRuby2 ..... 170
Heme binding assays ..... 171
Crystallization of mRuby2 ..... 171
Determination of crystal fluorescence ..... 171
Data collection and structure determination of mRuby2 ..... 172
Crystallization of mRuby2 with hemin ..... 173
Figures and tables ..... 174
References ..... 184

CHAPTER VII: THE UTILIZATION OF IRON BINDING PROTEINS TOWARD THE INVESTIGATION AND TREATMENT OF *PLASMODIUM* ..... 186

MFP as a novel potential therapeutic target ..... 186
Much is yet to be discovered about MFP ..... 187
mRuby2 could be used as a heme sensor during the study of *Plasmodium* ..... 188
Conclusion ..... 189
References ..... 190

# LIST OF FIGURES

## CHAPTER I

**Figure I.1.** Malaria is endemic to tropical countries ..... 19

**Figure I.2.** The complicated *Plasmodium* life cycle uses two distinct hosts, the mosquito and a vertebrate host ..... 20

**Figure I.3.** The inhibition of hemozoin formation liberates more free-heme which is thought to activate artemisinin and its derivatives ..... 21

**Figure I.4.** The blood-stage parasitophorous vacuole is formed when *Plasmodium* enters the erythrocyte ..... 22

## CHAPTER II

**Figure II.1.** *Mfp* is highly expressed during blood stage infection ..... 31

**Figure II.2.** Putative protein sequence encoded by *mfp* ..... 32

**Figure II.3.** Sequence alignment of all *Plasmodium* MFP sequences ..... 33

**Figure II.4.** A homology model of MFP shows a lipocalin-like structure ..... 35

**Figure II.5.** MFP contains a variation on the canonical lipocalin motifs ..... 36

**Figure II.6.** An alignment of MFP homologs from *P. falciparum* and non-*Plasmodium* species shows a strong sequence conservation of the four cysteines ..... 37

## CHAPTER III

**Figure III.1.** Sequence alignment of MFP variants used in this work ..... 62

**Figure III.2.** Recombinantly expressed His-MFP from *Plasmodium falciparum* is soluble ..... 63

**Figure III.3.** Aerobically purified His-MFP is brown when purified with reducing agent ..... 64

**Figure III.4.** MFP binds an iron-sulfur cluster ..... 65

**Figure III.5.** Comparison of Analytical SEC traces of His-MFP and PfMFPΔ ..... 66

**Figure III.6.** Cluster incorporation appears to stabilize the dimer state ..... 67

**Figure III.7.** The presence of an iron-sulfur cluster is confirmed by EPR spectroscopy ..... 68

**Figure III.8.** The homology model of MFP supports the biochemically-determined iron-binding properties ..... 69

**Figure III.9.** Heme shows a characteristic red-shift by UV-Vis upon binding to protein ..... 70

**Figure III.10.** PfMFPΔ binds hemin under both aerobic and anoxic conditions ..... 71

**Figure III.11.** Heme bound PfMFPΔ runs as a trimer by SEC ..... 72

**Figure III.12.** Heme binding is reduced when PfMFPΔ is reconstituted with a [4Fe-4S] cluster under anoxic conditions ..... 73

**Figure III.13.** A small peak for nitrosyl heme is observed when 50% reconstituted PfMFPΔ is incubated with 0.5 equivalents of heme ..... 74

**Figure III.14.** The 4CA MFP variant shows similar binding of iron(II) heme as does PfMFPΔ, but a decrease in binding for iron(III) heme ..... 75

**Figure III.15.** The individual cysteine variants exhibit similar oligomeric states as wild-type ..... 76

**Figure III.16.** Under aerobic conditions, Cys68 and Cys93 MFP variants bind heme similarly to wild-type and 4CA MFP, respectively ..... 77

**Figure III.17.** Both the C70A and C216A variants display altered heme binding when compared to wild-type under aerobic conditions ..... 78

**Figure III.18.** Oxidized MFP runs lower on a non-reducing SDS-page gel ..... 79

**Figure III.19.** Substituting both Cys68 and Cys93 with alanine stabilizes the disulfide bond ..... 80

**Figure III.20.** Potential model of MFP cofactor binding ..... 81

**Figure III.21.** Structural alignment of the homology model of MFP with Nitrophorin 7 (NP7) shows Cys93 of the homology model near the Nitrophorin 7 His ligand ..... 82

**CHAPTER IV**

**Figure IV.1.** His-MFP crystallizes in high alcohol conditions ..... 104

**Figure IV.2.** His-MFP crystals diffract to ~2.0 Å resolution ..... 105

**Figure IV.3.** Crystals of His-MFP run lower on a gel than the expected molecular weight of apo-MFP ..... 106

**Figure IV.4.** A digest-mass spectrometry analysis confirms that the low molecular weight band is His-MFP ..... 107

**Figure IV.5.** Representative crystals from screens [Fe-S] cluster bound MFP ..... 108

**Figure IV.6.** Representative quasicrystals from screens with heme bound MFP ..... 109

**CHAPTER V**

**Figure V.1.** Non-*Plasmodium* homologs are insoluble under our purification conditions ..... 156

**Figure V.2.** Oligomeric profiles of MFP homologs by size exclusion chromatography are vastly different ..... 157

**Figure V.3.** Sequence alignment of PfMFP and PbMFP ..... 158

**Figure V.4.** PbMFP purifies aerobically with a [2Fe-2S] cluster ..... 159

**Figure V.5.** PbMFP will fully reconstitute to a [4Fe-4S] cluster under anaerobic conditions ..... 160

**Figure V.6.** PbMFP only binds heme in the apo state ..... 161

**Figure V.7.** Crystallization trials of PbMFP ..... 162

**CHAPTER VI**

**Figure VI.1.** mRuby2 differs from mRuby by 4 amino acids ..... 174

**Figure VI.2.** Strep-II tagged mRuby2 fluoresces under UV-light ..... 175

**Figure VI.3.** mRuby2 fluorescence is quenched by the presence of heme ..... 176

**Figure VI.4.** Titration of mRuby2 with heme shows that heme binds to mRuby2 ..... 177

**Figure VI.5.** Crystals of mRuby2 fluoresce under UV light ..... 178

**Figure VI.6.** The structure of mRuby2 shows a strong resemblance to mRuby and other fluorescent proteins ..... 179

**Figure VI.7.** The altered amino acids of mRuby2 are present in the structure of mRuby2 ..... 180

**Figure IV.8.** Some of the quasi-crystals that form when mRuby2 is incubated with heme fluoresce whereas others do not ..... 181

**Figure VI.9.** Asp214 could be affecting the fluorophore through its adjacent residue, His215 ..... 182

# LIST OF TABLES

## CHAPTER II

**Table II.1.** Sequence identity of MFP across different *Plasmodium* species compared to that of *Plasmodium falciparum* ..... 38

**Table II.2.** Genes identified as potential protein homologs of *mfp* ..... 39

## CHAPTER III

**Table III.1.** PfMFP and its variants can adopt many different oligomeric states ..... 83

**Table III.2.** Individual cysteine variants display different heme-binding properties ..... 84

## CHAPTER IV

**Table IV.1.** Many variants of MFP have been used for attempted crystallization ..... 110

**Table IV.2.** Apo-MFP has been screened against many aerobic conditions in an attempt to obtain data-quality crystals ..... 111

**Table IV.3.** Crystallization conditions of crystals for which continuous scan data sets have been collected ..... 120

**Table IV.4.** Data statistics from a single His-MFP crystal (N5900_748_13) processed in different space groups ..... 126

**Table IV.5.** Data statistics for all static collection data sets collected as reported by RAPD automatic processing suite and comments from Xtriage ..... 127

**Table IV.6.** Multiple heavy atoms were attempted for derivatizing crystals ..... 133

**Table IV.7.** Data statistics for data sets collected by inverse beam in search of anomalous signal ..... 134

**Table IV.8.** Anaerobically purified and reconstituted MFP has been screened against many anaerobic conditions in an attempt to obtain data-quality crystals of a [4Fe-4S] cluster-bound MFP ..... 136

**Table IV.9.** Many known structures of lipocalins have been used in an attempt to solve the structure of MFP by molecular replacement ..... 141

## CHAPTER VI

**Table VI.1.** Data and refinement statistics for the current model of mRuby2 ..... 183

# CHAPTER I: UTILIZATION OF HEME BY *PLASMODIUM* AND ITS UTILITY TOWARD NOVEL THERAPEUTICS

## CHAPTER SUMMARY

Malaria is an endemic disease caused by the eukaryotic parasite *Plasmodium*. Increasing resistance to current therapeutics is driving the need for new methods of control as well as novel drug target identification. Given the utilization of chloroquine derivatives in artemisinin combination therapies as well as the increasing resistance of *Plasmodium* to chloroquine, investigation into the heme regulation of *Plasmodium* could be of use to develop novel therapeutics for pairing with artemisinin. Herein, we address *Plasmodium* as a parasite and describe the heme utilization of *Plasmodium* and its potential utility toward novel therapeutic development.

## MALARIA AND ITS SOCIETAL BURDEN

Malaria, a potentially deadly disease transmitted to humans by mosquitos, is costly to human health as well as economies(1-3). Nearly half of the world's population is at risk of contracting malaria; every year, almost half a million people die from infection (**Figure I.1**) (1,4). The global cost of malaria is at least $12 billion per year, with $3.1 billion of that being spent on control and elimination efforts alone (1,4,5). The majority of infections occur in regions lacking the monetary resources to handle this burden, such as Sub-Saharan Africa and South Asia (1,6). Although there has been a general reduction in malaria cases in recent years, these improvements have stalled, prompting a need for additional methods of eradicating the disease (1).

### MALARIA IS CAUSED BY THE PARASITE *PLASMODIUM*

The causative agent of malaria is *Plasmodium*, a genus of parasitic unicellular eukaryotes (1,5-8). The over 120 species of plasmodium have the ability to infect various different vertebrate hosts, but all of them share common features, including the method of infection and transmission (2,6). Five species of *Plasmodium* are known to infect humans: *P. falciparum*, *P. vivax*, *P. malariae*, *P. ovale*, and *P. knowlesi* (1,6,9-11). *Plasmodium* has three main life stages, the mosquito stage, the liver stage, and the blood stage, with the latter two stages occurring while in the vertebrate host (**Figure I.2**) (3). Because of this, transmission of malaria requires both the intermediate mosquito vector as well as the main vertebrate host (6).

The multiple life stages of *Plasmodium* provide a wealth of potential targets; however, the ability of the parasite to lie dormant in different life stages, especially *P. vivax*, complicates the therapeutic process (6). This ability to temporarily evade therapeutics, coupled with compliance issues based on structural and economic challenges in areas where malaria is endemic, has caused growing drug resistance, and prompted the necessity for alternative methods of vector control and novel therapeutics (1,12).

### EMERGING DRUG RESISTANCE

There are many pharmaceuticals currently available for the treatment of malaria; however, increasing resistance by *Plasmodium* is prompting the necessity of combination therapies and the

development of novel therapeutics (1). Problems of resistance are compounded by the fact that the areas in which malaria is endemic tend to lack the monetary resources to obtain new or more expensive therapeutics (13,14). Single drug therapies are the least expensive option and therefore highly utilized in these countries, but resistance to these compounds develops quickly (12). Due to the high relative cost, many patients may not finish a full course of medication or may be fooled into purchasing fake medication, leading to quicker resistance development (12,15). Additionally, since clinical symptoms of malaria can be observed during the blood stage of *Plasmodium* infection, including fever, chills, nausea, fatigue, and, in severe cases, death, many patients will discontinue therapy when symptoms subside, even though asymptomatic infection may still be present (16,17). This practice adds to drug resistance due to compliance issues in a similar fashion to antibiotic resistance (17).

*P. falciparum* is generally thought to be the most dangerous species, causing the majority of yearly deaths attributed to *Plasmodium* infection, likely because it also has conferred the most drug resistance of all human-infecting species (1). This and other resistant strains have prompted the need for combination therapies. Artemisinin combination therapy is the current gold standard, with partial resistance to artemisinin only being reported in one high-risk malaria region, the Greater Mekong subregion (1). These combination therapies often utilize an artemisinin derivative and a quinine derivative, and are currently highly effective for treating most strains of *Plasmodium* (3,18). The two drugs have different clearance rates, with the artemisinin derivative causing rapid clearance of the parasite and the quinine derivative acting on a longer time scale to clear more latent parasites (3). An additional utility of this combination lies in the synergy of function of the two drugs. Quinine and its derivatives prevent the formation of hemozoin crystals, a storage mechanism of *Plasmodium* to reduce the amount of toxic free-heme available to the cell, and artemisinin derivatives seem to be activated by heme-derived iron (**Figure I.3**) (3,19-22). An important challenge in the future of malaria treatment and eradication is that *P. falciparum* and *P. vivax* have evolved wide-spread resistance to chloroquine and its derivatives, prompting the need for additional studies into the metalome and iron trafficking within *Plasmodium* to design novel therapeutics for paired use with artemisinin (12).

## *PLASMODIUM* PROTECTION FROM THE ERYTHROCYTE

During the blood stage of malaria, the parasite resides within the host erythrocyte, a location not generally conducive to organismal growth (3). The host immune system response and the iron rich environment of the erythrocyte have prompted *Plasmodium* to develop strategies to cope with the harsh conditions. Two important strategies during the blood stage of malaria are heme regulation and the formation of a parasitophorous vacuole (PV).

### HEME ACQUISITION AND UTILIZATION BY *PLASMODIUM*

During the blood-stage of infection, *Plasmodium* can acquire approximately 80% of the hemoglobin present in host erythrocytes (23). The main utility of this uptake is to acquire amino acids for protein synthesis; however, the heme byproduct of degradation provides an interesting challenge for the protozoan. Although there are multiple potential pathways for heme detoxification, the main pathway is the formation of hemozoin crystals, large, solid crystals made up of heme, which are non-toxic to the parasite (24).

In addition to acquisition of heme from the host erythrocyte and potential utilization of heme from the hemozoin crystals, *Plasmodium* contains a heme biosynthetic pathway that is essential during the liver and mosquito stage of the parasitic life cycle (**Figure I.2**) (25,26). Interestingly, although *Plasmodium* contains two distinct pathways to gain heme, the *Plasmodium* genome contains only a small number of proteins identified as hemoproteins based on sequence homology (23). Most are cytochrome or cytochrome b-like proteins, and although a few additional proteins have been identified as binding heme *in vitro*, no evidence of actual heme binding has been observed *in vivo* (23,27). Additionally, no known heme degradation enzyme has been identified in the *Plasmodium* genome as of yet (28). Given the relative dearth of identified pathways for heme utilization or any known function for the hemozoin crystals other than detoxification, it is worth investigating additional potential heme utilization or degradation avenues within *Plasmodium*.

### THE PARASITOPHOROUS VACUOLE (PV) AS A FIRST LINE OF DEFENSE

In addition to the hemozoin formation, *Plasmodium* utilize an additional method of protection, formation of the PV (29). During the blood stage of infection, the parasite enters the erythrocyte by co-opting a part of the erythrocytic membrane and creating a PV (**Figure I.4**) (30). This vacuole is thought to provide the first line of defense for *Plasmodium* against the toxicity of the extracellular space of the erythrocyte (31). Although the PV is poorly understood, new techniques in bioinformatics are helping to identify the composition of the PV proteome (32,33).

## CONCLUSION AND THESIS OVERVIEW

Malaria, the potentially fatal disease caused by *Plasmodium*, is an important area of study to improve global health. Increasing resistance to current therapies is prompting the necessity for novel drugs (12). In order to develop these therapeutics, a stronger understanding of *Plasmodium* biology is needed. In this thesis, we describe the work performed toward furthering our understanding of *Plasmodium falciparum* through the study of two heme binding proteins. Chapter II describes the work performed to identify a particular protein of interest in *Plasmodium falciparum*, which we have termed MFP. Chapter III details biochemical studies performed on MFP, and the identification of two iron-containing binding partners. The nature of the relationship between these iron-containing binding partners is explored and we outline the effect of a few key amino acid substitutions on this protein. Chapter IV outlines our efforts toward obtaining a structure of this essential protein. Chapter V explores four homologs of MFP, including one from *Plasmodium berghei* and three from non-*Plasmodium* parasites. Chapter VI outlines our work toward understanding the heme-binding properties of mRuby2, an engineered fluorescent probe commonly used in protein localization assays, and the implications of its heme-binding properties on biological research. Chapter VII summarizes the key findings of this thesis and discusses next steps and the implications of this research.

FIGURES

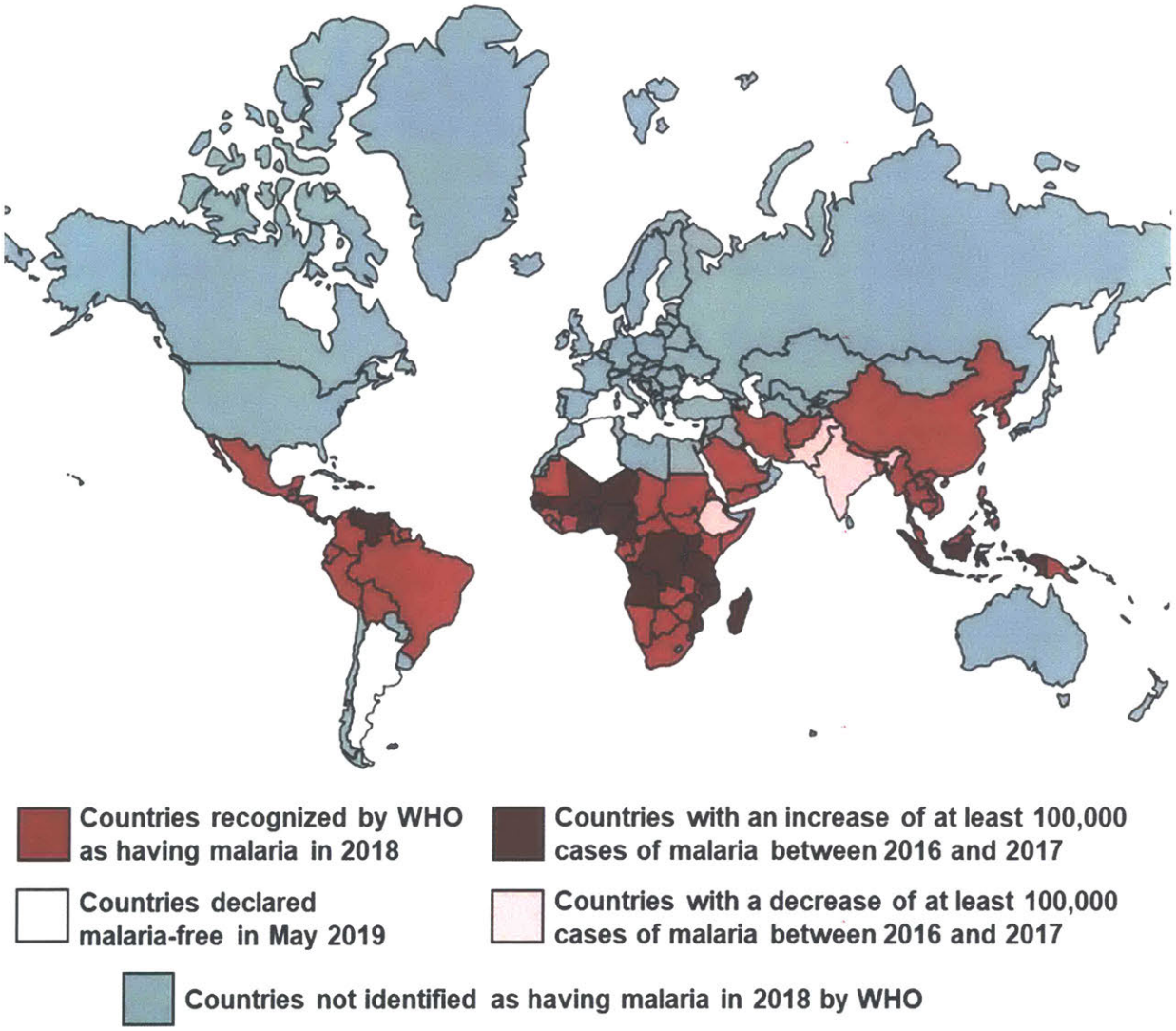



**Figure I.1. Malaria is endemic to tropical countries.** Countries displayed in red and salmon have high incidences of malaria. Dark red countries reported at least 300,000 cases and an increase of at least 100,000 cases between 2016 and 2017. Salmon countries reported at least 300,000 cases and a decrease of at least 100,000 cases between 2016 and 2017. Two countries, Algeria and Argentina, were declared malaria-free in May of 2019 and are displayed in light pink (34). Data displayed are from the World Health Organization (WHO) malaria country profile data as well as the World Malaria Report 2018 from WHO (1,35).

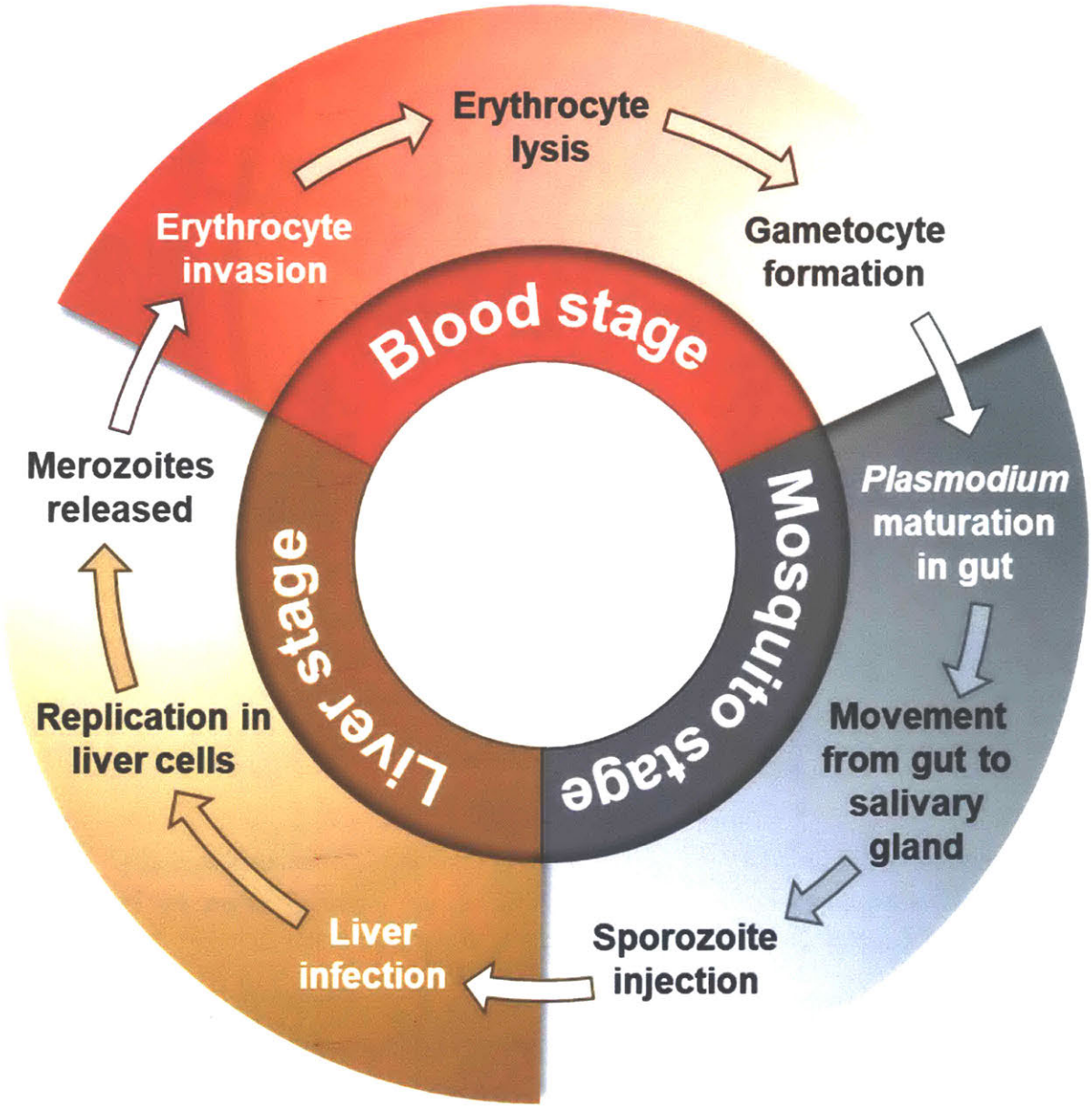



**Figure I.2. The complicated *Plasmodium* life cycle uses two distinct hosts, the mosquito and a vertebrate host.** The liver stage (brown) and the blood stage (red) are both within the vertebrate host while the mosquito stage is present in the mosquito transmission vector. White text signifies the beginning of the stage and white arrows signify the transition between life stages. Figure adapted from Mullard (2018) (36).

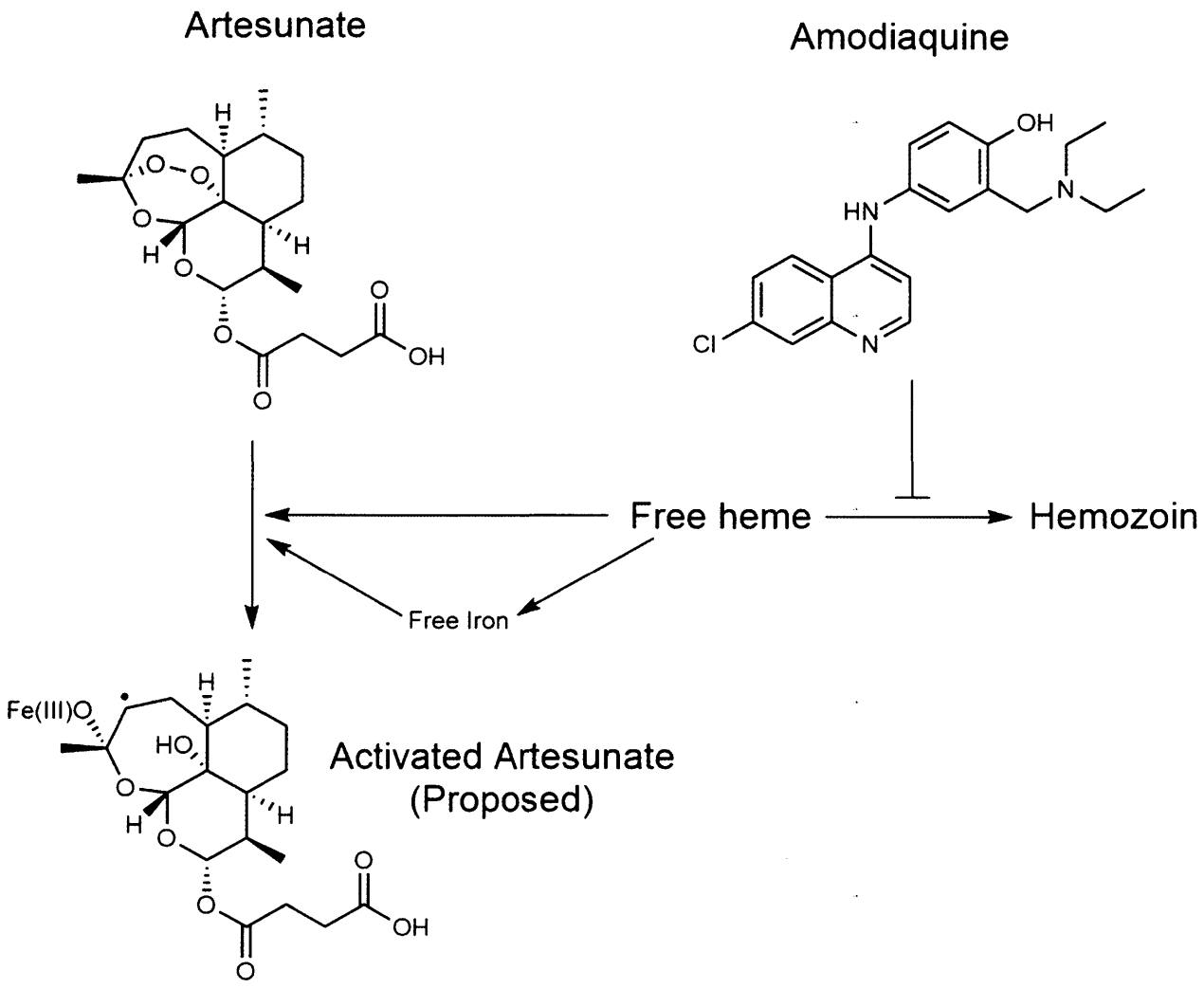



**Figure I.3. The inhibition of hemozoin formation liberates more free-heme which is thought to activate artemisinin and its derivatives.** It is thought that artemisinin and its derivatives, including artesunate, are activated by cleavage of the endoperoxide by either heme or heme-derived iron (37). The combination of a quinine derivative, such as amodiaquine, which inhibits hemozoin formation, and an artemisinin derivative, which is thought to be activated by free heme or heme derived iron, provides a synergistic effect for parasite clearance (20,37). The structures shown here are of artesunate (artemisinin derivative) and amodiaquine (quinine derivative), a pairing that has treated over 400 million patients (3).

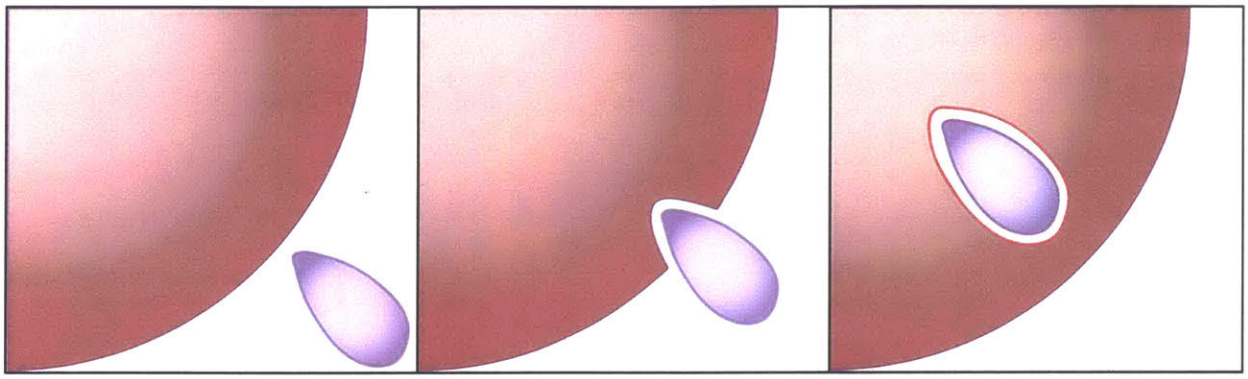


**Figure I.4. The blood-stage parasitophorous vacuole is formed when *Plasmodium* enters the erythrocyte. (Left)** *Plasmodium*, displayed in purple, attaches itself to the erythrocyte, displayed in red. **(Center)** The *Plasmodium* begins to enter the erythrocyte by pushing against the erythrocytic cell membrane. **(Right)** When the *Plasmodium* is fully within the erythrocyte, it closes off the portion of the erythrocytic cell membrane surrounding it, which forms the new parasitophorous vacuole membrane. The region between this new membrane and the *Plasmodium* is termed the parasitophorous vacuole.

# REFERENCES


1. World Health Organization. (2018) World malaria report 2018. World Health Organization
2. Cowman, A. F., Healer, J., Marapana, D., and Marsh, K. (2016) Malaria: Biology and Disease. *Cell* **167**, 610-624
3. Phillips, M. A., Burrows, J. N., Manyando, C., Huijsduijnen, R. H. v., Voorhis, W. C. V., and Wells, T. N. C. (2017) Malaria. *Nature Reviews Disease Primers* **3**
4. Centers for Disease Control and Prevention. (2019) Malaria's Impact Worldwide.
5. Greenwood, B. M., Bojang, K., Whitty, C. J., and Targett, G. A. (2005) Malaria. *The Lancet* **365**, 1487-1498
6. Ashley, E. A., Pyae Phyo, A., and Woodrow, C. J. (2018) Malaria. *The Lancet* **391**, 1608-1621
7. White, N. J., Pukrittayakamee, S., Hien, T. T., Faiz, M. A., Mokuolu, O. A., and Dondorp, A. M. (2014) Malaria. *The Lancet* **383**, 723-735
8. Cox, F. E. (2010) History of the discovery of the malaria parasites and their vectors. *Parasites & Vectors* **3**, 5
9. Cox-Singh, J., Davis, T. M. E., Lee, K. S., Shamsul, S. S. G., Matusop, A., Ratnam, S., Rahman, H. A., Conway, D. J., and Singh, B. (2008) Plasmodium knowlesi Malaria in Humans Is Widely Distributed and Potentially Life Threatening. *Clinical Infectious Diseases* **46**, 165-171
10. Ginouves, M., Veron, V., Musset, L., Legrand, E., Stefani, A., Prevot, G., Demar, M., Djossou, F., Brousse, P., Nacher, M., and Carme, B. (2015) Frequency and distribution of mixed Plasmodium falciparum-vivax infections in French Guiana between 2000 and 2008. *Malaria Journal* **14**
11. Howes, R. E., Reiner Jr, R. C., Battle, K. E., Longbottom, J., Mappin, B., Ordanovich, D., Tatem, A. J., Drakeley, C., Gething, P. W., Zimmerman, P. A., Smith, D. L., and Hay, S. I. (2015) Plasmodium vivax Transmission in Africa. **9**, e0004222
12. Menard, D., and Dondorp, A. (2017) Antimalarial Drug Resistance: A Threat to Malaria Elimination. a025619
13. Worrall, E., Hanson, K., and Basu, S. (2009) *The Relationship Between Socio-Economic Status and Malaria: A Review of the Literature*,
14. Girum, T., Hailemikael, G., and Wondimu, A. (2017) Factors affecting prevention and control of malaria among endemic areas of Gurage zone: an implication for malaria elimination in South Ethiopia, 2017. *Trop Dis Travel Med Vaccines* **3**, 17-17
15. Nayyar, G. M., Breman, J. G., Newton, P. N., and Herrington, J. (2012) Poor-quality antimalarial drugs in southeast Asia and sub-Saharan Africa. *The Lancet Infectious Diseases* **12**, 488-496
16. White, N. J., Pongtavornpinyo, W., Maude, R. J., Saralamba, S., Aguas, R., Stepniewska, K., Lee, S. J., Dondorp, A. M., White, L. J., and Day, N. P. (2009) Hyperparasitaemia and low dosing are an important source of anti-malarial drug resistance. *Malaria Journal* **8**, 253
17. White, N. J. (2004) Antimalarial drug resistance. *Journal of Clinical Investigation* **113**, 1084-1092
18. World Health Organization. (2015) *Guidelines for the treatment of malaria. Third edition*,

19. Olafson, K. N., Nguyen, T. Q., Rimer, J. D., and Vekilov, P. G. (2017) Antimalarials inhibit hematin crystallization by unique drug–surface site interactions. *Proceedings of the National Academy of Sciences* **114**, 7531-7536
20. Fong, K. Y., and Wright, D. W. (2013) Hemozoin and antimalarial drug discovery. *Future Medicinal Chemistry* **5**, 1437-1450
21. Klonis, N., Crespo-Ortiz, M. P., Bottova, I., Abu-Bakar, N., Kenny, S., Rosenthal, P. J., and Tilley, L. (2011) Artemisinin activity against Plasmodium falciparum requires hemoglobin uptake and digestion. **108**, 11405-11410
22. Tilley, L., Straimer, J., Gnädig, N. F., Ralph, S. A., and Fidock, D. A. (2016) Artemisinin Action and Resistance in Plasmodium falciparum. *Trends in Parasitology* **32**, 682-696
23. Sigala, P. A., and Goldberg, D. E. (2014) The Peculiarities and Paradoxes of Plasmodium Heme Metabolism. *Annual Review of Microbiology* **68**, 259-278
24. Egan, T. J., Combrinck, J. M., Egan, J., Hearne, G. R., Marques, H. M., Ntenteni, S., Sewell, B. T., Smith, P. J., Taylor, D., Van Schalkwyk, D. A., and Walden, J. C. (2002) Fate of haem iron in the malaria parasite Plasmodium falciparum. *Biochemical Journal* **365**, 343-347
25. Ke, H., Sigala, P. A., Miura, K., Morrisey, J. M., Mather, M. W., Crowley, J. R., Henderson, J. P., Goldberg, D. E., Long, C. A., and Vaidya, A. B. (2014) The Heme Biosynthesis Pathway Is Essential for Plasmodium falciparum Development in Mosquito Stage but Not in Blood Stages. **289**, 34827-34837
26. Goldberg, D. E., and Sigala, P. A. (2017) Plasmodium heme biosynthesis: To be or not to be essential? *PLOS Pathogens* **13**, e1006511
27. Kořený, L., Oborník, M., and Lukeš, J. (2013) Make It, Take It, or Leave It: Heme Metabolism of Parasites. *PLoS Pathogens* **9**, e1003088
28. Sigala, P. A., Crowley, J. R., Hsieh, S., Henderson, J. P., and Goldberg, D. E. (2012) Direct Tests of Enzymatic Heme Degradation by the Malaria ParasitePlasmodium falciparum. *Journal of Biological Chemistry* **287**, 37793-37807
29. Lingelbach, K., and Joiner, K. A. (1998) The parasitophorous vacuole membrane surrounding Plasmodium and Toxoplasma: an unusual compartment in infected cells. *Journal of Cell Science* **111**, 1467-1475
30. Cowman, A. F., and Crabb, B. S. (2006) Invasion of Red Blood Cells by Malaria Parasites. *Cell* **124**, 755-766
31. Schatten, H., and Ris, H. (2004) Three-Dimensional Imaging of Toxoplasma gondii–Host Cell Interactions within the Parasitophorous Vacuole. *Microscopy and Microanalysis* **10**, 580-585
32. Matz, J. M., and Matuschewski, K. (2018) An in silico down-scaling approach uncovers novel constituents of the Plasmodium-containing vacuole. *Scientific Reports* **8**
33. Schnider, C. B., Bausch-Fluck, D., Brühlmann, F., Heussler, V. T., and Burda, P.-C. (2018) BioID Reveals Novel Proteins of the Plasmodium Parasitophorous Vacuole Membrane. *mSphere* **3**, e00522-00517
34. (2019) Algeria and Argentina certified malaria-free by WHO. World Health Organization
35. World Health Organization. Malaria Country Profiles, 2018. World Health Organization
36. Mullard, A. (2018) Malaria medicine box expands. *Nature Reviews Drug Discovery* **17**, 693
37. O'Neill, P. M., Barton, V. E., and Ward, S. A. (2010) The Molecular Mechanism of Action of Artemisinin—The Debate Continues. *Molecules* **15**, 1705-1721

# CHAPTER II: BIOINFORMATICS OF MFP, A PUTATIVE DRUG TARGET FOR THE TREATMENT OF MALARIA

## CHAPTER SUMMARY

Emerging resistance of first-line therapeutics is one of the core challenges hindering the treatment and eradication of malaria. In order to combat this resistance, we identified a promising protein target for the development of novel therapeutics. Our collaborators in the Niles lab identified the gene PF3D7_0925900 as having high amounts of mRNA relative to other genes during the blood stage of malaria. Additionally, this gene is conserved across all *Plasmodium* species. The gene encodes a putative 24.7 kDa lipocalin-like protein. Homologous proteins are found in other apicomplexa as well as across many other organisms, including parasites and organisms of unknown function. Based on all of this information and a recent report that the homologous gene in *Plasmodium berghei* is essential, MFP could prove to be an interesting target for novel pharmaceutical development.

## CONTRIBUTIONS

The *mfp* gene was identified as a target of interest by Mai Nakashima (Niles Lab, MIT). An initial homology model was created by Mai Nakashima (Niles Lab, MIT). All other work in this chapter was performed by Allena Mistral Goren (Drennan Lab, MIT).

## INTRODUCTION

Malaria is a deadly disease causing approximately half a million deaths yearly worldwide (1,2). One of the largest problems associated with treatment of malaria is emerging drug resistance by *Plasmodium*, the family of parasites that causes malaria (1,3). *Plasmodium falciparum*, *Plasmodium vivax*, and *Plasmodium malariae* have all been reported to exhibit drug resistance (1,3,4). The majority of strains are currently immune to chloroquine, and a few multidrug resistant strains exist (3,4). This resistance is driving a need for novel therapeutics.

One way of driving the discovery of novel compounds for treatment of malaria is to study the biology of the parasite that causes the disease. *Plasmodium falciparum* is a eukaryotic parasite belonging to the apicomplexa phylum (5). *Plasmodium* parasites use mosquitos as the vector of transmission, and reside in both the liver and the blood stage in their vertebrate host (6,7). Although *Plasmodium* research has recently become of exceptional interest, we still have only scratched the surface toward understanding the biology of this parasite. In order to further efforts toward the prevention and cure of malaria, we undertook studies to identify potential new, biologically relevant drug targets for *Plasmodium*.

## RESULTS

### *MFP* IS A HIGHLY EXPRESSED, CONSERVED GENE ACROSS *PLASMODIUM* SPECIES EXPRESSING A 24.7 KDA PROTEIN

In our efforts to identify a protein of interest for therapeutic development, we identified a gene, PF3D7_0925900, which we have termed *mfp*, which is both highly expressed during the blood stage of malaria (**Figure II.1**) and conserved across *Plasmodium* species (**Table II.1**) (8). This gene, containing a single intron, expresses a putative 24.7 kDa protein (**Figure II.2**), which we have termed MFP (9). This gene encodes for a putative signal peptide and a single domain of unknown function (9). The MFP protein encoded by *mfp* exhibits high sequence conservation, containing many conserved residues and high overall sequence identity across all species of *Plasmodium* (**Figure II.3, Table II.1**).

## *MFP* ENCODES A PUTATIVE LIPOCALIN-LIKE PROTEIN

An initial homology model of MFP was generated by HHPred and Modeller (10), indicating that MFP is likely a lipocalin-like protein (**Figure II.4**). The lipocalin family of proteins consist of two subfamilies, the kernel lipocalins and the outlier lipocalins (11). Although kernel lipocalin proteins contain three conserved motifs with high sequence identity, many of the outlier lipocalin proteins possess variations on these three motifs (11). MFP possesses a very similar motif to that of motif 1 and motif 3, but does not seem to have motif 2 (**Figure II.5**), which is typical of the outlier lipocalins (11). The lipocalin family of proteins contains a structurally conserved β-barrel, usually with a larger opening on one side than the other, as well as a short $3_{10}$ -helix and a longer canonical α-helix. Many of the lipocalin-like proteins contain one or more disulfide bonds, so we additionally located the four cysteines present in MFP in our homology model. All four cysteines are within approximately 11 Å of each other (**Figure II.4C**) with a disulfide bond forming between Cys70 and Cys216.

## HOMOLOGS OF *MFP* ARE FOUND IN OTHER WELL-STUDIED APICOMPLEXA

Three additional putative MFP proteins were identified during a BLAST search (12) of non-*Plasmodium* proteins. The identified proteins were found in *Hammondia hammondi*, *Neospora caninum*, and *Toxoplasma gondii*. Each of these three organisms falls within the apicomplexa family (5). Like MFP, these proteins are annotated in gene databases as putative proteins of unknown function (13). An alignment of the three homologous proteins with MFP from *Plasmodium falciparum* shows approximately 20% sequence identity (**Figure II.6**). The three homologous proteins appear to have much stronger sequence conservation to each other than to MFP from *P. falciparum*.

## SIMILAR PROTEINS ARE PUTATIVELY PRESENT IN OTHER INFECTIVE ORGANISMS

To identify a broader swath of potential MFP homologs, we performed a BLAST search (12) using the three homologs identified in **Figure II.6**. A CXC motif observed in MFP from *Plasmodium* and all three homologs appeared in many of the proteins identified in this wider

search and was used to cull the sequences identified as potential homologs. In addition to the three homologs used for this search, 61 putative proteins were identified as potential homologs for MFP (**Table II.2**). Although many of these were found in *Bacteroidetes* of unknown origin or function, some were found in infective organisms including bacteria and fungi (5).

## DISCUSSION

Emerging drug resistance of *Plasmodium* is making the treatment of malaria more difficult (1,3). As such, we endeavored to identify a potential new drug target for the treatment of malaria caused by *Plasmodium falciparum*. We identified the conserved, highly expressed gene PF3D7_0925900, or *mfp*, as an interesting candidate for study. To inform our structural and biochemical experiments, we analyzed the putative protein MFP encoded by this gene using bioinformatic techniques and homology modeling.

The MFP protein was identified as a lipocalin-like protein by sequence similarity. This family of proteins has three conserved motifs, but contains minimal sequence conservation beyond that (11). Unfortunately, the diversity in function of known lipocalin proteins does not help us narrow down the *in vivo* function of this protein. As such, we continued our search for information on the function of MFP using its primary amino acid sequence.

Through a BLAST search (12), we were able to identify three additional homologs of MFP from *Hammondia hammondi*, *Neospora caninum*, and *Toxoplasma gondii*, all apicomplexa parasites (5). Interestingly, it was recently identified that *mfp* is an essential gene of *Plasmodium berghei* and localizes to the parasitophorous vacuole (PV), an extracellular compartment utilized by apicomplexa as protection from the host-cell environment (14). These findings could indicate a conserved function of this protein in either the infectivity or protection of these parasites. Although we were only able to locate this gene in four members of apicomplexa, we suspect the presence of this gene in more, less well-studied apicomplexa as many of these organisms have not had their genomes sequenced as of yet.

To determine whether homologs of this protein could be present in other organisms, we performed a BLAST search (12) using the three non-*Plasmodium* homologs. This search identified many more proteins, including 61 additional proteins that contained the same CXC

motif identified to be conserved across the proteins identified in apicomplexa. The 61 genes identified come from a broad swath of organisms, including bacteria, fungi, and multi-cellular organisms. Although many of the organisms identified to contain a putative MFP homolog are poorly studied, a few of these putative proteins are interesting due to the organisms in which they were found. The *Anopheles albimanus* is a mosquito known to carry *Plasmodium*, and though the potentially homologous protein identified is 992 amino acids, much larger than the 217 residues in MFP, it is noteworthy due to the organism's carrier status (15). Interestingly, *Chromera* and *Vitrella* are unicellular organisms with potential uses to study apicomplexa due to their evolutionary proximity to the phylum (16,17). Lastly, 16% of the species identified to contain a potential MFP homolog are obligate, facultative, or putative anaerobes (18-22). This finding, paired with the presence of the four conserved and physically close cysteine residues, could indicate the use of an iron-sulfur cluster by this protein.

## CONCLUSION

We identified a highly expressed gene of unknown function, PF3D7_0925900, which is conserved across all species of *Plasmodium*. This gene encodes for a putative lipocalin-like protein with four highly conserved cysteines predicted to be within close proximity of each other. An additional 64 homologs were also identified in other infective organisms, including three in other apicomplexa and two in a closely related organism used to study the evolution of apicomplexa. Taken together, our findings suggest that PF3D7_0925900 is a good candidate for study as a potential drug target for malaria.

## EXPERIMENTAL PROCEDURES

### ALIGNMENTS AND ANALYSES OF SEQUENCES

Sequences were aligned either by MultAlin (23) based on the primary structure or by PROMALS3D (24) based on the secondary structure of the proteins. All displays of sequence alignments were created using ESPript 3.0 (25). Identities and similarities were calculated using the Sequence Manipulation Suite (26) based on MFP from *Plasmodium falciparum* as a template.

## GENERATION OF MFP HOMOLOGY MODEL

A homology model of MFP from *Plasmodium falciparum* was generated using the sequence shown in **Figure II.2** without the signal peptide or putatively disordered region. Residues 33-216 of MFP from *P. falciparum* were input into the HHPred server (10). Sequences with greater than 90% probability, 53 in total, were forwarded to the Modeller server (10) to generate a homology model.

## IDENTIFICATION OF MFP HOMOLOGS

To identify potential homologs of MFP from *Plasmodium falciparum*, a BLAST search was carried out using the NCBI BLAST tool (12) and excluding all *Plasmodium* proteins. The three proteins identified are listed in **Figure II.6**. To identify more potential homologs, a BLAST search was carried out using the UniProt BLAST tool (27). Sequences found from this search were included in **Table II.2** if they contained the CXC motif found in MFP.

FIGURES AND TABLES

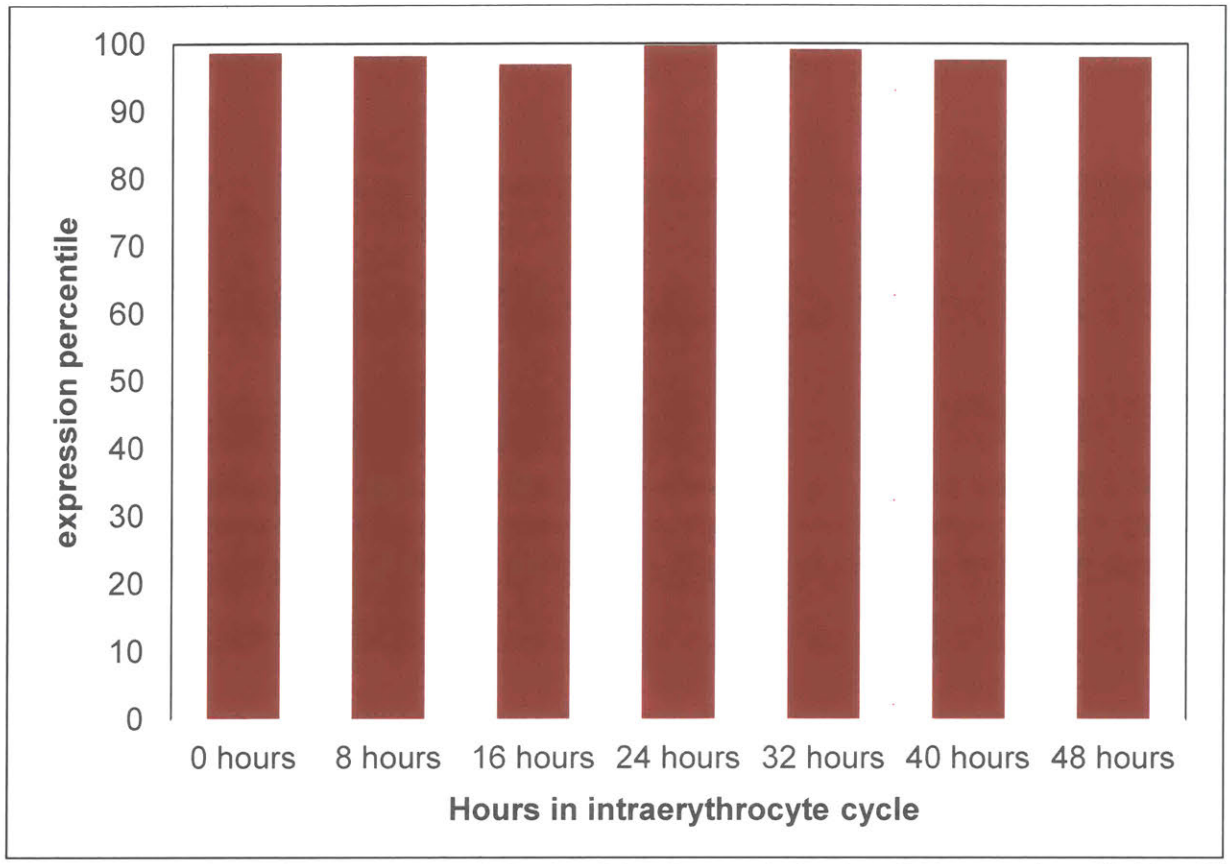



**Figure II.1. *Mfp* is highly expressed during blood stage infection.** RNA-seq data (8) identified MFP as being highly expressed (>96% percentile of expression for all time points) during the blood stage as compared to other genes monitored. Data was acquired through PlasmoDB (9).

```
1           11          21          31          41
MIKKAVSLFF IFFLFTIIRA SPGNDNNNVN NDGVVHVLSK NIDVKNLQGT

51          61          71          81          91
FYEIATNASD KIFPGLACRC TKYEFSGLKR DGNLGYVLIN FSCARNFIFG

101         111         121         131         141
EKKSEMTFKL ILNKPLDENT TTVEEFNASI YLVQGNQQIL LNGNINIIYA

151         161         171         181         191
ELNEQNEFEH LILGGQKSIE PMIIMSKYRT VLLDTYNKLI NSLYLAGYEP

201         211
SLLTWPFIIQ TDQTFCD
```

**Figure II.2. Putative protein sequence encoded by *mfp*.** The sequence has been annotated as containing a signal peptide thought to localize the protein to the parasitophorous vacuole (bold) and a single domain of unknown function (9). A putatively disorder, Asx region is highlighted in red.

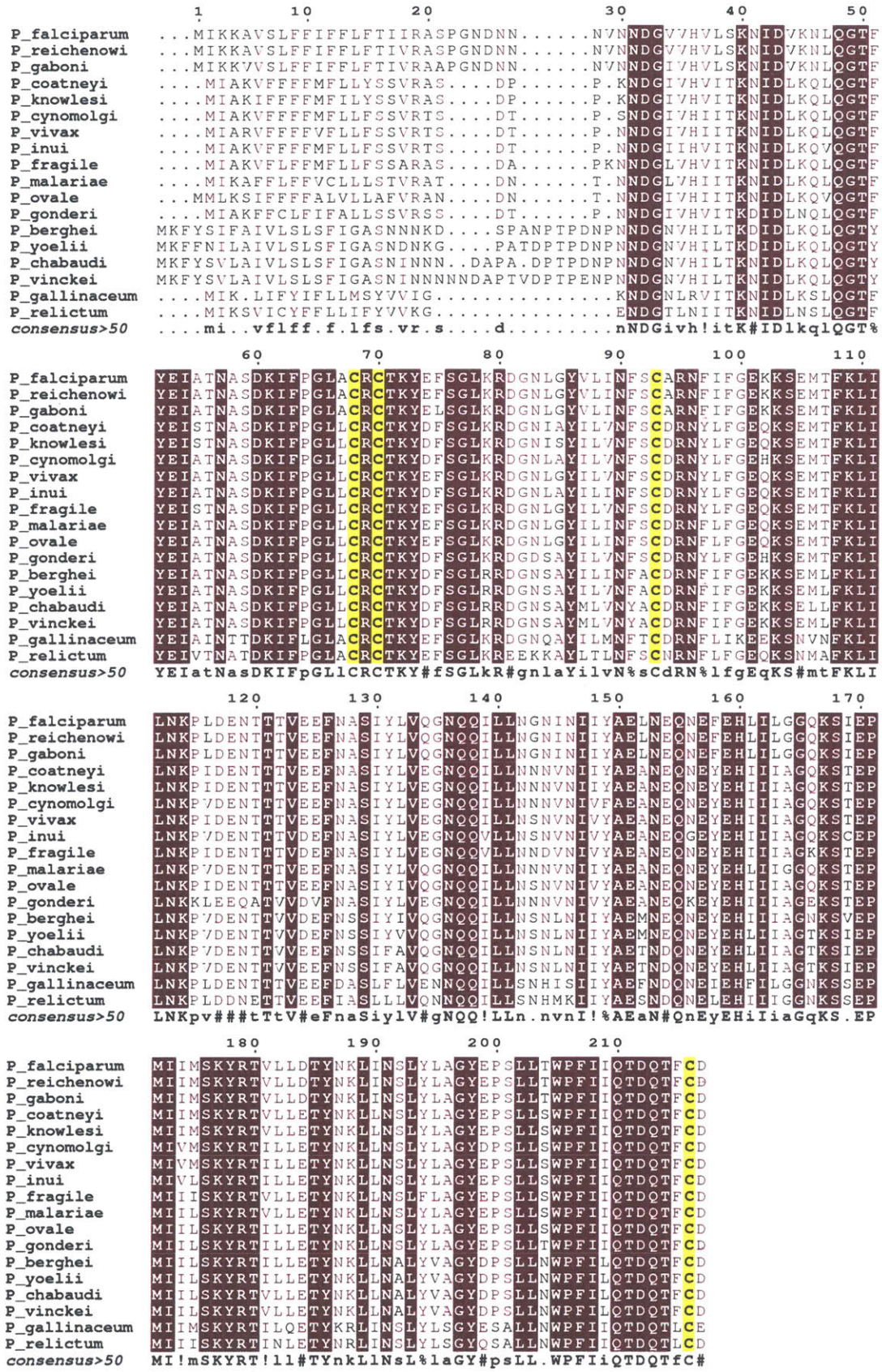

P_falciparum
P_reichenowi
P_gaboni
P_coatneyi
P_knowlesi
P_cynomolgi
P_vivax
P_inui
P_fragile
P_malariae
P_ovale
P_gonderi
P_berghei
P_yoelii
P_chabaudi
P_vinckei
P_gallinaceum
P_relictum
consensus>50

**Figure II.3. Sequence alignment of all *Plasmodium* MFP sequences.** There is a high sequence conservation among MFP from different species of *Plasmodium*, including four highly conserved cysteine residues highlighted in yellow. Residue numbers correspond to those from *P. falciparum*. Additional conserved residues are highlighted in red. Similar residues are in red text. Sequence alignment was generated with MultAlin (23) and figure was created using ESPript (25).

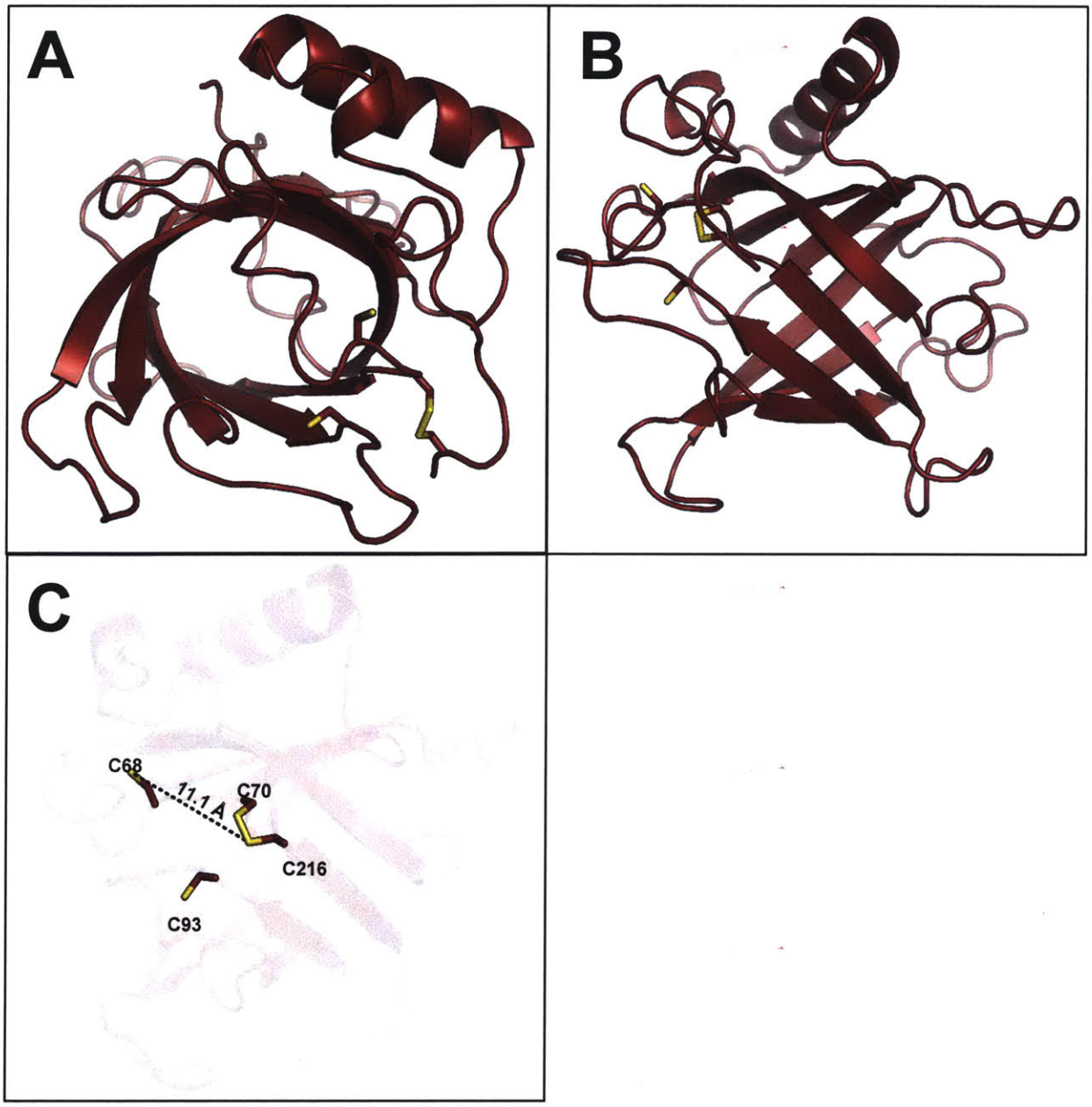



**Figure II.4. A homology model of MFP shows a lipocalin-like structure.** Based on sequence homology, it is likely that MFP contains a β-barrel as well as two α-helices forming a lipocalin-like fold. **(A)** Looking down the barrel, two flexible loops are observed, which could be acting as lids for the open β-barrel. **(B)** The side view of the homology model of MFP appears to be a typical lipocalin-like fold. **(C)** The four conserved cysteines are all less than 12 Å away from each other. The farthest distance between cysteines in the homology model is 11.1 Å between Cys68 and Cys216 and is shown in a black dashed line. The homology model of MFP was generated with HHpred (10). The four conserved cysteine residues of MFP are shown in stick representation.

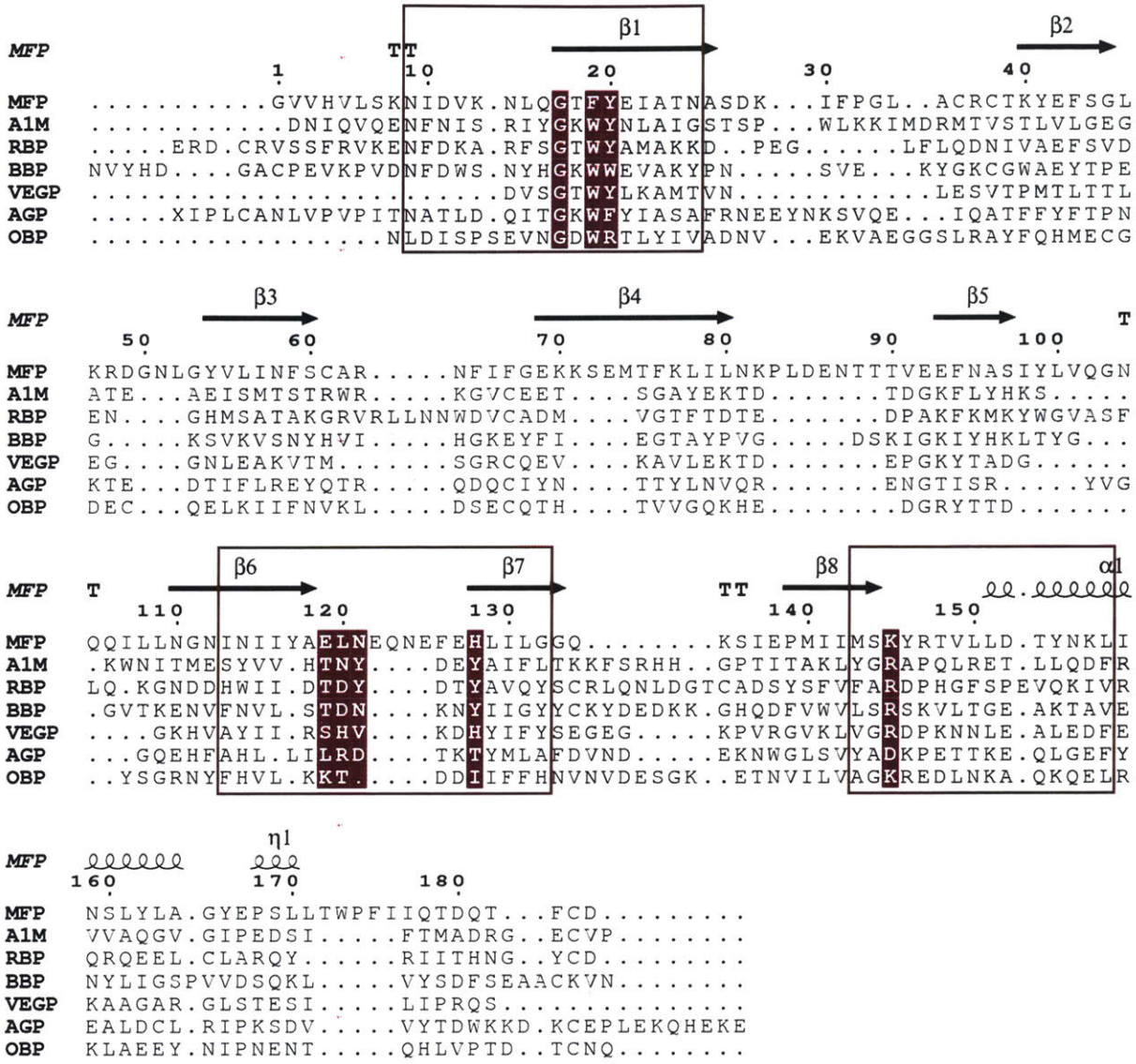



**Figure II.5. MFP contains a variation on the canonical lipocalin motifs.** Based on a structure-based sequence alignment (24) of MFP to other well-characterized lipocalins, there is minimal conservation of lipocalin motifs. Structurally conserved regions 1, 2, and 3 are boxed and labeled accordingly, and motifs 1, 2, and 3 are highlighted in red. The secondary structure is that of the MFP homology model. Motif 1 corresponds to residues near the $3_{10}$-helix, not present in the MFP homology model, and $\beta_1$; motif 2 corresponds to residues in $\beta_6$ and $\beta_7$; and motif 3 is a single amino acid toward the C-terminal side of $\beta_8$. Non-MFP sequences were chosen based on previously reported kernel and outlier members of the family (11). Sequence alignment was based on known structures of $\alpha_1$-microglobulin (PDB 3QKG), retinol-binding protein (PDB 1AQB), bilin-binding protein (PDB 1BBP), von Ebners gland protein (PDB 1XKY), $\alpha_1$-acid glycoprotein (PDB 3KQ0), odorant-binding protein (PDB 3FIQ), and the homology model of MFP.

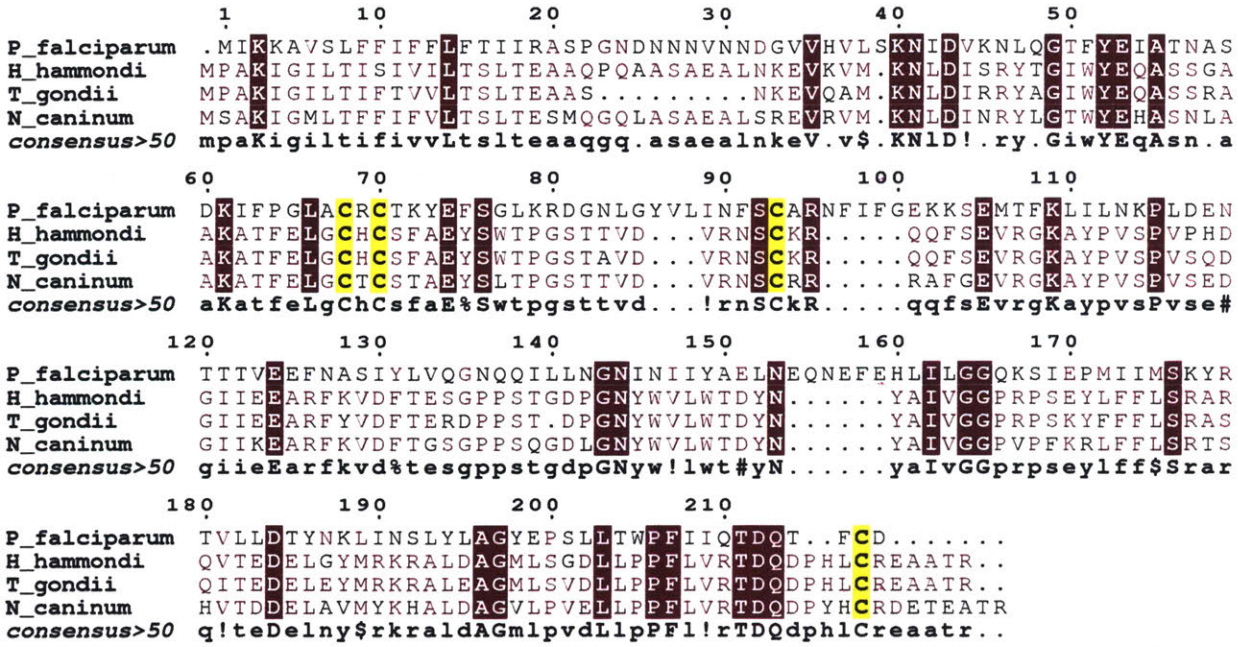



**Figure II.6. An alignment of MFP homologs from *P. falciparum* and non-*Plasmodium* species shows a strong sequence conservation of the four cysteines.** Only 40 residues total from the 217 amino acid protein are conserved across the four apicomplexa. Coloring is as in **Figure II.3.**

| ***Plasmodium* species** | **Identity to *P. falciparum*** | **Similarity to *P. falciparum*** | **Vertebrate host** |
|---|---|---|---|
| *P. reichenowi* | 99.54% | 100.00% | Primate (28) |
| *P. gaboni* | 97.70% | 99.08% | |
| *P. coatneyi* | 77.98% | 90.37% | Primate (28) |
| ***P. knowlesi*** | **77.52%** | **89.91%** | **Primate (28)** |
| *P. cynomalogi* | 76.50% | 91.24% | Primate (28) |
| ***P. vivax*** | **78.34%** | **91.71%** | **Primate (28)** |
| *P. inui* | 76.04% | 90.32% | Primate (28) |
| *P. fragile* | 77.98% | 90.37% | |
| ***P. malariae*** | **82.11%** | **91.28%** | **Primate (28)** |
| ***P. ovale*** | **78.28%** | **90.05%** | **Primate (28)** |
| *P. gonderi* | 72.94% | 87.16% | Primate (28) |
| *P. berghei* | 71.68% | 84.96% | Rodent (28) |
| *P. yoelii* | 69.87% | 82.10% | Rodent (28) |
| *P. chabaudi* | 70.61% | 82.46% | Rodent (28) |
| *P. vinckei* | 71.00% | 82.25% | Rodent (28) |
| *P. gallinaceum* | 66.52% | 77.83% | Bird (28) |
| *P. relictum* | 65.61% | 78.73% | Bird (28) |

**Table II.1. Sequence identity of MFP across different *Plasmodium* species compared to that of *Plasmodium falciparum*.** In addition to *Plasmodium falciparum*, bolded species are those that infect humans. Identity and similarity values were calculated using the Sequence Manipulation Suite (26).

| Gene | Organism | Taxonomic Lineage |
|---|---|---|
| D9V86_07125 | *Bacteroidetes/Chlorobi group bacterium ChocPot_Mid* | Bacteria › |
| A2041_02185 | *Bacteroidetes bacterium GWA2_31_9b* | Bacteria › Bacteroidetes |
| A2X13_09270 | *Bacteroidetes bacterium GWC2_33_15* | Bacteria › Bacteroidetes |
| A2W99_17515 | *Bacteroidetes bacterium GWF2_33_16* | Bacteria › Bacteroidetes |
| CVT94_11420 | *Bacteroidetes bacterium HGW-Bacteroidetes-11* | Bacteria › Bacteroidetes |
| CFE21_10720 | *Bacteroidetes bacterium B1(2017)* | Bacteria › Bacteroidetes › |
| CVU04_04735 | *Bacteroidetes bacterium HGW-Bacteroidetes-20* | Bacteria › Bacteroidetes › |
| DCL86_02100 | *Bacteroidales bacterium* | Bacteria › Bacteroidetes › Bacteroidia › Bacteroidales |
| DCQ24_02215 | *Bacteroidales bacterium* | Bacteria › Bacteroidetes › Bacteroidia › Bacteroidales |
| DF409_03380 | *Bacteroidales bacterium* | Bacteria › Bacteroidetes › Bacteroidia › Bacteroidales › |
| TBC1_11433 | *Lentimicrobium saccharophilum* | Bacteria › Bacteroidetes › Bacteroidia › Bacteroidales › Lentimicrobiaceae › Lentimicrobium |
| CJD36_021475 | *Flavipsychrobacter stenotrophus* | Bacteria › Bacteroidetes › Chitinophagia › Chitinophagales › Chitinophagaceae › Flavipsychrobacter |
| AHMF7616_03186 | *Adhaeribacter sp. HMF7616* | Bacteria › Bacteroidetes › Cytophagia › Cytophagales › Hymenobacteraceae › Adhaeribacter |
| AHMF7605_21020 | *Adhaeribacter sp. HMF7605* | Bacteria › Bacteroidetes › Cytophagia › Cytophagales › Hymenobacteraceae › Adhaeribacter |
| SAMN04487941_2179 | *Pontibacter akesuensis* | Bacteria › Bacteroidetes › Cytophagia › Cytophagales › Hymenobacteraceae › Pontibacter |
| DXT99_08695 | *Pontibacter diazotrophicus* | Bacteria › Bacteroidetes › Cytophagia › Cytophagales › Hymenobacteraceae › Pontibacter |

| Gene | Organism | Taxonomic Lineage |
|---|---|---|
| SAMN05444128_2323 | *Pontibacter indicus* | Bacteria › Bacteroidetes › Cytophagia › Cytophagales › Hymenobacteraceae › Pontibacter |
| PKOR_04810 | *Pontibacter korlensis* | Bacteria › Bacteroidetes › Cytophagia › Cytophagales › Hymenobacteraceae › Pontibacter |
| SAMN05421545_0678 | *Pontibacter lucknowensis* | Bacteria › Bacteroidetes › Cytophagia › Cytophagales › Hymenobacteraceae › Pontibacter |
| BD749_2036 | *Pontibacter ramchanderi* | Bacteria › Bacteroidetes › Cytophagia › Cytophagales › Hymenobacteraceae › Pontibacter |
| O71_14426 | *Pontibacter sp. BAB1700* | Bacteria › Bacteroidetes › Cytophagia › Cytophagales › Hymenobacteraceae › Pontibacter |
| D1627_09455 | *Pontibacter sp. KIRAN* | Bacteria › Bacteroidetes › Cytophagia › Cytophagales › Hymenobacteraceae › Pontibacter |
| A3841_14835 | *Pontibacter sp. S10-8* | Bacteria › Bacteroidetes › Cytophagia › Cytophagales › Hymenobacteraceae › Pontibacter |
| SAMN06296052_11784 | *Pontibacter ummariensis* | Bacteria › Bacteroidetes › Cytophagia › Cytophagales › Hymenobacteraceae › Pontibacter |
| C8E01_101437 | *Pontibacter virosus* | Bacteria › Bacteroidetes › Cytophagia › Cytophagales › Hymenobacteraceae › Pontibacter |
| EFA69_10880 | *Rufibacter immobilis* | Bacteria › Bacteroidetes › Cytophagia › Cytophagales › Hymenobacteraceae › Rufibacter |
| TH61_08865 | *Rufibacter sp. DG15C* | Bacteria › Bacteroidetes › Cytophagia › Cytophagales › Hymenobacteraceae › Rufibacter |
| DC20_05370 | *Rufibacter tibetensis* | Bacteria › Bacteroidetes › Cytophagia › Cytophagales › Hymenobacteraceae › Rufibacter |

| Gene | Organism | Taxonomic Lineage |
|---|---|---|
| TH63_15075 | *Rufibacter sp. DG31D* | Bacteria › Bacteroidetes › Cytophagia › Cytophagales › Hymenobacteraceae › Rufibacter |
| EFB08_03895 | *Rufibacter sp. R-22-1c-1* | Bacteria › Bacteroidetes › Cytophagia › Cytophagales › Hymenobacteraceae › Rufibacter |
| SAMN05421739_1093 | *Pontibacter chinhatensis* | Bacteria › Bacteroidetes › Cytophagia › Cytophagales › Hymenobacteraceae› Pontibacter |
| C8N40_101568 | *Pontibacter mucosus* | Bacteria › Bacteroidetes › Cytophagia › Cytophagales › Hymenobacteraceae› Pontibacter |
| C7N43_08660 | *Sphingobacteriales bacterium UPWRP_1* | Bacteria › Bacteroidetes › Sphingobacteriia › Sphingobacteriales |
| C7N43_14605 | *Sphingobacteriales bacterium UPWRP_1* | Bacteria › Bacteroidetes › Sphingobacteriia › Sphingobacteriales |
| blc_1 | *Mucilaginibacter gotjawali* | Bacteria › Bacteroidetes › Sphingobacteriia › Sphingobacteriales › Sphingobacteriaceae › Mucilaginibacter |
| blc | *Candidatus Aminicenantes bacterium ADurb.Bin147* | Bacteria › Candidatus Aminicenantes |
| A2Y56_13390 | *Candidatus Aminicenantes bacterium RBG_13_63_10* | Bacteria › Candidatus Aminicenantes › |
| COT43_02495 | *Candidatus Marinimicrobia bacterium CG08_land_8_20_14_0_20_45_22* | Bacteria › Candidatus Marinimicrobia |
| CMF23_04495 | *Ignavibacteriae bacterium* | Bacteria › Ignavibacteriae |
| CVV24_08535 | *Ignavibacteriae bacterium HGW-Ignavibacteriae-3* | Bacteria › Ignavibacteriae › |
| A2V66_13790 | *Ignavibacteria bacterium RBG_13_36_8* | Bacteria › Ignavibacteriae › Ignavibacteria |
| C4543_11040 | *Ignavibacteriales bacterium* | Bacteria › Ignavibacteriae › Ignavibacteria › Ignavibacteriales › |
| CO129_00065 | *Ignavibacteriales bacterium CG_4_9_14_3_um_filter_34_10* | Bacteria › Ignavibacteriae › Ignavibacteria › Ignavibacteriales › |

| Gene | Organism | Taxonomic Lineage |
|---|---|---|
| DCF63_07740 | *Planctomycetaceae bacterium* | Bacteria › Planctomycetes › Planctomycetia › Planctomycetales › P lanctomycetaceae › |
| C4563_07250 | *Desulfobulbus sp.* | Bacteria › Proteobacteria › Deltaproteobacteria › Desulfobacterales› Desulfobulbaceae › Desulfobulbus |
| EDC27_1891 | *Desulfosoma caldarium* | Bacteria › Proteobacteria › Deltaproteobacteria › Syntrophobacterales › Syntrophobacteraceae › Desulfosoma |
| CL816_00065 | *Coxiellaceae bacterium* | Bacteria › Proteobacteria › Gammaproteobacteria › Legionellales › Coxiellaceae › |
| SAMN05216202_4245 | *Pseudomonas mucidolens* | Bacteria › Proteobacteria › Gammaproteobacteria › Pseudomonadales › Pseudomonadaceae › Pseudomonas |
| CF386_09245 | *Paraphotobacterium marinum* | Bacteria › Proteobacteria › Gammaproteobacteria › Vibrionales › Vibrionaceae › Paraphotobacterium |
| SAMN06296036_11666 | *Pseudobacteriovorax antillogorgiicola* | Bacteria › Proteobacteria › Oligoflexia › Oligoflexales › Pseudobacteriovoracaceae › Pseudobacteriovorax |
| HHA_453540 | *Hammondia hammondi (Parasitic protozoan)* | Eukaryota › Alveolata › Apicomplexa › Conoidasida › Coccidia › Eucoccidiorida › Eimeriorina › Sarcocystidae › Hammondia |
| BN1204_001720 | *Neospora caninum (strain Liverpool)* | Eukaryota › Alveolata › Apicomplexa › Conoidasida › Coccidia › Eucoccidiorida › Eimeriorina › Sarcocystidae › Neospora › |
| BN1205_012995 | *Toxoplasma gondii (strain ATCC 50861 / VEG)* | Eukaryota › Alveolata › Apicomplexa › Conoidasida › Coccidia › Eucoccidiorida › Eimeriorina › Sarcocystidae › Toxoplasma |
| Cvel_10704 | *Chromera velia CCMP2878* | Eukaryota › Alveolata › Chromerida › Chromera |
| Cvel_134 | *Chromera velia CCMP2878* | Eukaryota › Alveolata › Chromerida › Chromera › |

| Gene | Organism | Taxonomic Lineage |
|---|---|---|
| Vbra_22046 | *Vitrella brassicaformis (strain CCMP3155)* | Eukaryota › Alveolata › Chromerida › Vitrella |
| PV10_00774 | *Exophiala mesophila (Black yeast)* | Eukaryota › Fungi › Dikarya › Ascomycota › Pezizomycotina › Eurotiomycetes › Chaetothyriomycetidae › Chaetothyriales › Herpotrichiellaceae › Exophiala |
| UCDDA912_g02941 | *Diaporthe ampelina* | Eukaryota › Fungi › Dikarya › Ascomycota › Pezizomycotina › Sordariomycetes › Sordariomycetidae › Diaporthales › Diaporthaceae › Diaporthe |
| VM1G_08678 | *Valsa mali* | Eukaryota › Fungi › Dikarya › Ascomycota › Pezizomycotina › Sordariomycetes › Sordariomycetidae › Diaporthales › Valsaceae › Valsa |
| 20346122 | *Gaeumannomyces graminis var. tritici (strain R3-111a-1) (Wheat and barley take-all root rot fungus)* | Eukaryota › Fungi › Dikarya › Ascomycota › Pezizomycotina › Sordariomycetes › Sordariomycetidae › Magnaporthales › Magnaporthaceae › Gaeumannomyces › |
| N/A | *Corethrella appendiculata* | Eukaryota › Metazoa › Ecdysozoa › Arthropoda › Hexapoda › Insecta › Pterygota › Neoptera › Holometabola › Diptera › Nematocera › Culicoidea › Chaoboridae › Corethrella |
| N/A | *Anopheles albimanus (New world malaria mosquito)* | Eukaryota › Metazoa › Ecdysozoa › Arthropoda › Hexapoda › Insecta › Pterygota › Neoptera › Holometabola › Diptera › Nematocera › Culicoidea › Culicidae › Anophelinae › Anopheles |
| ASZ90_005084 | *hydrocarbon metagenome* | unclassified sequences › metagenomes › ecological metagenomes |
| blc_1 | *mine drainage metagenome* | unclassified sequences › metagenomes › ecological metagenomes |

**Table II.2. Genes identified as potential protein homologs of *mfp*.** Genes were identified through a UniProt BLAST search against MFP from *H. hammondi*, *N. caninum*, and *T. gondii*.

# REFERENCES


1. World Health Organization. (2018) World malaria report 2018. World Health Organization
2. Centers for Disease Control and Prevention. (2019) Malaria's Impact Worldwide.
3. Menard, D., and Dondorp, A. (2017) Antimalarial Drug Resistance: A Threat to Malaria Elimination. a025619
4. World Health Organization. (2018) Antimalarial drug efficacy and drug resistance.
5. Sayers, E. W., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., Church, D. M., Dicuccio, M., Edgar, R., Federhen, S., Feolo, M., Geer, L. Y., Helmberg, W., Kapustin, Y., Landsman, D., Lipman, D. J., Madden, T. L., Maglott, D. R., Miller, V., Mizrachi, I., Ostell, J., Pruitt, K. D., Schuler, G. D., Sequeira, E., Sherry, S. T., Shumway, M., Sirotkin, K., Souvorov, A., Starchenko, G., Tatusova, T. A., Wagner, L., Yaschenko, E., and Ye, J. (2009) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* **37**, D5-D15
6. Phillips, M. A., Burrows, J. N., Manyando, C., Huijsduijnen, R. H. v., Voorhis, W. C. V., and Wells, T. N. C. (2017) Malaria. *Nature Reviews Disease Primers* **3**
7. Ashley, E. A., Pyae Phyo, A., and Woodrow, C. J. (2018) Malaria. *The Lancet* **391**, 1608-1621
8. Otto, T. D., Wilinski, D., Assefa, S., Keane, T. M., Sarry, L. R., Böhme, U., Lemieux, J., Barrell, B., Pain, A., Berriman, M., Newbold, C., and Llinás, M. (2010) New insights into the blood-stage transcriptome of Plasmodium falciparum using RNA-Seq. *Molecular Microbiology* **76**, 12-24
9. Aurrecoechea, C., Brestelli, J., Brunk, B. P., Dommer, J., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G., Harb, O. S., Heiges, M., Innamorato, F., Iodice, J., Kissinger, J. C., Kraemer, E., Li, W., Miller, J. A., Nayak, V., Pennington, C., Pinney, D. F., Roos, D. S., Ross, C., Stoeckert, C. J., Treatman, C., and Wang, H. (2009) PlasmoDB: a functional genomic database for malaria parasites. **37**, D539-D543
10. Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A. N., and Alva, V. (2018) A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *Journal of Molecular Biology* **430**, 2237-2243
11. Flower, D. R. (1996) The lipocalin protein family: structure and function. *Biochemical Journal* **318**, 1-14
12. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-410
13. Gajria, B., Bahl, A., Brestelli, J., Dommer, J., Fischer, S., Gao, X., Heiges, M., Iodice, J., Kissinger, J. C., Mackey, A. J., Pinney, D. F., Roos, D. S., Stoeckert, C. J., Wang, H., and Brunk, B. P. (2007) ToxoDB: an integrated Toxoplasma gondii database resource. **36**, D553-D556
14. Matz, J. M., and Matuschewski, K. (2018) An in silico down-scaling approach uncovers novel constituents of the Plasmodium-containing vacuole. *Scientific Reports* **8**
15. Giraldo-Calderon, G. I., Emrich, S. J., Maccallum, R. M., Maslen, G., Dialynas, E., Topalis, P., Ho, N., Gesing, S., Madey, G., Collins, F. H., and Lawson, D. (2015)

VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases. **43**, D707-D713

16. Moore, R. B., Oborník, M., Janouškovec, J., Chrudimský, T., Vancová, M., Green, D. H., Wright, S. W., Davies, N. W., Bolch, C. J. S., Heimann, K., Šlapeta, J., Hoegh-Guldberg, O., Logsdon, J. M., and Carter, D. A. (2008) A photosynthetic alveolate closely related to apicomplexan parasites. *Nature* **451**, 959-963
17. Obornik, M., Modry, D., Lukes, M., Cernotikova-Stribrna, E., Cihlar, J., Tesarova, M., Kotabova, E., Vancova, M., Prasil, O., and Lukes, J. (2012) Morphology, ultrastructure and life cycle of Vitrella brassicaformis n. sp., n. gen., a novel chromerid from the Great Barrier Reef. *Protist* **163**, 306-323
18. Sun, L., Toyonaga, M., Ohashi, A., Tourlousse, D. M., Matsuura, N., Meng, X.-Y., Tamaki, H., Hanada, S., Cruz, R., Yamaguchi, T., and Sekiguchi, Y. (2016) Lentimicrobium saccharophilum gen. nov., sp. nov., a strictly anaerobic bacterium representing a new family in the phylum Bacteroidetes, and proposal of Lentimicrobiaceae fam. nov. *International Journal of Systematic and Evolutionary Microbiology* **66**, 2635-2642
19. Kadnikov, V. V., Mardanov, A. V., Beletsky, A. V., Karnachuk, O. V., and Ravin, N. V. (2019) Genome of the candidate phylum Aminicenantes bacterium from a deep subsurface thermal aquifer revealed its fermentative saccharolytic lifestyle. *Extremophiles* **23**, 189-200
20. Podosokorskaya, O. A., Kadnikov, V. V., Gavrilov, S. N., Mardanov, A. V., Merkel, A. Y., Karnachuk, O. V., Ravin, N. V., Bonch-Osmolovskaya, E. A., and Kublanov, I. V. (2013) Characterization of Melioribacter roseusgen. nov., sp. nov., a novel facultatively anaerobic thermophilic cellulolytic bacterium from the class Ignavibacteria, and a proposal of a novel bacterial phylum Ignavibacteriae. *Environmental Microbiology* **15**, 1759-1771
21. Laanbroek, H. J., and Pfennig, N. (1981) Oxidation of short-chain fatty acids by sulfate-reducing bacteria in freshwater and in marine sediments. *Archives of Microbiology* **128**, 330-335
22. Baena, S., Perdomo, N., Carvajal, C., Diaz, C., and Patel, B. K. C. (2011) Desulfosoma caldarium gen. nov., sp. nov., a thermophilic sulfate-reducing bacterium from a terrestrial hot spring. *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY* **61**, 732-736
23. Corpet, F. (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research* **16**, 10881-10890
24. Pei, J., Kim, B.-H., and Grishin, N. V. (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Research* **36**, 2295-2300
25. Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research* **42**, W320-W324
26. Stothard, P. (2000) The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* **28**, 1102-1104
27. (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research* **47**, D506-D515
28. Perkins, S. L., and Schall, J. J. (2002) A molecular phylogeny of malarial parasites recovered from cytochrome *b* gene sequences. *Journal of Parasitology* **88**, 972-978

# CHAPTER III: BIOCHEMICAL AND BIOPHYSICAL CHARACTERIZATION OF *PLASMODIUM* PROTEIN MFP

## CHAPTER SUMMARY

In Chapter II, we identified a putative protein of interest in *Plasmodium falciparum*. The gene encoding for this protein, PF3D7_0925900, is thought to be highly expressed during the blood stage due to the high amounts of mRNA from this gene during the blood stage of *Plasmodium* infection (1). A homologous gene in *Plasmodium berghei* has also recently been found to be essential for viability during this period (2). Interestingly, this gene is conserved across *Plasmodium* species, highlighting its potential importance for *Plasmodium* infection. Characterization of this gene at the biochemical level could lead to further understanding of the Plasmodium infection cycle leading to putative therapies. In this work, we describe the biochemical characterization of the protein expressed by PF3D7_0925900, a potential drug target for combatting *Plasmodium* infection. Under both aerobic and anoxic conditions, the 25 kDa protein binds a heme cofactor with an apparent 1:1 stoichiometric ratio. Under anoxic, reducing conditions, the recombinant protein also binds a [4Fe-4S] cluster in an apparent 1:1 ratio. Incorporation of the [4Fe-4S] cluster into MFP negatively impacts heme binding. The oligomeric profiles of the heme-bound and cluster-bound protein differ, implying that a shift of oligomeric state could be the mechanism precluding dual cofactor binding. Our investigations of a cluster-binding deficient variant as well as amino acid substituted variants also provide insight into the residues used by this protein to bind heme. Given the interesting iron binding properties of the protein expressed by PF3D7_0925900, we have named this gene *mfp* which stands for Malarial Ferrous Protein (MFP).

## CONTRIBUTIONS

Plasmids were provided by Mai Nakashima and Khan Osman (Niles Lab, MIT). Recombinant expression and purification of all proteins were performed by Hannah Grupe and Allena Mistral Goren (Drennan Lab, MIT). Anaerobic work was performed by Allena Mistral Goren. Mary C. Andorfer (Drennan Lab, MIT) helped with the collection of EPR spectra. Heme binding assays were performed by Hannah Grupe, Dante Avalos, and Allena Mistral Goren (Drennan Lab, MIT). EPR experiments were performed in the Department of Chemistry Instrumentation facility under the supervision of John Grimes and Walt Massefski (MIT). Khan Osman (Niles lab, MIT) discovered the disulfide bond. We would like to thank Professor Daniel Suess (MIT), Constantin Voll (The 3M Company), and Alex Taguchi (RubrYc Therapeutics) for their help in learning EPR spectroscopy techniques and theory.

## INTRODUCTION

Malaria is a parasitic infection endemic to tropical locations causing fever, chills, nausea and fatigue (3-5). These symptoms are usually observed during the blood stage of *Plasmodium* infection (5-7). During this stage of infection, the parasite must adapt and thrive under conditions not normally conducive to cellular growth: the extracellular matrix in which the *Plasmodia* survive is iron-rich and the host's immune system readily identifies the parasite as a foreign invader. In order to survive, *Plasmodia* have evolved different strategies to deal with such conditions.

Although many unicellular organisms would not survive in the iron rich environment of the erythrocyte, *Plasmodia* survive by tightly regulating their iron homeostasis. During this blood stage, the heme-biosynthesis pathway in *Plasmodium* is not necessary for cellular viability as the organism scavenges over 80% of the heme from individual host erythrocytes (8). Much of this free heme is converted to hemozoin, thus greatly reducing the toxicity (9).

In addition to this iron regulation-based protection, *Plasmodia* form an additional buffer zone known as the parsitophorous vacuole (PV, **See Figure I.4**) (10). This extracellular compartment is currently not well characterized, but is thought to contain the first line of defense for the parasite against the host immune response (11). The known PV proteome is still likely under-annotated; however new advances in gene and protein identification continue to expand our knowledge of protein localization within this specialized compartment(2,12).

Recently, a homolog of PF3D7_0925900 from *Plasmodium berghei* has been identified as essential for the parasite's survival during the blood stage of infection, and was termed parasitophorous vacuole protein 5, or PV5, due to the protein's localization to the PV (2). The identification of PF3D7_0925900 as a highly conserved gene across all species of *Plasmodium* and the protein's cellular localization to the PV (2,13) prompted us to explore the physiological function of the encoded protein. We undertook biochemical characterization of the protein of interest, and found that it can bind a [4Fe-4S] cluster and a heme moiety, but apparently not at the same time. The interesting iron-binding properties of this protein prompted us to generate a more descriptive name for the protein: the Malarial Ferrous Protein (MFP).

# RESULTS

## RECOMBINANT MFP IS A SOLUBLE, 24.7 KDA PROTEIN

In order to probe the function of this newly discovered protein *in vitro*, we recombinantly expressed the 24.7 kDa protein MFP from *Plasmodium falciparum* using *E. coli* expression strains. Initial investigation of MFP was performed on a 24 kDa $His_6$-tagged construct, termed His-MFP (**Figure III.1**). His-MFP expresses in high relative abundance in BL21 *E. coli* strains and is very soluble under the purification conditions used (**Figure III.2A**). By size exclusion chromatography (SEC), His-MFP at relatively low concentrations runs as approximately 90% monomer with the remaining His-MFP being trimeric (**Figure III.2B-D**).

## MFP BINDS AN OXYGEN-SENSITIVE IRON-SULFUR CLUSTER

During purification, a light brown color was observed; however, this color was not strong enough under the initial purification conditions used to obtain a UV-Visible spectrum to probe the identity of the chromophore. Including dithiothreitol (DTT) as a reducing agent during purification increased the amount of chromophore that co-purified with the protein, allowing for spectroscopic analysis (**Figure III.3A**). The UV-Visible spectrum of this chromophore shows a feature around 412 nm, usually indicative of a [4Fe-4S] cluster (**Figure III.3B**). The presence of an iron-sulfur cluster is further indicated by the oxygen sensitivity of the chromophore and the observed UV-Vis spectrum of a [2Fe-2S] cluster, a semi-stable oxidative intermediate (**Figure III.4**).

As histidine-tags are known to bind iron, we replaced the $His_6$-tag with a StrepII-tag for affinity purification. This new construct was termed PfMFP (**see Figure III.1**) and did not appear to drastically alter the oligomeric profile of the protein (data not shown). Additionally, we truncated 11 residues in order to remove the putatively disordered Asx-rich N-terminus, a construct we termed PfMFPΔ, which also did not drastically alter the oligomeric profile, with the primary species for both variants being the dimer at 5 mg/mL (**Figure III.5, Table III.1**). The major difference was a reduction of higher order oligomers that are likely not biologically relevant. Anaerobic reconstitution of PfMFPΔ was found to incorporate more chromophore and to alter the oligomeric profile of the sample at low protein concentrations (less than 1 mg/mL), going

from a monomeric protein at low concentrations in the apo-state to a dimeric protein at low concentrations in the cluster-bound state (**Figure III.6, Table III.1**). Metal analysis of reconstituted PfMFPΔ yielded 3.5 irons per protein, and temperature dependent electron paramagnetic resonance (EPR) spectroscopy confirmed the presence of a [4Fe-4S] cluster (**Figure III.8**). The g-value of 1.94 at 10 K and the lack of signal at 40 K are indicative of a [4Fe-4S] cluster species. As purified, His-MFP and PfMFP also show incorporation of either a [3Fe-4S] cluster or a [4Fe-4S] cluster, although to a much lesser extent. Taken together, these results suggest that each monomer of MFP binds a single [4Fe-4S] cluster. Interestingly, MFP only contains 4 cysteine residues, the amino acid most likely to bind an [Fe-S] cluster in a protein, and in the homology model of MFP all of these cysteines are positioned within 10 Å of each other, placing them in a prime location for cluster ligation (**Figure III.8**).

## MFP BINDS A HEME MOIETY

Although there were no canonical heme-binding motifs present, given the high expression of MFP across the blood stage of plasmodium, and the localization of MFP to the vacuole directly adjacent to the host red blood cell environment, we investigated whether heme could be an additional binding partner of MFP. Heme displays a characteristic red-shift (~386 nm to ~412 nm) by UV-Vis spectroscopy upon 6-coordinate binding to a protein (**Figure III.9**), although a 5-coordinate state can also be observed around 370 nm. Under aerobic conditions, PfMFPΔ with no cluster present (apo-PfMFPΔ) binds heme, indicated by the appearance of the soret band at approximately 412 nm (**Figure III.10A**). The spectra observed are indicative of a 6-coordinate iron (III) heme species (14). Under anoxic, reducing conditions, apo-PfMFPΔ also binds iron (II) heme, with an observed soret band at 425 nm upon binding (**Figure III.10B**). Ferrous heme binding was examined under anoxic conditions, due to the instability of heme in the presence of reductant under aerobic conditions (15). Both titrations indicate a 1:1 binding stoichiometry (**Figure III.10C-D**). To test the role of the oligomeric state of MFP in heme binding, we re-purified PfMFPΔ that had been incubated with heme overnight by affinity purification using the StrepII-tag. The re-purified heme-bound PfMFPΔ appears trimeric by analytical SEC (**Figure III.11, Table III.1**).

## THE PRESENCE OF A [4FE-4S] CLUSTER LIMITS HEME BINDING BY MFP

In order to determine whether the binding of these two iron-containing cofactors by MFP occur simultaneously, we attempted to perform similar heme titration assays on reconstituted PfMFPΔ under anoxic conditions; however, the appearance of a soret band at 425nm was minimal. Thus, instead of a titration, we incubated cluster-bound PfMFPΔ with a single equivalent of heme for up to 4 hours, and found only a minimal soret band at 425 nm in the absorption spectrum, even after 4 hours of incubation (**Figure III.12A**). In contrast, a strong soret band at 425 nm is visible upon incubation of apo-PfMFPΔ with one equivalent of hemin (**Figure III.12B**). Taken together, these results suggest the presence of an [4Fe-4S] cluster interferes with heme binding. Two potential methods of inhibition were investigated: inhibition due to both cofactors being ligated by the same residues and inhibition due to a change in oligomeric profile.

Upon incubation of reconstituted PfMFPΔ with approximately 50% cluster content (50% apo-MFP and 50% holo-MFP) with 0.5 equivalents of hemin we do observe an interesting species in the EPR spectrum in addition to the [4Fe-4S] cluster (**Figure III.13**). Although we are unaware of where nitric oxide was introduced into the system, this species has been identified as a nitrosyl heme species due to the triplet centered around g = 2.01 (16,17), but no change in the overall EPR spectrum of the [4Fe-4S] cluster was observed, indicating little to no interaction between the heme and the iron-sulfur cluster. We have attributed the nitrosyl-heme species as being bound to the apo-form of the protein rather than the cluster-bound form and the lack of interaction is the strongest evidence to date that heme and cluster cannot bind MFP at the same time.

## SUBSTITUTION OF CYSTEINES TO ALANINE ALTERS HEME-BINDING PROPERTIES UNDER AEROBIC CONDITIONS

To investigate the residues likely to bind the [4Fe-4S] cluster, we chose to investigate an MFP variant with alanines substituted for all four cysteines, which was termed 4CA. This protein variant does not contain the spectral fingerprint of an [Fe-S] cluster, with no absorbance at 412 nm, both as purified and after attempts at reconstitution, implying that it does not bind [Fe-S] cluster (data not shown). This protein variant still binds heme anaerobically with a similar soret

to protein ratio during titration, consistent with similar amounts of binding (**Figure III.14A and C, Table III.2**). Aerobically, however, the 4CA mutant appears to bind less heme in comparison to the PfMFPΔ construct, at least in this single experiment (**Figure III.14B and D, Table III.2**).

## INDIVIDUAL CYSTEINE VARIANTS DISPLAY ALTERED AEROBIC HEME BINDING

We also generated individual cysteine to alanine variants of MPF to investigate the effect that each individual residue might have on heme binding. A non-reducing gel of MFP variants shows a band at ~25 kDa for all four variants and an additional band around 22 kDa for three of the four variants (**Figure III.15A**). The relevance of this lower band is discussed in the next section. Three of the four variants ran as dimers by analytical SEC at 5 mg/mL (**Figure III.15B, Table III.1**). Interestingly, the C93A variant ran at a molecular weight consistent with an oligomeric state between a dimer (~50 kDa) and a trimer (~75 kDa). The C68A variant appears to bind Fe(III) heme in a similar fashion to wild-type, with a similar soret to protein ratio (**Figure III.16A, Table III.2**). The C93A variant displays apparent reduced heme similar to that of the 4CA variant (**Figure III.16B, Table III.2**).

The C70A and C216A MFP variants, however, display drastically different Fe(III) heme binding properties as compared to wild-type. Upon initial incubation of either variant with heme, a new distinct peak was observed, with a peak at 370 nm in addition to the soret, which shifts to 420 nm instead of 412 nm (**Figure III.17**). The $A_{420}/A_{280}$ ratio of both variants suggests 1:1 protein heme binding.

## MFP CONTAINS A DISULFIDE BOND BETWEEN CYS70 AND CYS216

Upon non-reducing SDS-page gel analysis, a second, smaller band was observed in addition to the band at ~25 kDa attributed to MFP. These bands run at approximately 22 and 25 kDa for His-MFP, PfMFP and variants of MFP and at approximately 18 and 22 kDa for PfMFPΔ. Neither of these bands is large enough to indicate an intermolecular interaction stable under SDS-page gel conditions, leading us to the hypothesis that the bands are a more compact (lower) and less compact (upper) state of the MFP monomer. Upon reduction of samples with β-mercaptoethanol (BME), the smaller band was no longer observed, indicating the potential of an

intramolecular disulfide bond. To verify this, samples of PfMFPΔ were incubated with dithiothreitol (DTT), potassium ferricyanide, or buffer and analyzed by non-reducing SDS-page gel electrophoresis. The samples oxidized with potassium ferricyanide contained only the lower band, the samples reduced with DTT contained only the upper band, and the as-purified samples incubated with buffer contained both bands (**Figure III.18**). Interestingly, these same two bands are present in only three of the four individual cysteine variants (**Figure III.15A**). The C216A variant eliminates the lower band, indicating that Cys216 is one of the two relevant cysteines in the disulfide bond. Our collaborators have established that Cys70 and Cys216 are the primary cysteines involved in disulfide formation (Khan and Niles, personal communication), consistent with the proposed disulfide bond in the homology model of MFP (**Figure III.8**).

Although some amount of disulfide bond is present in the wild-type MFP, we were able to isolate a variant of MFP that purifies with a substantial amount of the disulfide bond intact by substituting Cys68 and Cys93 with alanines, an MFP variant we termed C68A/C93A. This variant behaves similarly to wild-type MFP by analytical SEC, but a non-reducing SDS-page gel analysis shows only the lower band attributed to the disulfide bonded form of MFP (**Figure III.19**). We are excited to investigate the biophysical properties of this more structurally homogeneous form of MFP.

## DISCUSSION

In order to combat growing drug resistance in *Plasmodium*, a greater understanding of proteins imperative to the organism's survival should be pursued. Identifying the utility of MFP, an essential *Plasmodium* protein (2), could yield insights into the mode of infection of *Plasmodium*. Given the observed UV-Vis spectra of chromophores present in MFP that are characteristic of iron prosthetic groups, we endeavored to biochemically characterize the metal-binding properties of MFP.

In this work, we determined that MFP binds two different iron-containing prosthetic groups: a [4Fe-4S] cluster and a heme moiety. Our initial hypothesis upon discovery of the iron-sulfur cluster was that the cluster was a [2Fe-2S] cluster, given the $C_{68}$-X-$C_{70}$ motif, which has not been previously shown to bind a [4Fe-4S] cluster. Many proteins bind a [2Fe-2S] cluster along a single plane using a CXC motif, including quinoline 2-oxidoreductase from *Pseudamonas*

*putida*, CODH from *Oligotropha carboxidovorans*, and SkfB from *Bacillus subtilis* (18-20). Interestingly, within MFP the [2Fe-2S] cluster has only been observed as an oxidative degradation product (**Figure III.4**). Although it is possible that [4Fe-4S] cluster is produced erroneously upon the addition of excess iron, the fact that as purified MFP anaerobically co-purifies with either a [3Fe-4S] cluster or an intact [4Fe-4S] cluster seems to indicate the relevance of the cubane iron-sulfur cluster. However, without an assay to assess protein function, we cannot determine the relevance of any of the cluster species.

As for the heme moiety, *in vitro* both ferrous and ferric heme bind to apoMFP in an apparent 1:1 stoichiometry, with observable sorets at 425 and 412 nm, respectively. The *in vitro* incorporation of a [4Fe-4S] cluster by MFP also appears to inhibit heme binding. Interestingly, the major oligomeric states for both the heme-bound and cluster-bound forms of PfMFPΔ differ, implying some quaternary structure dependence of cofactor binding (**Figure III.20**). This shift in oligomeric profile may contribute to inhibition of heme binding when the cluster is present, but we cannot additionally rule out a shared binding site for these metallocofactors or some other structural rearrangement upon binding. Both oligomeric states can be accessed by the apo-form of the protein (**Figure III.5**), potentially indicating that the apo-form of the protein can access binding of either cofactor, but more investigation is needed to determine the malleability of the oligomeric profile of the apo-protein. Additionally, results from our studies of 4CA MFP suggest that the four conserved cysteines, which are required for binding of [4Fe-4S] cluster, are not ligands to the heme, as no change in heme coordination is apparent for 4CA MFP. Thus, the [4Fe-4S] cluster does not seem to hinder heme binding through a mechanism of ligand competition.

Interestingly, based on the homology model, there appears to be a disordered region that could be acting as a lid on the open side of the β-barrel (**Figure III.8**). It is possible that an open versus closed position of this lid could lead to the different mobility observed for MFP on a non-denaturing gel. We find that this different mobility is linked to oxidation and thus indicative of disulfide formation. Notably, Cys68 and Cys70 are positioned near the end of this putative lid region, suggesting that disulfide formation could 'lock' the lid closed, leading to a more compact state of MFP, and the observed change in mobility on a non-denaturing gel. Assuming the [4Fe-4S] cluster causes this same lid closure, the cluster could be preventing heme binding in the β-

barrel of the protein. It is worth noting that we did not look for features indicative of heme binding other than the soret for holoMFP heme binding assays, so it is possible that the cluster only precludes heme binding that causes the appearance of a soret band, a likely 6-coordinate state, but that MFP can still bind heme in a different fashion.

In order to investigate the contribution of individual cysteines to ferric heme binding, we generated cysteine to alanine variants of each individual cysteine. Three of the four variants appeared dimeric by analytical SEC. The fourth, C93A, appeared to run between a dimer and trimer by analytical SEC. This behavior has been observed previously during purification of wild-type PfMFP, which tends to display peaks between hexamer and monomer, depending on the preparation conditions. The left shifted SEC peak for C93A potentially indicates some intermolecular interactions of this variant that differ from the other three variants, and are also present in the wild-type protein. Interestingly, this variant also displays decreased iron(III) heme binding similar to the 4CA variant, potentially indicating that Cys93 is involved in iron(III) heme binding, although direct ligation is unlikely since substitution of this cysteine does not appear to change the coordination state of the heme. In order to gain structural insight into whether this residue could be contributing to heme binding, we returned to the homology model and searched the literature for lipocalin proteins binding heme or other metal cofactors. Although no characterized lipocalins have been reported to bind an [Fe-S] cluster previously, a few different proteins from this family have been identified to bind heme (21,22). To date, however, the only structures of lipocalin proteins with heme bound are those of the nitrophorin (NP) family, a protein from the insect *Rhodnius prolixus* which delivers nitric oxide (NO) for vasodilation and sequesters histamine to prevent allergic reaction during the insect's blood meal (22-27). The overlay of a structure of nitrophorin 7 (NP7, PDB 4XMH) and the homology model of MFP indicates the possibility of Cys93 being close to an iron (III) heme binding site (**Figure III.21**). A crystal structure of MFP will be important to establish where and how heme binds to MFP.

The other three individual cysteine-to-alanine MFP variants also bind heme (**Table III.2**). The C68A MFP variant behaves similarly to wild-type, indicating that Cys68 is likely not involved in heme binding, consistent with the results from our studies of heme binding to C4A MFP. Surprisingly, the C70A and C216A variants have very different binding behavior from wild-type and from 4CA MFP, with two separate heme-bound peaks observed by UV-Vis. The peak at 370

nm is generally indicative of a 5-coordinate heme species in biological systems (28,29), and we observe a strong contribution at 370 nm to the absorbance of the sample for these MFP variants. The second observed peak is at 420 nm. This peak could be attributed to many different heme species in biological systems, one of which is a nitrosyl-heme (30,31). This possibility is intriguing due to the ability of nitrophorins to bind nitric oxide and a previously observed nitrosyl heme species was visualized by EPR spectroscopy (**Figure III.13**). It is worth noting that we have not attempted EPR on MFP that was incubated with heme only. All EPR experiments to date were done on protein reconstituted with an iron-sulfur cluster. Additional EPR experiments are clearly warranted to further investigate the possibility of a nitrosyl heme species in MFP.

We also do not understand why individual C70A and C216A variants have such different spectra compared to the 4CA MFP variant in which all four cysteines are substituted. It seems that substituting certain cysteines with alanine is more disruptive to heme binding than substituting all cysteines, possibility due to formation of novel disulfide bonds when some but not all cysteines are present. For example, when Cys70 cannot form a disulfide with Cys216 due to a C216A substitution, it might form a new disulfide to Cys68 or Cys93 of the same monomer or a cysteine or a different monomer. The resulting change in structure could alter heme binding. These findings must be investigated further.

## CONCLUSION

An understanding of MFP, an essential protein found in *Plasmodium* and conserved across apicomplexa, could provide insight into the mode of infection and establish MFP as a new drug target for malaria and other parasitic diseases. Here we identify two iron-containing potential cofactors for MFP. Each of these iron prosthetic groups precludes the binding of the other, thus providing insight into the *in vitro* biochemical characteristics of this novel protein. Although the biological relevance of these cofactors has yet to be determined, these two iron-containing cofactors could be important to the *Plasmodium* during the blood stage in which it is exposed to the most external iron.

# EXPERIMENTAL PROCEDURES

## EXPRESSION OF PFMFP VARIANTS

All plasmids were obtained through collaboration with the Niles Lab at The Massachusetts Institute of Technology. NiCo21(DE3) competent cells (New England Biolabs) were transformed with *Pfmfp* in a pET28b vector and plated on an LB (Lennox) Agar plate using 50 μg/mL kanamycin as a selection marker. A single colony was used to inoculate a 5 mL overnight culture with 50 μg/mL kanamycin. The culture was grown overnight at 37 °C and 220 rpm.

One milliliter of the overnight culture was added to a 50 mL start culture with 50 μg/mL kanamycin. This start culture was grown at 37 °C and 220 rpm for approximately 3 hours.

Two liters of LB with 50 μg/mL kanamycin was inoculated with the full 50 mL start culture. This solution was split into two 2.8 L Fernbach flasks, each containing 1 L of culture. These cultures were grown at 37 °C and 220 rpm until the $OD_{600}$ reached approximately 0.6.

*Pfmpf* expression was induced by the addition of a final concentration of 400 μM isopropyl β-D-1-thiogalactopyranoside (IPTG, Gold Biotechnologies). Apo-PfMFP cultures were grown overnight at 18 °C and 220 rpm. Cluster-bound PfMFP was induced with 400 μM IPTG in the presence of 50 mg of solid ferric citrate (Sigma) per liter. Cluster-bound PfMFP cultures were grown overnight at 18 °C and 70 rpm.

Cells were harvested by centrifugation at 4,000 *g* for 30 minutes at 4 °C. The pellets were collected and flash frozen in liquid nitrogen for storage at -70 °C until purification.

## AEROBIC PURIFICATION OF HIS-MFP

Five grams of cells were resuspended in 25 mL of lysis buffer containing 50 mM HEPES, pH 8.0 and 300 mM sodium chloride (NaCl) either with or without 5 mM dithiothreitol (DTT). Forty milligrams of lysozyme (Gold Biotechnologies), one cOmplete protease cocktail tablet (Roche), and 2 μL of benzonase (Millipore) were added to the solution. Freeze-thaw lysis was performed five times to lyse cells. The cell lysate was purified by centrifugation at 25,000 *g* and 4 °C for 30 minutes. The cell lysate was diluted from 25 mL to 250 mL using lysis buffer. The diluted lysate was batch bound to an equilibrated 5 mL HisTrap HP column (GE life sciences) for 3 hours. His-MFP was eluted from the column by a step-wise gradient in the following manner: 25 mL of

lysis buffer with 20 mM imidazole (elution fraction 1), 25 mL of lysis buffer with 50 mM imidazole (elution fraction 2), 25 mL of lysis buffer with 100 mM imidazole (elution fraction 3), and 50 mL of lysis buffer with 500 mM imidazole (elution fraction 4 and 5). Each fraction was analyzed by gel electrophoresis. Fractions containing His-MFP were pooled and run on a Hi Load 26/60 Superdex$^{TM}$ 200 preparatory grade column (GE life sciences) pre-equilibrated with storage buffer (50 mM HEPES, pH 8.0 and 100 mM NaCl). Fractions with absorbance at 280 nm were analyzed by gel electrophoresis. Fractions containing pure His-MFP were pooled and concentrated to 40-100 mg/mL measured by $A_{280}$ ($\varepsilon_{280,\ His\text{-}MFP}$ = 19,160 $M^{-1}$ $cm^{-1}$, this value and all subsequent were obtained through ExPASy ProtParam (32)) using an Amicon 10 kDa molecular weight cut off centrifugal filter and stored at 4 °C in storage buffer for further use.

### Anaerobic purification of His-MFP

In addition to the aerobic purification, His-MFP was purified under anoxic conditions. In an MBraun anaerobic chamber under $N_2$ atmosphere in the presence of less than 0.1 ppm $O_2$, 5 g of cells, 40 mg of solid lysozyme (Gold Biotechnologies), 2 µL of benzonase (Millipore), and one cOmplete protease cocktail tablet (Roche) were resuspended in anaerobic lysis buffer containing 50 mM HEPES, pH 8.0, 300 mM NaCl, and 10 mM DTT. The cells were allowed to lyse for 3 hours at 4 °C and clarified by centrifugation. The oxygen-free centrifuge tubes were sealed in the anaerobic chamber before being removed for centrifugation at 25,000 g and 4 °C for 30 minutes. The sealed centrifuge bottles were brought back into the anaerobic chamber. The clarified lysate was filtered with a 22-micron sterile syringe filter and loaded onto a 5 mL HisTrap HP column (GE life sciences) that had been pre-equilibrated with lysis buffer. The solution was batch bound to the column for 3 hours. His-MFP was eluted from the column using a step-wise gradient in the following manner: 50 mL of anaerobic lysis buffer with 50 mM imidazole added (elution fractions 1 and 2) and 25 mL of anaerobic lysis buffer with 500 mM imidazole added (elution fraction 3). Fractions were analyzed by aerobic gel electrophoresis. Fractions containing His-MFP were pooled and run on a Hi Load 16/60 Superdex$^{TM}$ 200 preparatory grade column (GE life sciences) pre-equilibrated with anaerobic storage buffer (50 mM HEPES, pH 8.0, 100 mM NaCl, and 10 mM DTT). Fractions with absorbance at 280 nm were analyzed by gel electrophoresis. Fractions containing pure His-MFP were pooled and concentrated to 5-20 mg/mL measured by $A_{280}$ ($\varepsilon_{280,\ His\text{-}MFP}$ = 19,160 $M^{-1}$ $cm^{-1}$ (32)) using an Amicon 10 kDa

molecular weight cut off centrifugal filter and stored under anaerobic conditions at 4 °C in storage buffer for further use.

### AEROBIC PURIFICATION OF STREP-TAGGED MFP

Approximately 10 g of cells were resuspended in 50 mL of 50 mM HEPES, pH 8.0 and 300 mM NaCl (lysis buffer). Approximately 40 mg of lysozyme (Gold Biotechnologies), 2 μL of benzonase (Millipore), and one cOmplete protease cocktail tablet (Roche) were added to the solution. The cells were allowed to lyse at 4 °C on a rotator for 2 hours. The cell lysate was clarified by centrifugation at 25,000 *g* and 4 °C for 30 minutes. The clarified lysate was filtered with a 22-micron sterile syringe filter. The lysate was loaded onto 10 mL of Streptactin-XT Superflow resin (IBA) in a gravity column, which was then washed with 50 mL of lysis buffer. Protein was subsequently eluted in three fractions using 6, 16, and 8 mL of elution buffer (lysis buffer with 50 mM biotin added). Fractions were collected and analyzed by gel electrophoresis. Fractions containing pure MFP were pooled and concentrated to 5-30 mg/mL measured by $A_{280}$ ($\varepsilon_{280,\ PfMFP}$ = 24,660 $M^{-1}$ $cm^{-1}$ (32)) using an Amicon 10 kDa molecular weight cut off centrifugal filter and stored at 4 °C in storage buffer for further use.

### ANAEROBIC PURIFICATION OF STREP-TAGGED MFP

PfMFP was purified anaerobically in addition to the aerobic purification. In an MBraun anaerobic chamber at 4 °C under $N_2$ atmosphere, approximately 10 mg frozen cell pellet with 40 mg of solid lysozyme (Gold Biotechnologies), 2 μL of benzonase (Millipore) and one cOmplete protease cocktail tablet (Roche) were resuspended in 50 mL of 50 mM HEPES, pH 8.0, 300 mM NaCl, and 10 mM DTT (anaerobic lysis buffer). The cells were chemically lysed for 3 hours and clarified by centrifugation. The oxygen-free centrifuge tubes were sealed in the anaerobic chamber before being removed for centrifugation at 25,000 g and 4 °C for 30 minutes. The sealed centrifuge bottles were brought back into the anaerobic chamber. The clarified lysate was filtered with a 22-micron sterile syringe filter and loaded onto 10 mL of Streptactin-XT superflow resin (IBA) in a gravity column. The protein was eluted from the column as previously stated with anaerobic elution buffer (anaerobic lysis buffer with 50 mM biotin added). All fractions were collected for analysis by aerobic SDS gel electrophoresis. Fractions containing pure MFP were pooled and concentrated to 5-30 mg/mL measured by $A_{280}$ ($\varepsilon_{280,\ PfMFP}$ = 24,660

$M^{-1}$ $cm^{-1}$ (32)) using an Amicon 10 kDa molecular weight cut off centrifugal filter and stored under anaerobic conditions at 4 °C in storage buffer for further use.

## OLIGOMERIC ANALYSIS OF MFP

Purified protein in 50 mM HEPES, pH 8.0 and 100 mM NaCl was diluted to approximately 5 mg/mL protein (measured by $A_{280}$, $\varepsilon_{280,\text{ His-MFP}}$ = 19,160 $M^{-1}$ $cm^{-1}$, $\varepsilon_{280,\text{ PfMFP}}$ = 24,660 $M^{-1}$ $cm^{-1}$ (32)) and filtered through a costar Spin-X 22 μm cellulose spin filter by centrifugation. The filtered solution was loaded onto a Superdex$^{TM}$ 200 increase 10/300 GL equilibrated in protein storage buffer (50 mM HEPES, pH 8.0 and 100 mM NaCl) using a 100 μL sample loop on a BioRad FPLC instrument. The sample progress through the column was observed at 280 and 410 nm by an inline multi-wavelength detector to monitor protein absorbance and iron-sulfur cluster absorbance, respectively.

Calibration of the column was performed using five protein standards with the following molecular weights: 75 kDa, 43 kDa, 29 kDa, 13.7 kDa, and 6.5 kDa. The retention volume was observed for each protein by monitoring $A_{280}$ and the following linear calibration equation for the column was obtained:

$$[\text{retention time}] = -0.6305\ [\log (\text{MW})] + 4.7482$$

where [retention time] is the retention time through the column in minutes and [log (MW)] is the $\log_{10}$ of the protein molecular weight in daltons.

## RECONSTITUTION OF A [4FE-4S] CLUSTER IN PFMFP AND METAL ANALYSIS

In an MBraun anaerobic chamber at 4 °C under $N_2$ atmosphere, PfMFP in 50 mM HEPES, pH 8.0, 300 mM NaCl, and 10 mM DTT was incubated with 2 molar equivalents of $FeCl_3$ for approximately 5 minutes before the addition of 2 molar equivalents of $Na_2S$. The solution was incubated overnight, approximately 16 hours, and was desalted to remove excess iron using a HiPrep 26/10 Sephadex G-25 desalting column (GE). UV-Vis spectra were collected using a Thermo Fisher NanoDrop$^{TM}$ 2000c. A small aliquot of protein was removed from the anaerobic chamber for iron analysis. Briefly, 100 μL of ~80 μM reconstituted protein was added to 125 μL of a 0.4% (w/v) ascorbate solution, 7.5 μL concentrated hydrochloric acid, and 12.5 μL of 5 mg/mL ferene. This solution was vortexed for 30 second, after which 500 μL of 8 M guanidinium chloride and 100 μL of 10 M ammonium acetate were added. The solution was

incubated for at least 30 minutes, after which the absorbance was read and compared to the extinction coefficient for iron bound ferene of 34,500 $M^{-1}cm^{-1}$ at 593 nm. The calculated concentrations were increased by 8.45-fold to account for dilution during the assay preparation. The concentration of iron was compared to the concentration of protein calculated by a standard Bradford assay with no correction coefficient.

## ELECTRON PARAMAGNETIC RESONANCE SPECTROSCOPY

Anaerobic, reconstituted PfMFPΔ was concentrated to ~200 μM. Samples incubated with heme contained ~100 μM [4Fe-4S] cluster by $A_{412}$ ($\varepsilon \sim 15{,}000\ M^{-1}cm^{-1}$) and were incubated with a final concentration 100 μM hemin dissolved in 0.5 M NaOH. Samples were either reduced for EPR spectroscopy using 10 mM sodium dithionite (Riedel-de Haën) or frozen with no additional reductant. Samples were flash frozen in liquid nitrogen inside of a Coy anaerobic chamber under ~2.5% hydrogen, 97.5% argon atmosphere and measured by X-band (~9.4 GHz) EPR spectroscopy in the Department of Chemistry Instrumentation Facility at The Massachusetts Institute of Technology. Spectra were collected using a Bruker EMX spectrometer equipped with an ER 4119HS high sensitivity resonator, a Gunn diode microwave source, and a Bruker/CoolEdge 4K waveguide cryogen-free cryostat. Spectra were measured at 9.37 GHz microwave frequency, 200 μW power, 8 G modulation amplitude, and the temperatures indicated.

## HEME BINDING ASSAYS

To investigate the stoichiometry of heme binding for all PfMFP constructs, heme was titrated against apo-PfMFP, both aerobically and anaerobically. To obtain anaerobic apo-PfMFPΔ, 250 mM EDTA was added to a 100 μM sample of 50 μM PfMFPΔ in the MBraun anaerobic chamber ($O_2$ = 0.1 - 0.6 ppm) and incubated until no [Fe-S] cluster peak was observed by UV-Vis spectroscopy, approximately 5 days. The protein was desalted on a PD-10 desalting column (Amersham Biosciences) and protein concentration was measured by the absorbance at 280 nm ($\varepsilon_{280,\ PfMFP}$ = 24,410 $M^{-1}\ cm^{-1}$ (32)). For the aerobic samples, an initial spectrum of 5 or 10 μM apo-protein was taken using a 1 cm path length cuvette on a Thermo Fisher NanoDrop 2000c. Subsequently, 1 mL of 5 or 10 μM apo-protein was spiked with 1 or 2 μL of 0.59 mM hemin (Sigma) in 0.5 M NaOH, respectively, and allowed to incubate for 5 minutes. At this time, a UV-

Visible spectrum was taken of the solution blanked against protein buffer. The process of spiking the solution with hemin was continued until 17 or 34 μL of hemin, approximately 2 equivalents, had been added to the 5 or 10 μM protein sample, respectively. A heme standard dissolved in protein storage buffer was run in the same manner in parallel and subtracted to remove the contribution of free-hemin in solution from the UV-Vis spectrum. The anaerobic titration was performed in the same manner with the following change: the experiment was conducted in an MBraun anaerobic chamber under a nitrogen atmosphere with 10 mM DTT as an additional component in the buffer. Protein used was prepared anaerobically as described above, and the hemin solution was allowed to gas exchange in the anaerobic chamber for 24 hours before the experiment. In order to compare the heme titrations with the heme-binding time courses, the absorbance of free-heme was not removed from the solution.

To investigate the time dependence of apo-PfMFPΔ heme binding, 10 μM of anaerobically purified apo-PfMFPΔ or 5 μM of reconstituted-PfMFPΔ was incubated with 10 μM or 5 μM hemin, respectively, in 50 mM HEPES, pH 8.0, 100 mM NaCl, and 10 mM DTT in an MBraun anaerobic chamber. UV-Vis spectra blanked against 50 mM HEPES, pH 8.0, 100 mM NaCl and 10 mM DTT were collected using a cuvette on a Thermo Fisher NanoDrop 2000c at 5 minutes, 10 minutes, 30 minutes, 1 hour, 2 hours, and 4 hours.

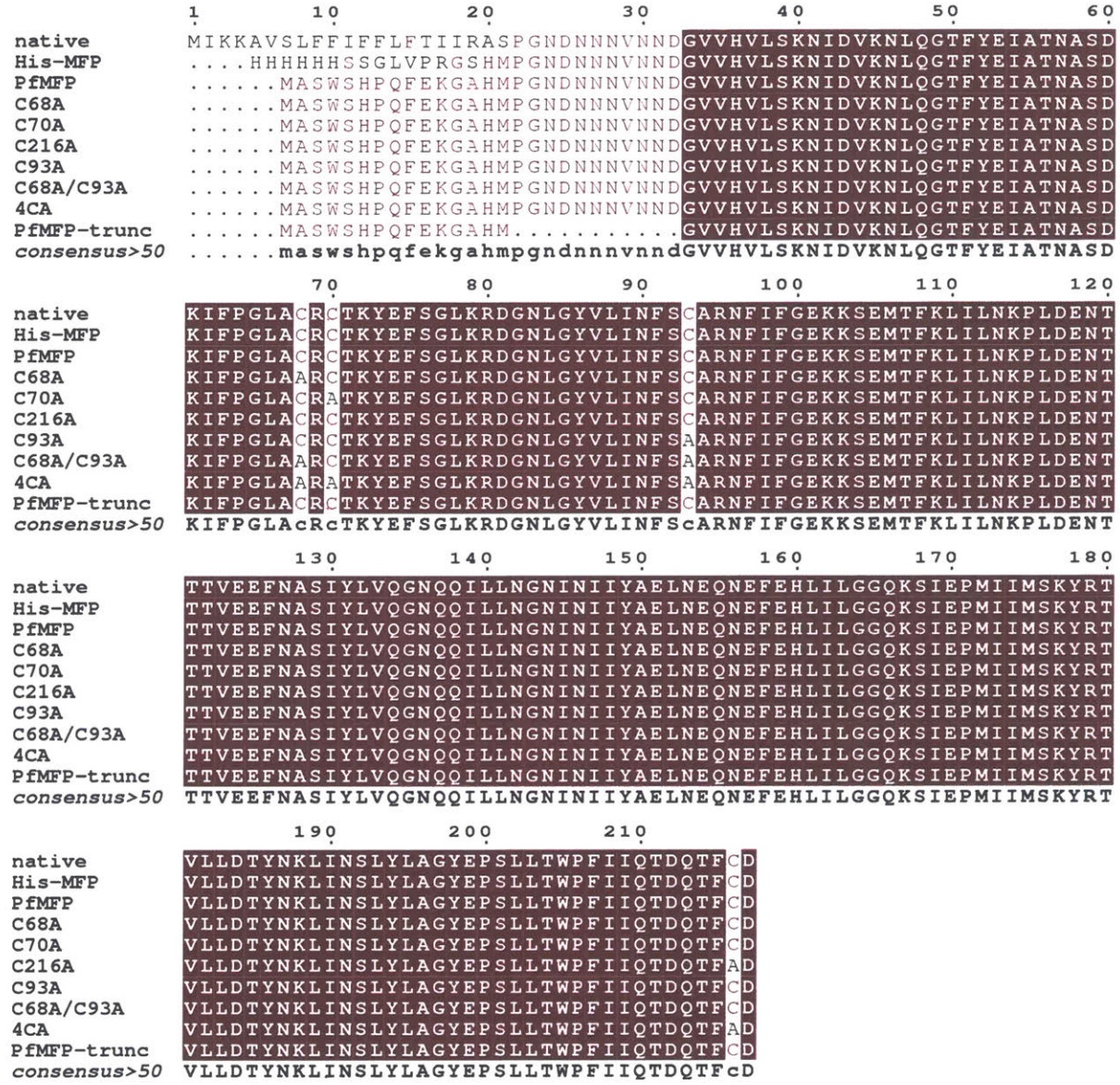



**Figure III.1. Sequence alignment of MFP variants used in this work.** The putative signal peptide was removed for all expression constructs. **Native** refers to the protein expressed in *Plasmodium falciparum*; **His-MFP** has no signal peptide and a $His_6$-affinity purification tag; **PfMFP** has no signal peptide and a StrepII-affinity purification tag; **C68A, C70A, C93A,** and **C216A** have no signal peptide, have a StrepII-affinity purification tag, and contain a single substitution from cysteine to alanine at the given residue number; **4CA** has no signal peptide, a StrepII-affinity purification tag, and all four cysteines are mutated to alanine; and **PfMFP-trunc**, also written as PfMFPΔ in the text of this chapter, has no signal peptide, a StrepII-affinity purification tag, and is missing the disordered N-terminal tail. All additional variants were based on PfMFP. Conserved residues are highlighted in red. Similar residues are in red text. Figure generated with MultAlin (33) and ESpript (34).

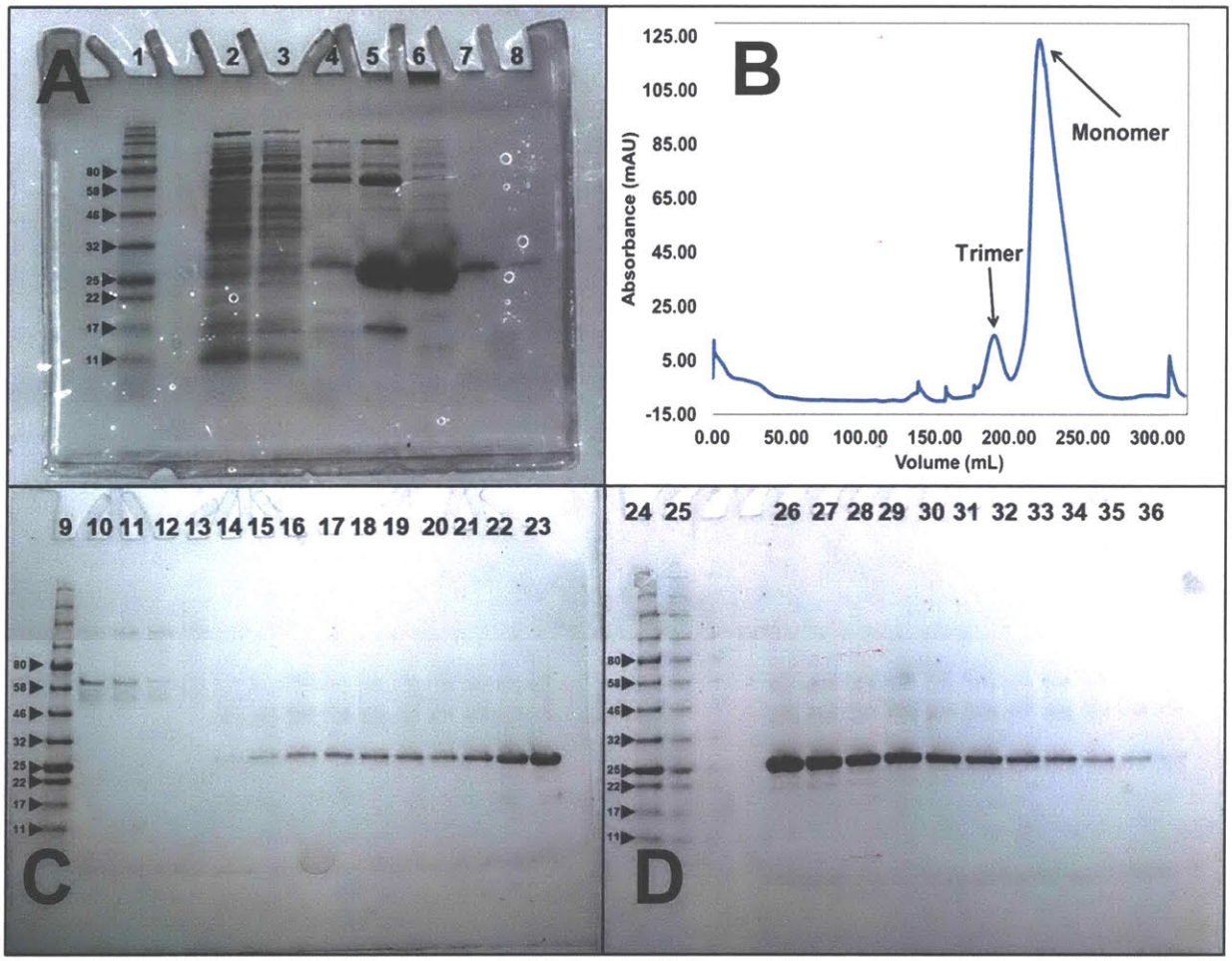



**Figure III.2. Recombinantly expressed His-MFP from *Plasmodium falciparum* is soluble.** **(A)** The purification of the 24 kDa His-MFP yields a large amount of relatively pure protein. Lanes 1-8 are as follows: NEB pre-stained ladder, column flow through, column wash, and elution fractions 1-5. **(B)** The SEC trace of His-MFP after initial affinity purification indicates a mostly monomeric protein under the purification conditions tested. **(C & D)** The peaks labeled monomer and trimer are relatively pure samples of His-MFP. Lanes 9, 24, and 25 correspond to NEB pre-stained ladder, lanes 10-13 correspond to the peak around 140 mL, and lanes 14-23 and 26-36 correspond to the trimer and monomer peak of His-MFP.

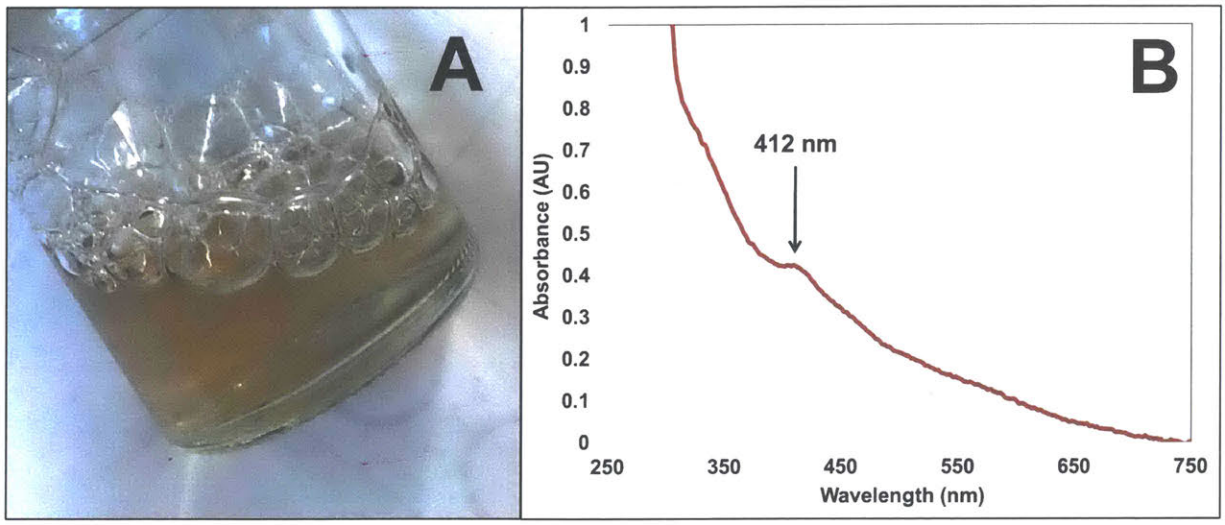



**Figure III.3. Aerobically purified His-MFP is brown when purified with reducing agent.** The unknown chromophore appears to be an [Fe-S] cluster. **(A)** MFP appears reddish-brown above approximately 50 mg/mL when purified aerobically in the presence of DTT. **(B)** The UV-Visible absorption spectrum potentially indicates the presence of a [4Fe-4S] cluster.

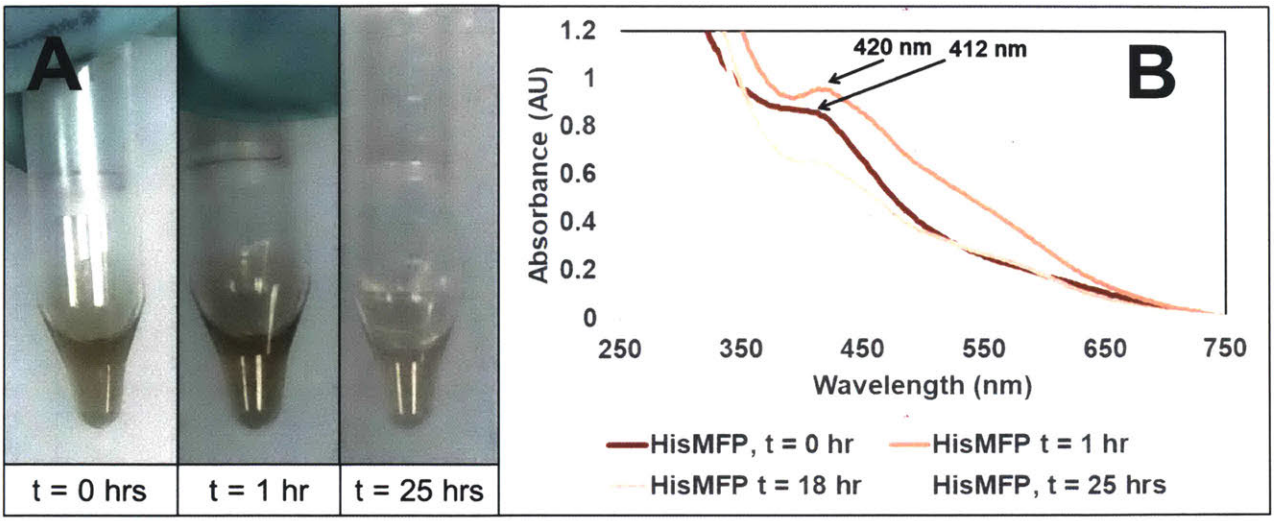



**Figure III.4. MFP binds an iron-sulfur cluster.** The brown chromophore was oxygen sensitive, observed both **(A)** visually and **(B)** spectroscopically. The shift in the UV-Vis spectra observed between t = 0 hr and t = 1 hr is consistent with a [4Fe-4S] cluster degrading to a [2Fe-2S] cluster. The t = 18 hrs and t = 25 hrs spectra indicate a loss of the [2Fe-2S] cluster, consistent with oxygen sensitivity.

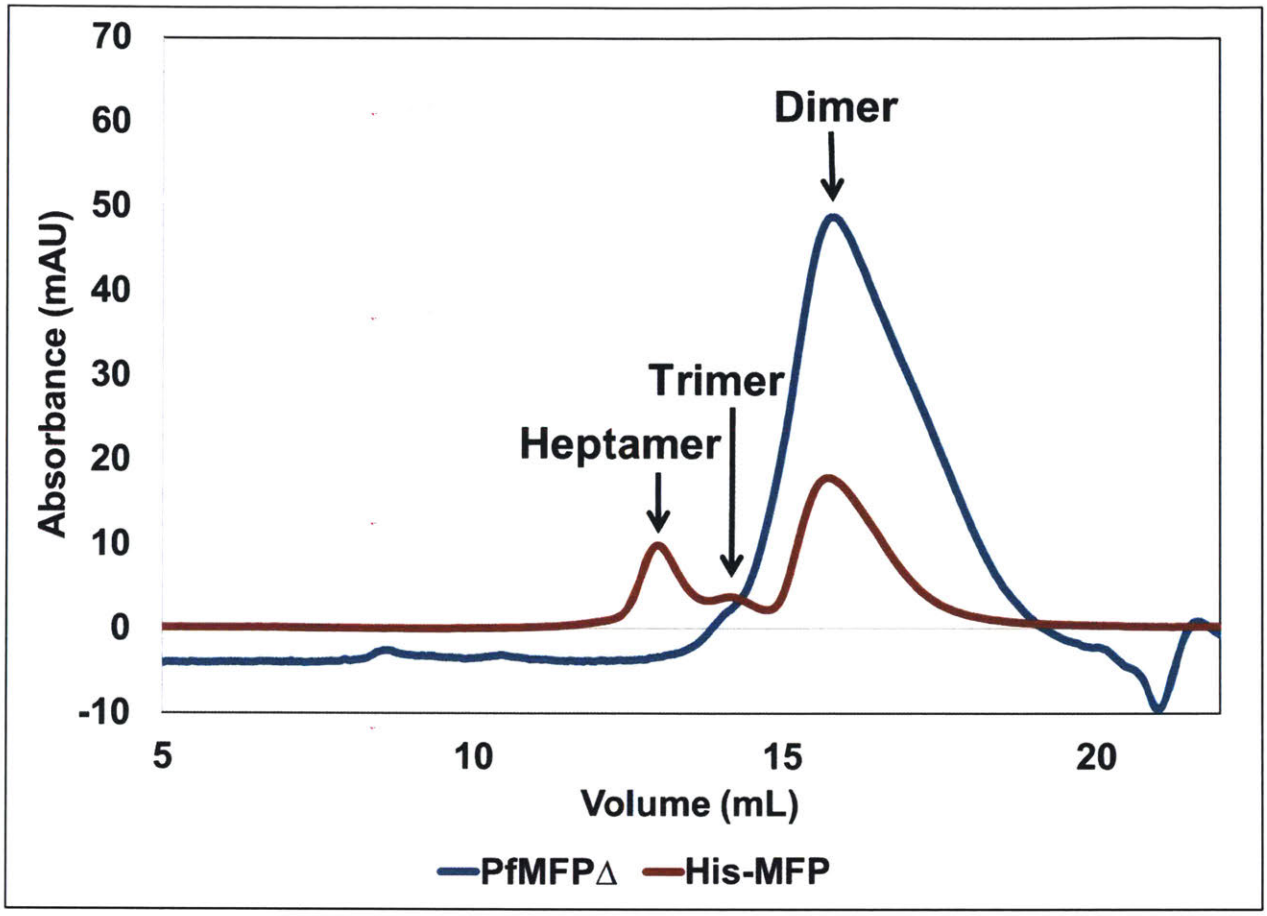



**Figure III.5. Comparison of Analytical SEC traces of His-MFP and PfMFPΔ.** The His-MFP and PfMFPΔ variants of MFP run similarly on a sizing column at 5 mg/mL. The His-MFP contains more higher order oligomers than the PfMFPΔ, indicating that the truncated construct might be more useful in biochemical and biophysical assays.

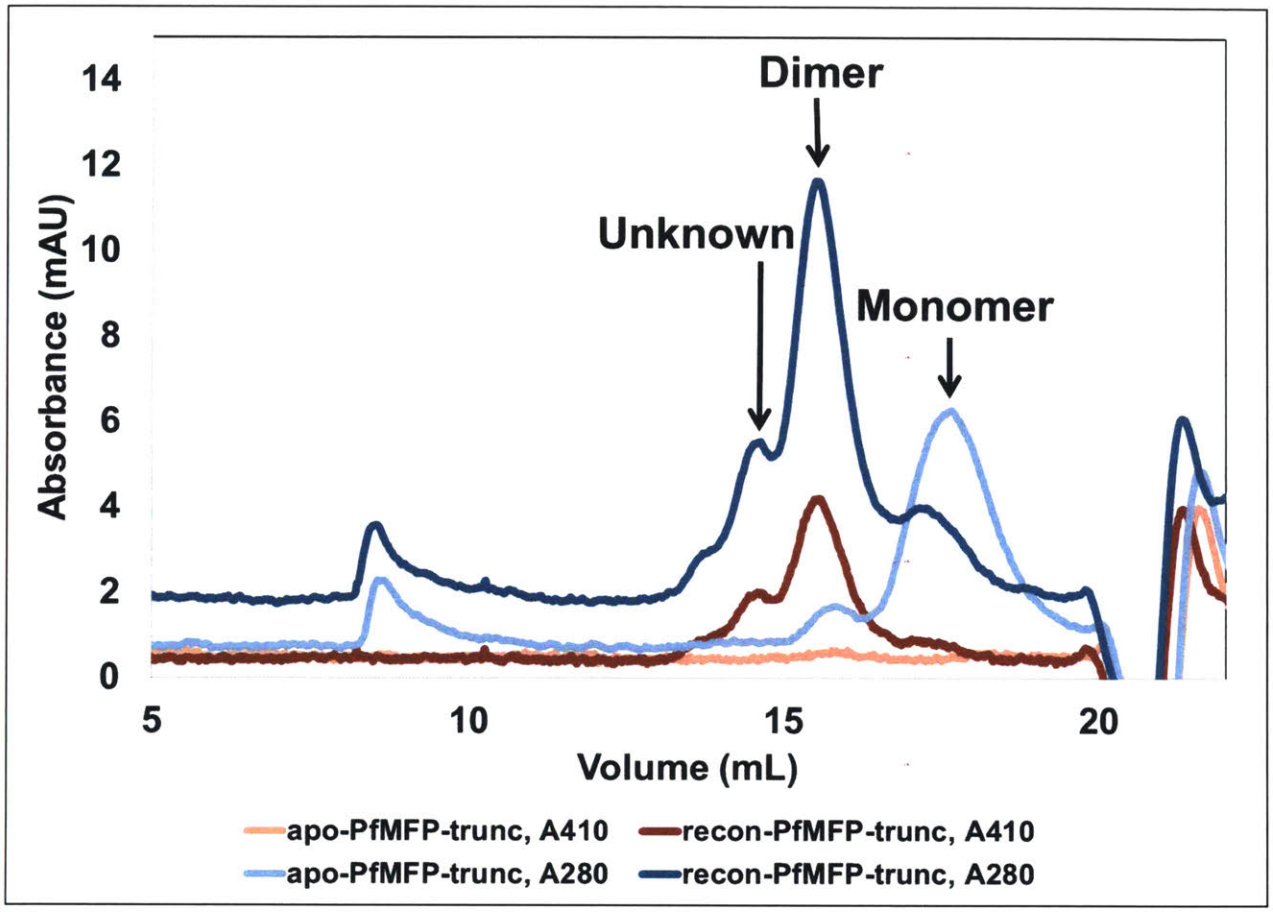



**Figure III.6. Cluster incorporation appears to stabilize the dimer state.** At concentrations of less than 1 mg/mL of MFP, apo-PfMFPΔ appears mostly monomeric by SEC, but reconstituted PfMFPΔ appears mostly dimeric. The absorbance a $A_{280}$ and $A_{410}$ are shown in blue and red, respectively, for the as purified, apo sample and the reconstituted sample in light and dark colors, respectively. There is an unknown peak that appears upon reconstitution, labeled above, that does not appear to correspond with a molecular weight of either the dimer or a trimeric species, running at an apparent molecular weight of 64.4 kDa, as opposed to 45.5 kDa for the dimer and 68.3 kDa for the trimer.

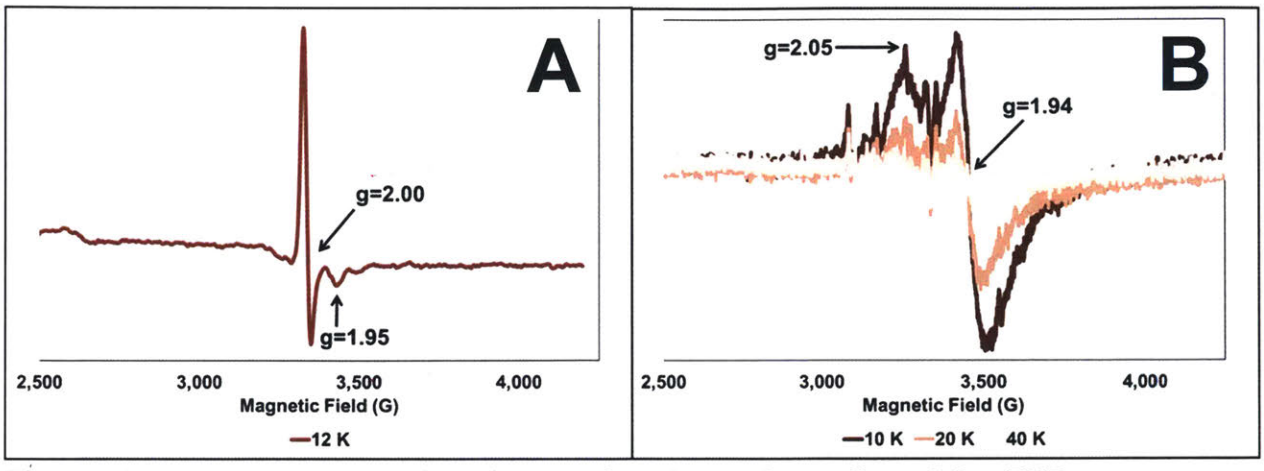



**Figure III.7. The presence of an iron-sulfur cluster is confirmed by EPR spectroscopy.** Clear EPR spectral finger prints for both a [3Fe-4S] cluster and a [4Fe-4S] cluster are observed. **(A)** The EPR spectrum of as purified His-MFP with no additional reducing agent shows a clear signal for a [3Fe-4S cluster]. **(B)** In the presence of sodium dithionite, the EPR spectrum of reconstituted PfMFPΔ indicates the presence of a [4Fe-4S] cluster.

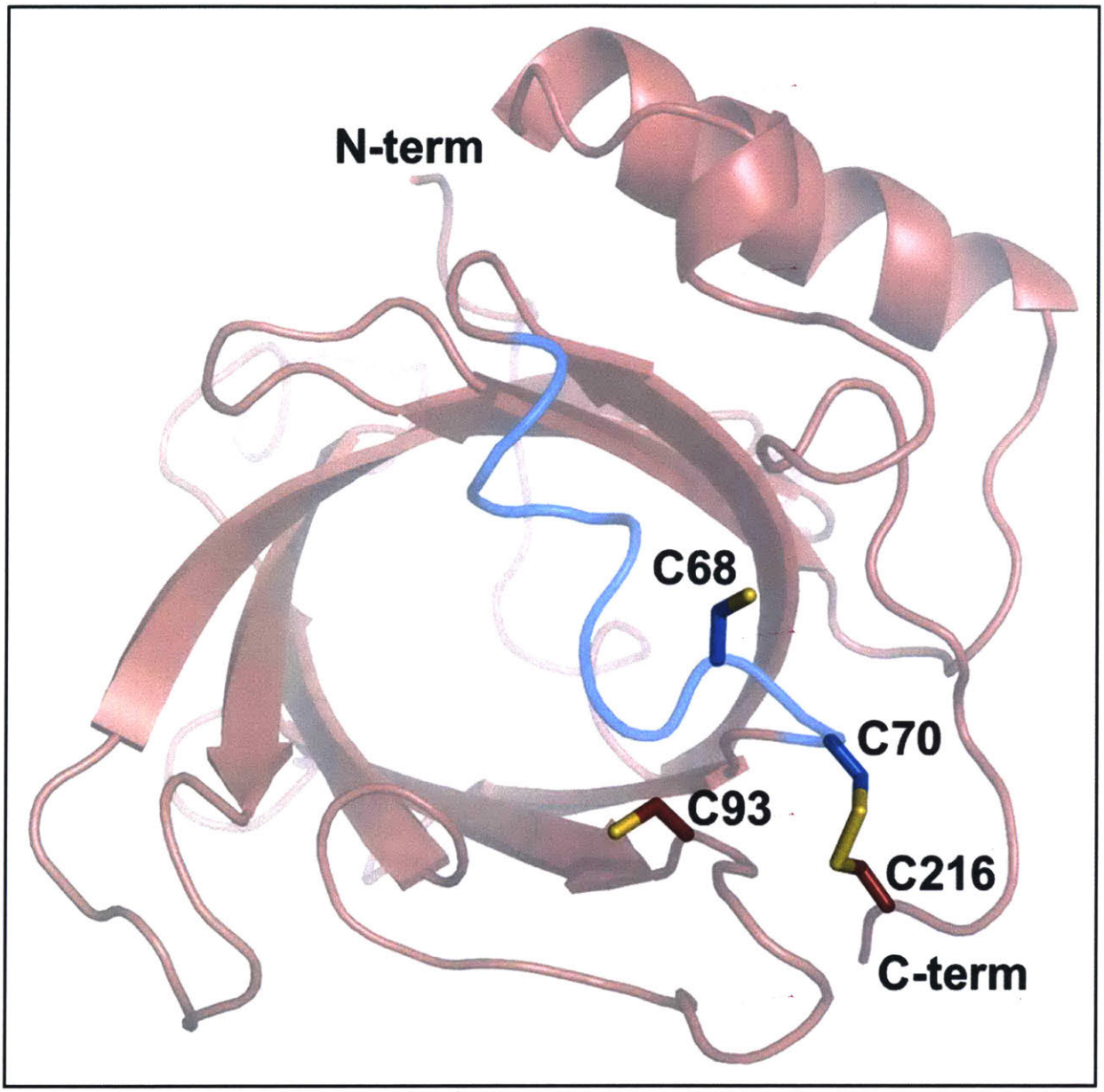



**Figure III.8. The homology model of MFP supports the biochemically-determined iron-binding properties.** All four cysteines are less than 12.0 Å away from each other and perfectly poised to bind an [Fe-S] cluster. The putative "lid" region is highlighted in blue. Cysteine residues are shown in stick.

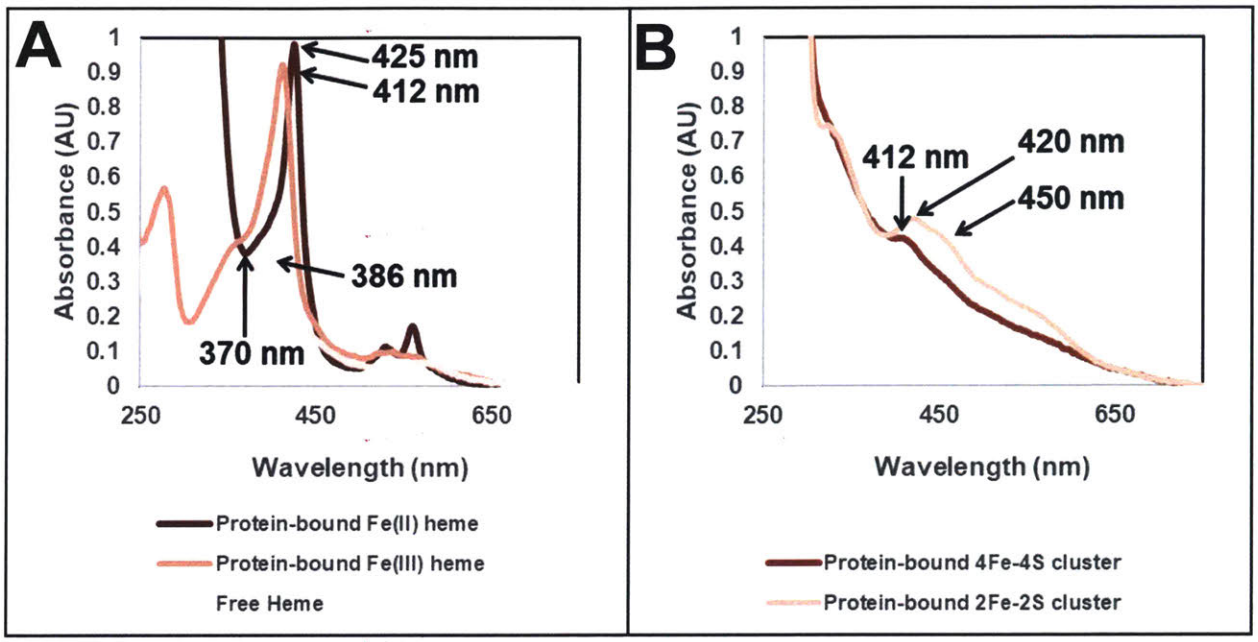



**Figure III.9. Heme shows a characteristic red-shift by UV-Vis upon binding to protein.** Sample spectra to show heme and cluster binding were generated using MFP as the protein and sample buffer for the appropriate condition. **(A)** Upon six-coordinate binding to the protein, the heme absorbance shifts toward a sharp peak between 400 and 450 nm instead of the broad peak observed around 386 nm in the buffer conditions used. **(B)** In contrast, the absorbance of an [Fe-S] cluster is closer to a shoulder than a peak, with a local maximum value of ~412 nm for a [4Fe-4S] cluster and a local maximum of ~420 nm for a [2Fe-2S] cluster. The soret absorbance is much stronger than that of an [Fe-S] cluster absorbance, with molar absoptivities close to 150,000 $M^{-1}cm^{-1}$ for heme and close to 15,000 $M^{-1}cm^{-1}$ for a typical [4Fe-4S] cluster.

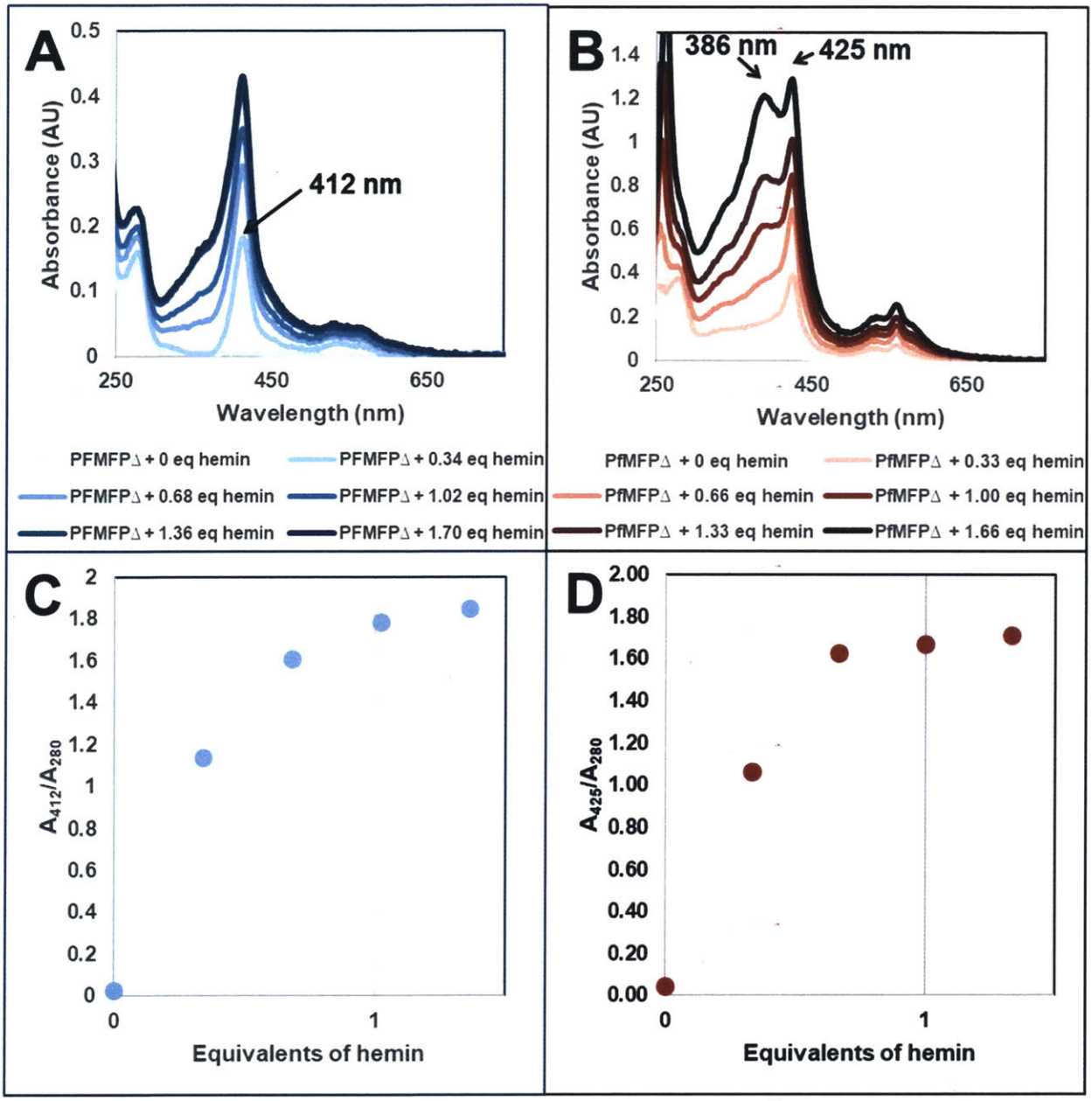



**Figure III.10. PfMFPΔ binds hemin under both aerobic and anoxic conditions.** Upon titration with hemin, apo-PfMFPΔ binds hemin both **(A)** aerobically and **(B)** anaerobically. A strong soret band is observed for both samples. Based on the soret to protein ratio, apo-PfMFPΔ appears to bind a single hemin per monomer both **(C)** aerobically and **(D)** anaerobically. Plots represent a single experiment. Protein concentration in (A) is 5 μM and in (B) is 10 μM.

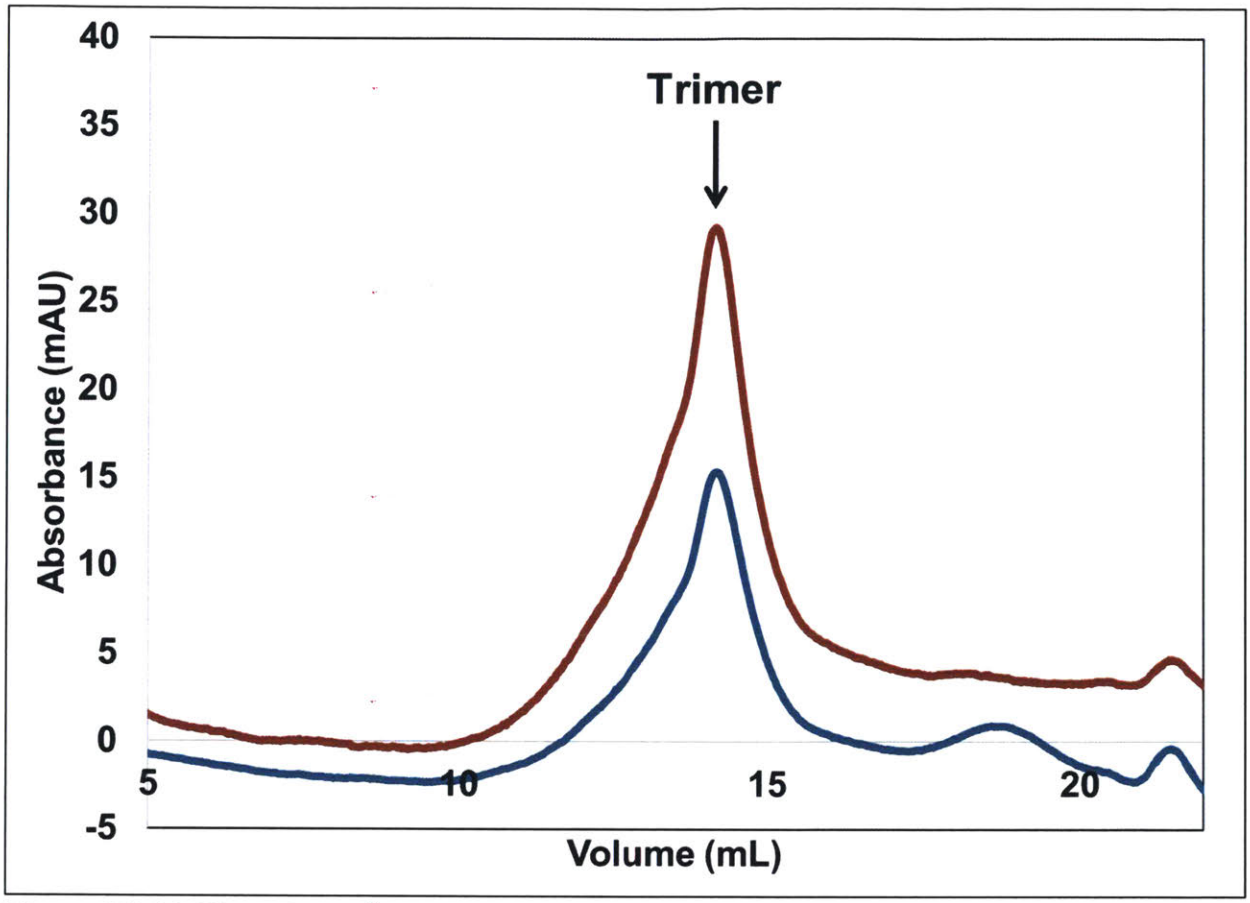



**Figure III.11. Heme bound PfMFPΔ runs as a trimer by SEC.** Blue and red traces correspond to the absorbance at 280 nm (protein) and absorbance at 410 nm (heme soret) of the sample, respectively.

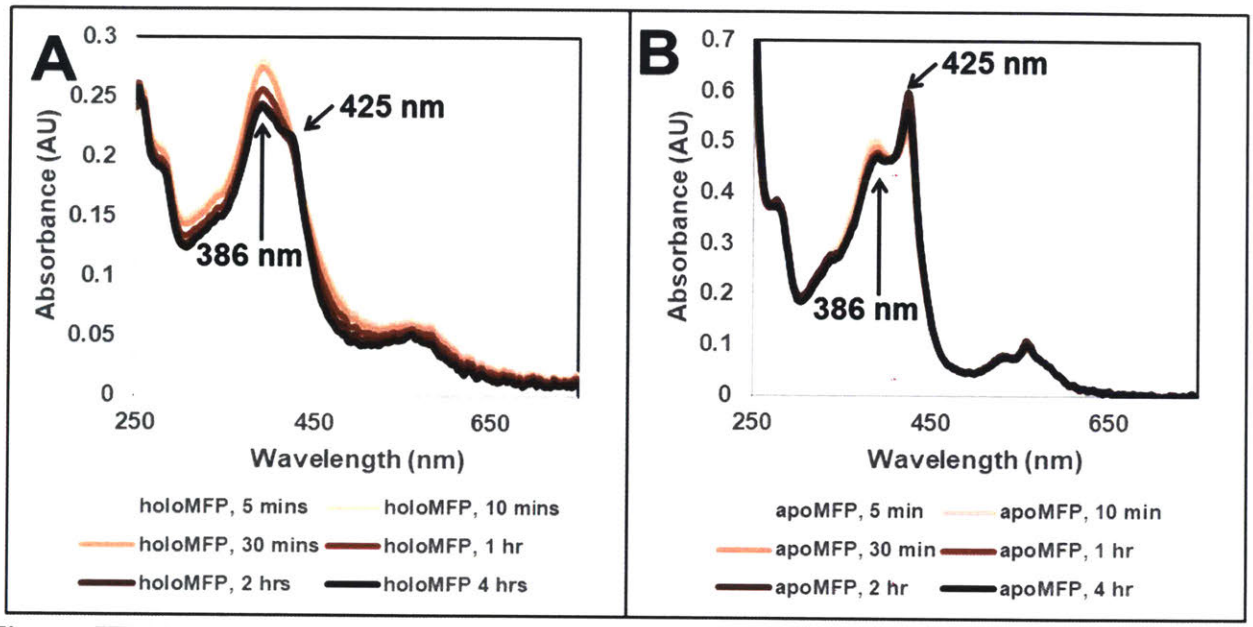



**Figure III.12. Heme binding is reduced when PfMFPΔ is reconstituted with a [4Fe-4S] cluster under anoxic conditions. (A)** Reconstituted PfMFPΔ binds very little heme, even after 4 hrs, based on the minimal appearance of a soret at 425 nm, indicating bound heme, and the larger peak at 386 nm indicating the presence of free heme. It is possible that it is the fraction of apo-protein left after reconstitution that is binding heme and that the presence of the cluster blocks heme binding. **(B)** apo-PfMFPΔ binds heme as observed by the peak at 425 nm. The soret peaks at 425 nm are labeled in both (A) and (B). The $A_{425}/A_{280}$ ratios are 1.09 and 1.58 for (A) and (B), respectively. The $A_{425}/A_{280}$ ratio for the titration of apo-PfMFPΔ under anoxic conditions (Figure III.10B) is 1.67. The reduction in absorbance over the time course in (A) is likely due to a lack of mixing over the 4 hour experiment. Protein concentrations were (A) 5 μM and (B) 10 μM. Both were incubated with a single equivalent of hemin.

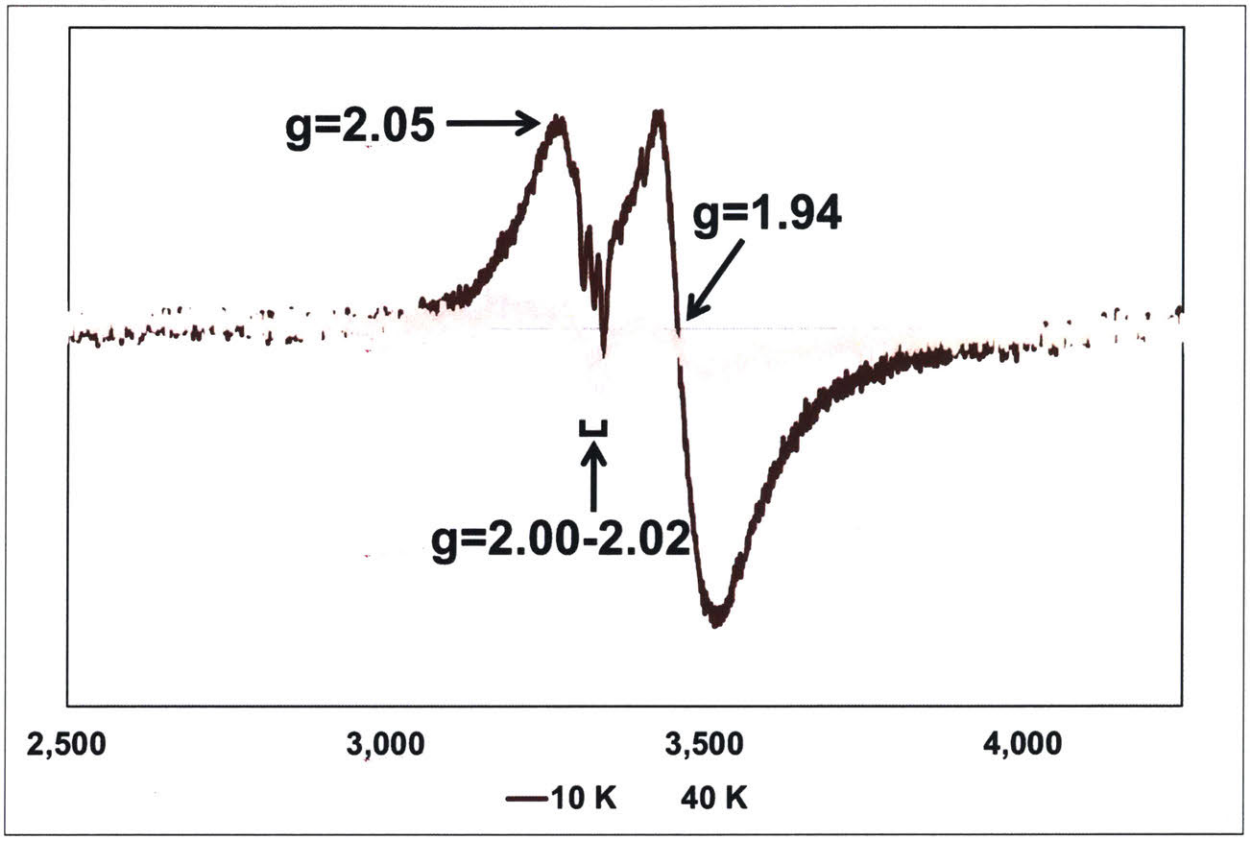



**Figure III.13. A small peak for nitrosyl heme is observed when 50% reconstituted PfMFPΔ is incubated with 0.5 equivalents of heme.** The triplet centered around g=2.01 is indicative of a NO-bound heme species (16,17).

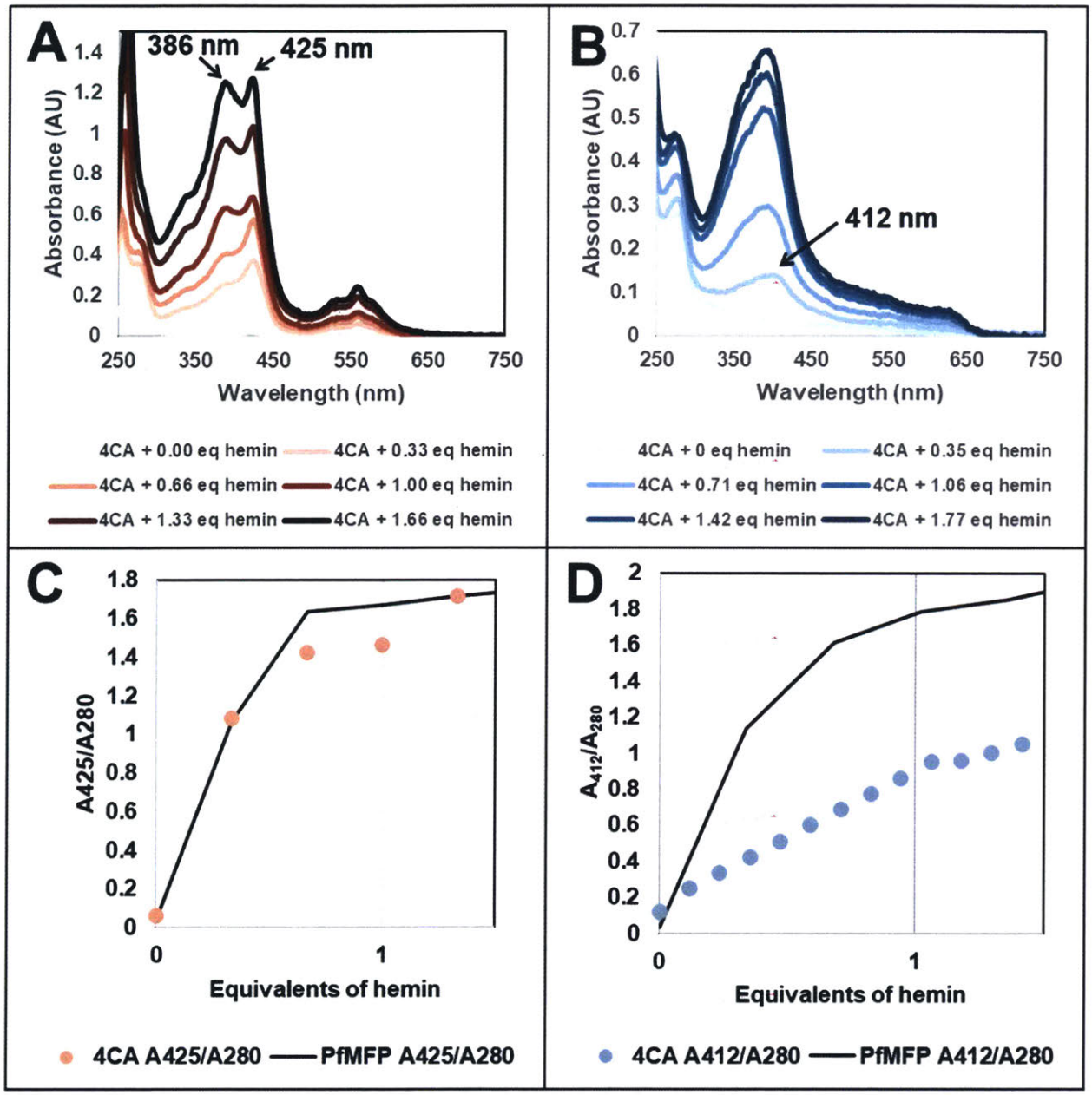



**Figure III.14. The 4CA MFP variant shows similar binding of iron(II) heme as does PfMFPΔ, but a decrease in binding for iron(III) heme. (A)** To determine a cysteine residue ligates heme, the 4CA MFP variant was titrated with heme. Under anaerobic conditions, heme binds to 4CA-MFP similarly to wild-type protein. **(B)** Under aerobic conditions, though, the appearance of a soret band is minimal. The **(C)** $A_{425}/A_{280}$ or **(D)** $A_{412}/A_{280}$ of the 4CA heme binding titration (circles) compared to the same ratio for wild-type (solid black line) indicate a similar binding of iron (II) heme and a reduced binding of iron (III) heme. Plots represent a single experiment. Both titrations were performed with 10 μM protein.

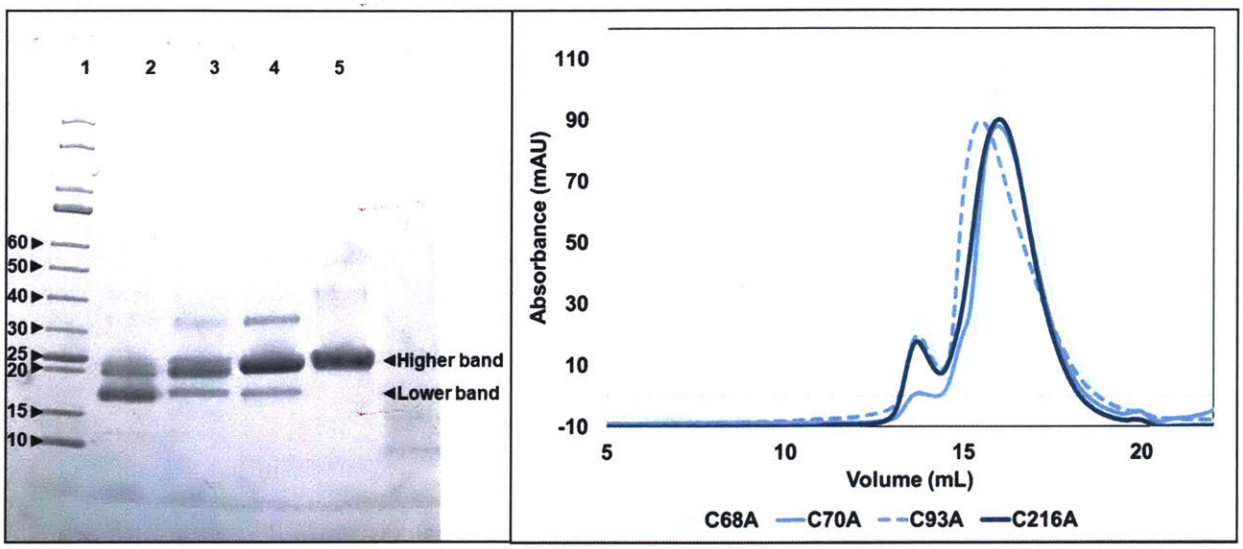



**Figure III.15. The individual cysteine variants exhibit similar oligomeric states as wild-type. (A)** A higher and a lower band can be seen in three of the four variants by non-reducing SDS-page gel. All samples were 0.5 mg/mL. Lane 1 is the protein ladder, lane 2 is the C68A MFP variant, lane 3 is the C70A MFP variant, lane 4 is the C93A MFP variant, and lane 5 is the C216A variant. C216A is the only variant that completely ablates the lower band. **(B)** Three cysteine to alanine variants run as dimers by analytical SEC when aerobically purified. The C93A variant runs between a dimer (50 kDa) and a trimer (75 kDa).

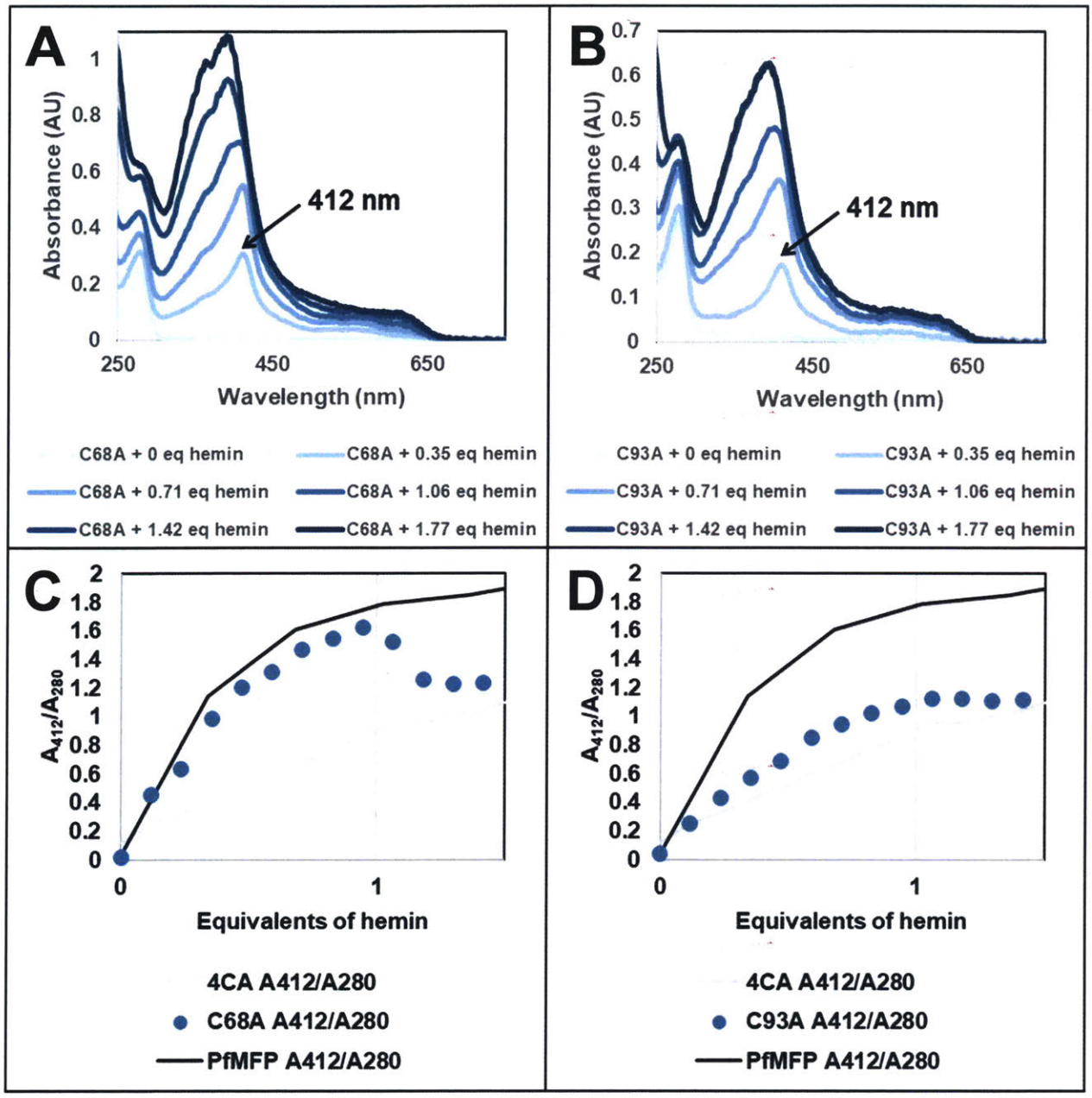



**Figure III.16. Under aerobic conditions, Cys68 and Cys93 MFP variants bind heme similarly to wild-type and 4CA MFP, respectively.** The observed binding for **(A)** C68A and **(B)** C93A are similar to when either all or none of the cysteines are present, respectively. This binding can be observed in the $A_{412}/A_{280}$ ratio for **(C)** C68A and **(D)** C93A (blue dots) as compared with the same values for PfMFPΔ (solid black line) and 4CA (solid gray line). Both titrations were performed with 10 μM protein.

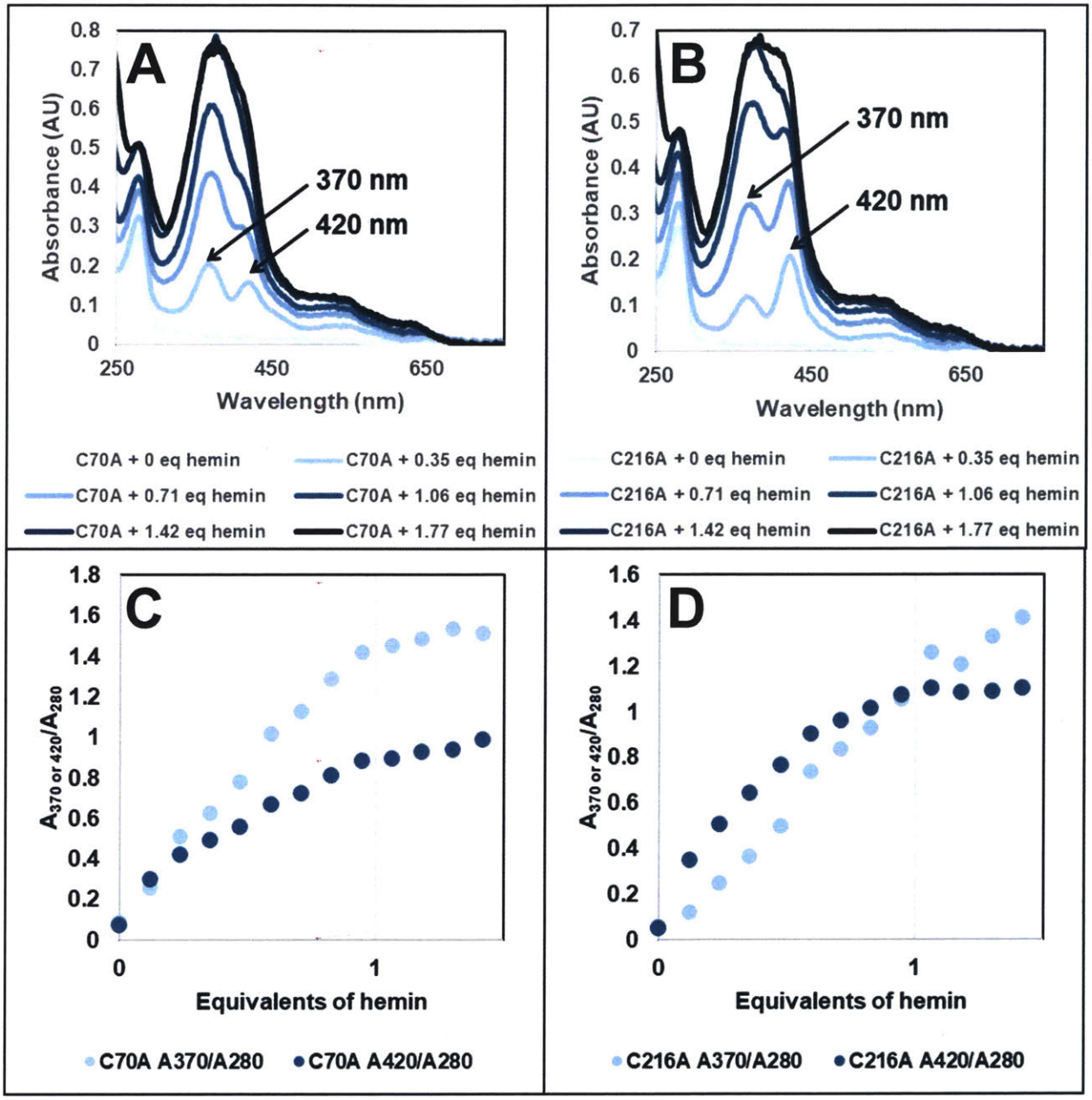



**Figure III.17. Both the C70A and C216A variants display altered heme binding when compared to wild-type under aerobic conditions.** The soret for both **(A)** C70A and **(B)** C216A is shifted to 420 nm and there appears to be an additional binding event present for both C70A and C216A with a strong absorbance at 370 nm. **(C-D)** The $A_{420}/A_{280}$ ratio (dark blue dots) of both variants suggest a 1:1 protein to heme ratio. The $A_{370}/A_{280}$ ratio of C70A also suggests a 1:1 protein to heme ratio, however the increasing $A_{370}/A_{280}$ ratio of of C216A suggests continued binding, even after a single equivalent of hemin has been added. Both titrations were performed with 10 μM protein.

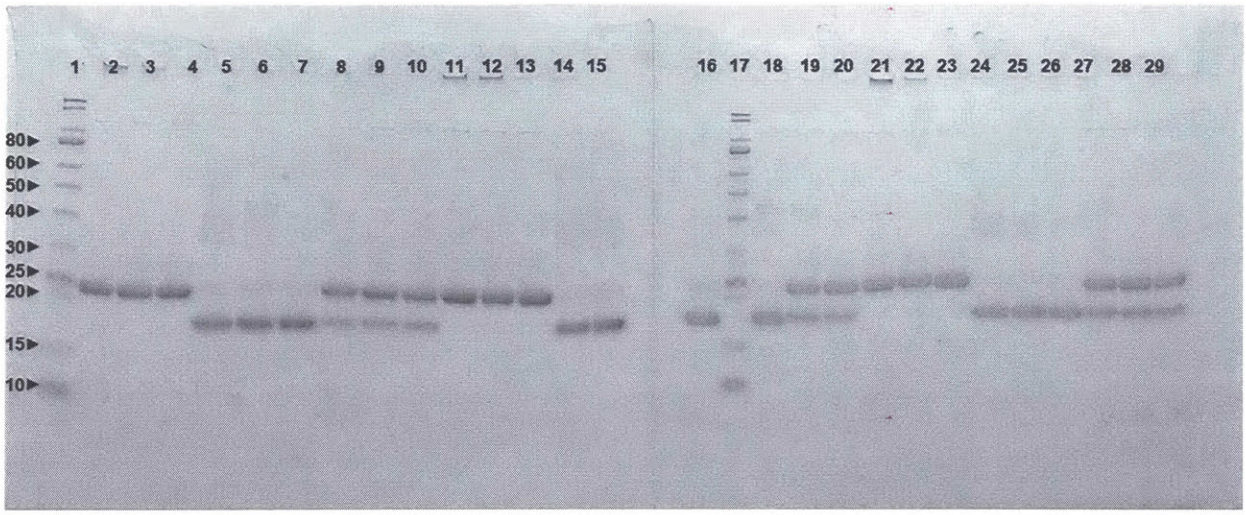



**Figure III.18. Oxidized MFP runs lower on a non-reducing SDS-page gel.** The lower band is likely due to a disulfide bond between two cysteine residues. Each group of three bands corresponds to 0.5 mg/mL PfMFPΔ incubated with 10 mM DTT, 10 mM potassium ferricyanide, or buffer for 1 hour, 2 hours, and 4 hours (left to right) at different temperatures. Lanes 2-4 were incubated with DTT at 4 °C, lanes 5-7 were incubated with potassium ferricyanide at 4 °C, lanes 8-10 were incubated with buffer at 4 °C, lanes 11-13 were incubated with DTT at room temperature, lanes 14-16 were incubated with potassium ferricyanide at room temperature, lanes 18-20 were incubated with buffer at room temperature, lanes 21-23 were incubated with DTT at 37 °C, lanes 24-26 were incubated with potassium ferricyanide at 37 °C and lanes 27-29 were incubated with buffer at 37 °C. Lane 1 and 17 are NEB unstained broad range protein ladder.

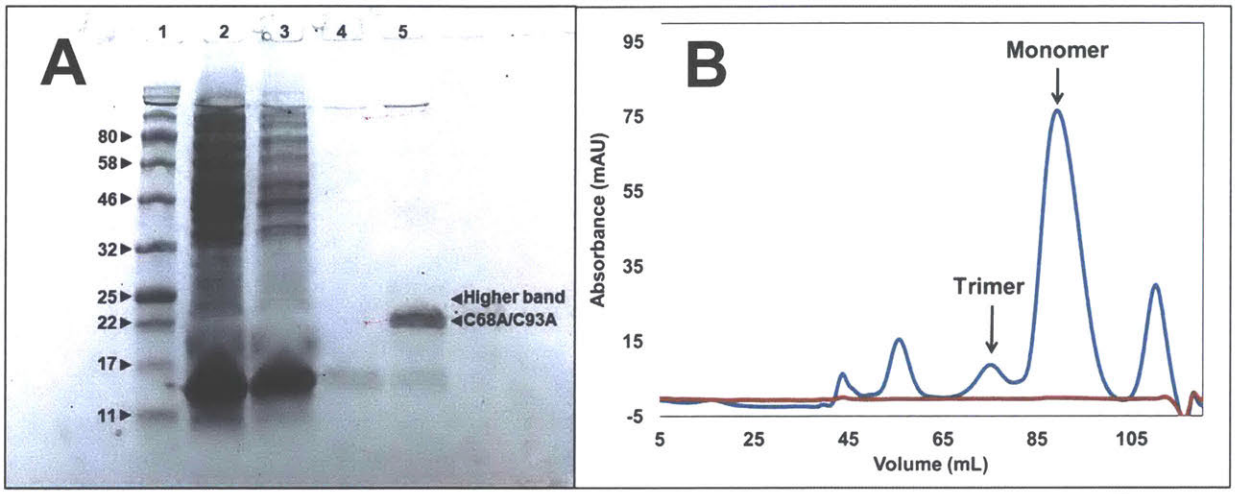



**Figure III.19. Substituting both Cys68 and Cys93 with alanine stabilizes the disulfide bond.** Initial purifications of PfMFP C68A/C93A show **(A)** a single, lower band by non-reducing SDS-page gel electrophoresis and **(B)** a relatively homogenous sample profile by SEC.

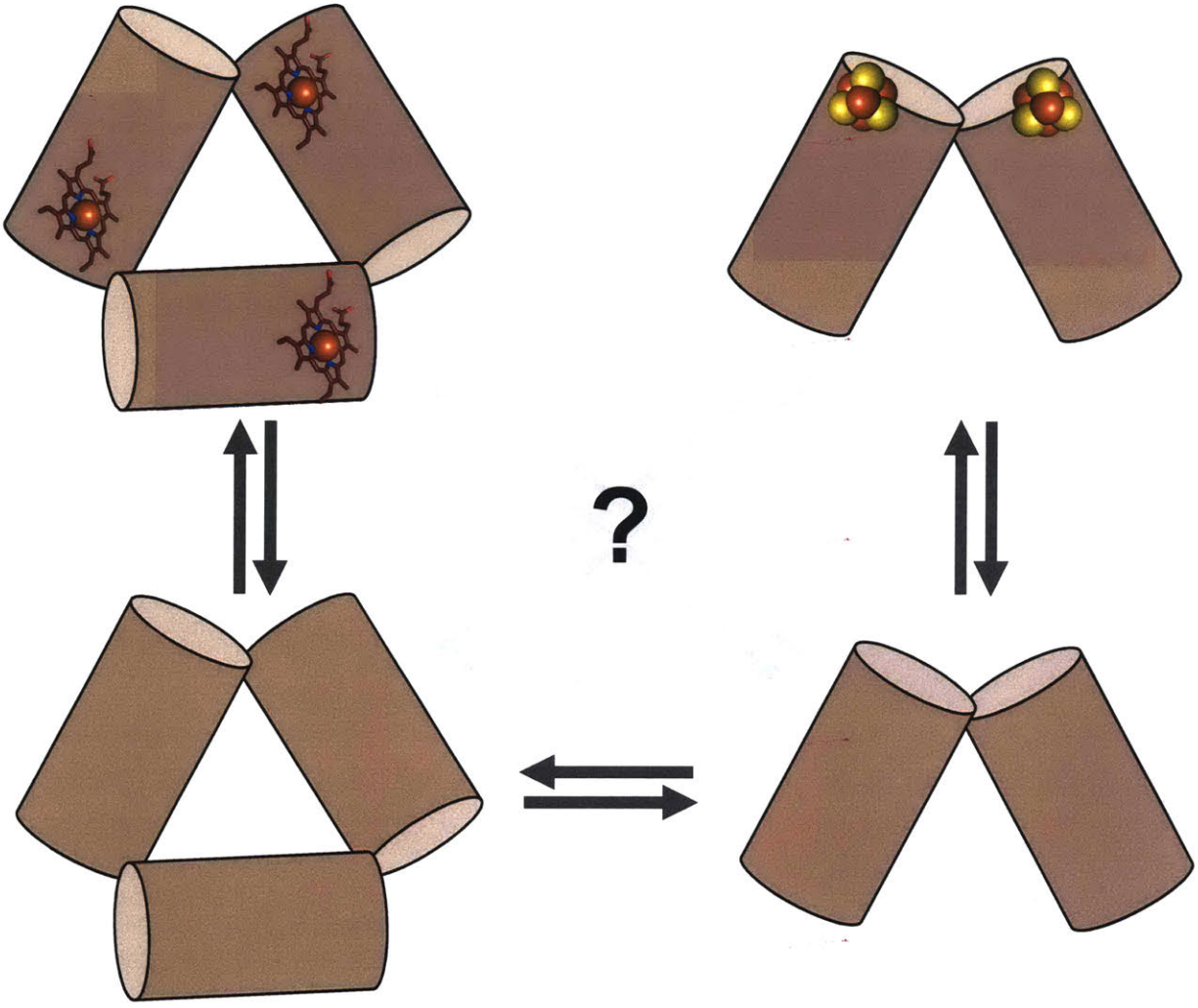



**Figure III.20. Potential model of MFP cofactor binding.** The iron-sulfur cluster bound protein adopts a dimeric form, whereas the heme bound adopts a trimeric form. The binding of one cofactor inhibits binding of the other, but the apo-protein appears to be able to adopt multiple oligomeric states.

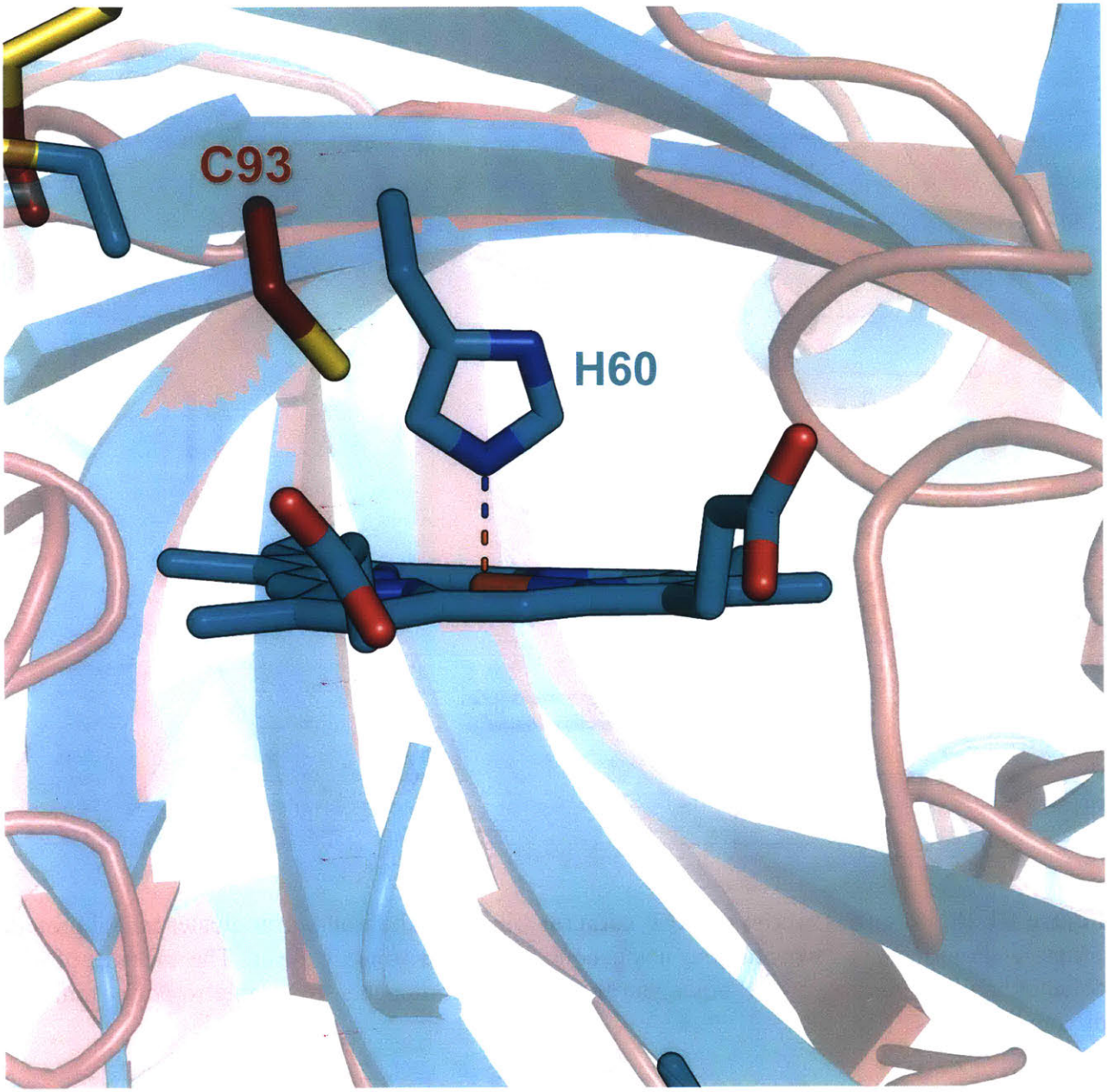



**Figure III.21. Structural alignment of the homology model of MFP with Nitrophorin 7 (NP7) shows Cys93 of the homology model near the Nitrophorin 7 His ligand.** The position of Cys93 in the homology model of MFP (fire brick) aligns well with the position of His60 in NP7 (cyan), the residue that ligates the heme in the nitrophorin family. All cysteines for both structures as well as His60 and the heme moiety in NP7 are shown in stick form.

| Variant | | PfMFPΔ | PfMFP | 4CA | C68A | C70A | C93A | C216A |
|---|---|---|---|---|---|---|---|---|
| Apo-protein (aerobic) | Monomer | Observed | Observed | Observed | | | | |
| | Dimer | Primary | Primary | Primary | Primary | Primary | Primary** | Primary |
| | Trimer | Primary | Primary | | Observed | Observed | Observed | Observed |
| Apo-protein (anaerobic) | Monomer | Observed | Observed | Observed | | | | |
| | Dimer | Primary | Primary | Primary | Primary | Primary | Primary | Primary |
| | Trimer | Observed | Observed | | Observed | Observed | Observed | Observed |
| Cluster-bound protein | Monomer | | | N/A | | | | |
| | Dimer | Primary | Primary | N/A | | | | |
| | Trimer | Observed | Observed | N/A | | | | |
| Heme-bound protein | Monomer | | | | | | | |
| | Dimer | | | | | | | |
| | Trimer | Primary | | | | | | |

**Table III.1. PfMFP and its variants can adopt many different oligomeric states.** For each variant, the primary observed oligomeric state(s) at 5 mg/mL of protein are indicated in red as well as other commonly observed states, indicated by "Observed". Boxes grayed out indicate that this condition has not been visualized by analytical SEC. Note that the absence of an oligomeric profile does not mean it is not accessible, just that we do not typically observe it during analytical SEC experiments. These data are specific to 5 mg/mL protein, as lower or higher concentrations shift the profiles toward lower or higher order oligomers, respectively. **C93A has a main oligomeric state between that of a dimer and a trimer.

| Variant | 5-coordinate species | | 6-coordinate species | |
|---|---|---|---|---|
| | λmax (nm) | Apeak/A280 | λmax (nm) | Apeak/A280 |
| **PfMFPΔ** | | | 412 | 1.85 |
| **4CA** | | | 412 | 1.41** |
| **C68A** | | | 412 | 1.62 |
| **C70A** | 370 | 1.51 | 420 | 0.99 |
| **C93A** | | | 412 | 1.13 |
| **C216A** | 370 | 1.41** | 420 | 1.11 |
| **C68A/C93A** | 370 | 1.58** | 420 | 0.96 |

**Table III.2. Individual cysteine variants display different heme-binding properties.** The presence or absence of a 5- and 6-coordinate heme species is displayed as well as the $\lambda_{max}$ for each species. The maximum $A_{peak}/A_{280}$ for each species observed indicates the intensity of the soret.

** Indicates that the value had not leveled off after the addition of 1 equivalent of heme, suggesting potential binding of more than 1 equivalent of this species.

# REFERENCES


1. Otto, T. D., Wilinski, D., Assefa, S., Keane, T. M., Sarry, L. R., Böhme, U., Lemieux, J., Barrell, B., Pain, A., Berriman, M., Newbold, C., and Llinás, M. (2010) New insights into the blood-stage transcriptome of Plasmodium falciparum using RNA-Seq. *Molecular Microbiology* **76**, 12-24
2. Matz, J. M., and Matuschewski, K. (2018) An in silico down-scaling approach uncovers novel constituents of the Plasmodium-containing vacuole. *Scientific Reports* **8**
3. World Health Organization. (2018) World malaria report 2018. World Health Organization
4. World Health Organization. Malaria Country Profiles, 2018. World Health Organization
5. Ashley, E. A., Pyae Phyo, A., and Woodrow, C. J. (2018) Malaria. *The Lancet* **391**, 1608-1621
6. Phillips, M. A., Burrows, J. N., Manyando, C., Huijsduijnen, R. H. v., Voorhis, W. C. V., and Wells, T. N. C. (2017) Malaria. *Nature Reviews Disease Primers* **3**
7. Miller, L. H., Baruch, D. I., Marsh, K., and Doumbo, O. K. (2002) The pathogenic basis of malaria. *Nature* **415**, 673-679
8. Sigala, P. A., and Goldberg, D. E. (2014) The Peculiarities and Paradoxes of Plasmodium Heme Metabolism. *Annual Review of Microbiology* **68**, 259-278
9. Egan, T. J., Combrinck, J. M., Egan, J., Hearne, G. R., Marques, H. M., Ntenteni, S., Sewell, B. T., Smith, P. J., Taylor, D., Van Schalkwyk, D. A., and Walden, J. C. (2002) Fate of haem iron in the malaria parasite Plasmodium falciparum. *Biochemical Journal* **365**, 343-347
10. Lingelbach, K., and Joiner, K. A. (1998) The parasitophorous vacuole membrane surrounding Plasmodium and Toxoplasma: an unusual compartment in infected cells. *Journal of Cell Science* **111**, 1467-1475
11. Schatten, H., and Ris, H. (2004) Three-Dimensional Imaging of Toxoplasma gondii–Host Cell Interactions within the Parasitophorous Vacuole. *Microscopy and Microanalysis* **10**, 580-585
12. Schnider, C. B., Bausch-Fluck, D., Brühlmann, F., Heussler, V. T., and Burda, P.-C. (2018) BioID Reveals Novel Proteins of the Plasmodium Parasitophorous Vacuole Membrane. *mSphere* **3**, e00522-00517
13. Aurrecoechea, C., Brestelli, J., Brunk, B. P., Dommer, J., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G., Harb, O. S., Heiges, M., Innamorato, F., Iodice, J., Kissinger, J. C., Kraemer, E., Li, W., Miller, J. A., Nayak, V., Pennington, C., Pinney, D. F., Roos, D. S., Ross, C., Stoeckert, C. J., Treatman, C., and Wang, H. (2009) PlasmoDB: a functional genomic database for malaria parasites. **37**, D539-D543
14. Hofbauer, S., Hagmüller, A., Schaffner, I., Mlynek, G., Krutzler, M., Stadlmayr, G., Pirker, K. F., Obinger, C., Daims, H., Djinović-Carugo, K., and Furtmüller, P. G. (2015) Structure and heme-binding properties of HemQ (chlorite dismutase-like protein) from Listeria monocytogenes. *Archives of Biochemistry and Biophysics* **574**, 36-48
15. Knipp, M., Taing, J. J., and He, C. (2011) Reduction of the lipocalin type heme containing protein nitrophorin — Sensitivity of the fold-stabilizing cysteine disulfides toward routine heme-iron reduction. *Journal of Inorganic Biochemistry* **105**, 1405-1412

16. Weichsel, A., Maes, E. M., Andersen, J. F., Valenzuela, J. G., Shokhireva, T. K., Walker, F. A., and Montfort, W. R. (2005) Heme-assisted S-nitrosation of a proximal thiolate in a nitric oxide transport protein. **102**, 594-599
17. Wayland, B. B., and Olson, L. W. (1974) Spectroscopic studies and bonding model for nitric oxide complexes of iron porphyrins. *Journal of the American Chemical Society* **96**, 6037-6041
18. Bonin, I., Martins, B. M., Purvanov, V., Fetzner, S., Huber, R., and Dobbek, H. (2004) Active Site Geometry and Substrate Recognition of the Molybdenum Hydroxylase Quinoline 2-Oxidoreductase. *Structure* **12**, 1425-1435
19. Dobbek, H., Gremer, L., Kiefersauer, R., Huber, R., and Meyer, O. (2002) Catalysis at a dinuclear [CuSMo(O)OH] cluster in a CO dehydrogenase resolved at 1.1-A resolution. *Proceedings of the National Academy of Sciences* **99**, 15971-15976
20. Grell, T. A. J., Kincannon, W. M., Bruender, N. A., Blaesi, E. J., Krebs, C., Bandarian, V., and Drennan, C. L. (2018) Structural and spectroscopic analyses of the sporulation killing factor biosynthetic enzyme SkfB, a bacterial AdoMet radical sactisynthase. *Journal of Biological Chemistry*, jbc.RA118.00536
21. Karnaukhova, E., Rutardottir, S., Rajabi, M., Wester Rosenlöf, L., Alayash, A. I., and Åkerström, B. (2014) Characterization of heme binding to recombinant $\alpha_1$-microglobulin. **5**
22. Champagne, D. E., Nussenzveig, R. H., and Ribeiro, J. M. C. (1995) Purification, Partial Characterization, and Cloning of Nitric Oxide-carrying Heme Proteins (Nitrophorins) from Salivary Glands of the Blood-sucking InsectRhodnius prolixus. *Journal of Biological Chemistry* **270**, 8691-8695
23. Weichsel, A., Andersen, J. F., Champagne, D. E., Walker, F. A., and Montfort, W. R. (1998) Crystal structures of a nitric oxide transport protein from a blood-sucking insect. *Nature structural biology* **5**, 304-309
24. Andersen, J. F., and Montfort, W. R. (2000) The Crystal Structure of Nitrophorin 2: A TRIFUNCTIONAL ANTIHEMOSTATIC PROTEIN FROM THE SALIVA OFRHODNIUS PROLIXUS. **275**, 30496-30503
25. Andersen, J. F., Weichsel, A., Balfour, C. A., Champagne, D. E., and Montfort, W. R. (1998) The crystal structure of nitrophorin 4 at 1.5 A resolution: transport of nitric oxide by a lipocalin-based heme protein. *Structure* **6**, 1315-1327
26. Knipp, M., Ogata, H., Soavi, G., Cerullo, G., Allegri, A., Abbruzzetti, S., Bruno, S., Viappiani, C., Bidon-Chanal, A., and Luque, F. J. (2015) Structure and dynamics of the membrane attaching nitric oxide transporter nitrophorin 7. *F1000Research*
27. Ribeiro, J. M., and Walker, F. A. (1994) High affinity histamine-binding and antihistaminic activity of the salivary nitric oxide-carrying heme protein (nitrophorin) of Rhodnius prolixus. *The Journal of Experimental Medicine* **180**, 2251-2257
28. Brewitz, H. H., Kühl, T., Goradia, N., Galler, K., Popp, J., Neugebauer, U., Ohlenschläger, O., and Imhof, D. (2015) Role of the Chemical Environment beyond the Coordination Site: Structural Insight into Fe (IIIP) rotoporphyrin Binding to Cysteine-Based Heme-Regulatory Protein Motifs. *ChemBioChem* **16**, 2216-2224
29. Wißbrock, A., Paul George, A. A., Hans, Kühl, T., and Imhof, D. (2019) The molecular basis of transient heme-protein interactions: analysis, concept and implementation. *Bioscience Reports* **39**, BSR20181940

30. Russwurm, M., and Koesling, D. (2004) NO activation of guanylyl cyclase. *The EMBO Journal* **23**, 4443-4450
31. Makino, R., Matsuda, H., Obayashi, E., Shiro, Y., Iizuka, T., and Hori, H. (1999) EPR Characterization of Axial Bond in Metal Center of Native and Cobalt-substituted Guanylate Cyclase. **274**, 7714-7723
32. Gateiger, E., Hoogland, C., Gattiker, A., Devaud, S., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005) Protein identification and analysis tools on the ExPASy server. in *The Proteomics Protocols Handbook* (Walker, J. M. ed.), Humana Press. pp 571-607
33. Corpet, F. (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research* **16**, 10881-10890
34. Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research* **42**, W320-W324

# CHAPTER IV: ATTEMPTS TO STRUCTURALLY CHARACTERIZE MFP

## CHAPTER SUMMARY

Although we are beginning to understand the nature of cluster and heme binding to MFP, structural information could add insight into the biological function of this essential protein. Herein, we describe our attempts at obtaining structural information about MFP. We endeavored to obtain structures for the three forms of the protein discussed in Chapter III: apo-MFP, cluster-bound MFP, and heme-bound MFP. As of now, the only reliably reproducible crystals have been obtained using apo-MFP, and no usable data has been obtained for this particular crystal form. Recently, we have identified a reproducible condition of the [Fe-S] cluster bound form of MFP; however, data quality crystals have not been obtained. Additionally, conditions identified to crystallize MFP with heme bound lend more evidence to the coordination of heme by a cysteine residue. Although we have learned many interesting properties of MFP through our attempts to obtain a structure, more work is still needed in order to structurally characterize this elusive protein.

## CONTRIBUTIONS

All plasmids were obtained from Mai Nakashima and Khan Osman (Niles Lab, MIT). Protein was expressed and purified by Allena Goren, Dante Avalos, and Hannah Grupe (Drennan Lab, MIT). All preparations for protein crystallography were performed by Allena Goren and Hannah Grupe (Drennan Lab, MIT). Crystallization experiments were performed by Allena Goren, Hannah Grupe, Phoebe Li, Mary C. Andorfer, Francesca Vaccaro, Gerardo Perez, and Dante Avalos (Drennan Lab, MIT). Synchrotron data was collected at the Northeastern Collaborative Access Team beamlines at Argonne National Labs.

## INTRODUCTION

Thus far, we have described our work toward understanding the cofactor binding of MFP, an essential *Plasmodium* protein. The preclusion of heme binding by the presence of an iron-sulfur cluster could be due to a shared binding site or some structural rearrangement that occurs upon cluster binding that prevents the heme from accessing its native site. Although we can make some conclusions regarding these two hypotheses based on the homology model and the biochemical data, having a directly observed structure could lead to a better understanding of this protein.

Crystallography has long been used as a tool for understanding the structural reasons for molecular and biological phenomena. By taking snapshots of a protein of interest in different states, we can begin to piece together the dynamic actions of biomolecules of interest *in vivo*. For MFP, these snapshots could provide insight into the nature of heme coordination or the mechanism by which cluster-bind precludes heme-binding. As such, we endeavored to structurally characterize MFP by crystallography with the goal of obtaining structures of the three relevant states discussed in Chapter III: apo-MFP, cluster-bound MFP, and heme-bound MFP. Multiple constructs were utilized in an attempt to obtain these structures (**Table IV.1**).

## RESULTS

### HIS-MFP CRYSTALLIZES IN HIGH ALCOHOL SOLUTIONS

Over 20,000 conditions of apo-MFP from *Plasmodium falciparum* have been screened aerobically in an attempt to structurally characterize this essential protein (**Table IV.2**) (1). Very few results of interest have been obtained, with the most promising crystallization conditions containing high amounts of volatile liquid and requiring incubation in a cold room. Ethanol-based conditions produce micro-crystals upon screening but were not reproducible upon scale up (**Figure IV.1A**). Although occasional single crystals form at low protein concentrations in screening experiments under ethanol-based conditions, we are unable to harvest crystals from these conditions as they readily dissolve back into solution upon opening of the crystallization environment.

Many additional conditions have been observed with high amounts of 2-propanol (**Figure IV.1B-D**). The morphology of the crystals obtained during incubation against lower pH buffers appears more multi-lattice by eye, and have proved less reliably reproducible. Crystallization experiments of MFP against 2-propanol, HEPES buffer, and salt incubated in the cold room yield reproducible hexagonal rods (**Figure IV.1C-D**), so these crystals were pursued further to try to obtain data quality crystals (**Table IV.3**).

When cryoprotected in either ethylene glycol or glucose, crystals of MFP yield diffraction indicative of protein crystals (**Figure IV.2**). Upon processing of the data using standard processing software, however, it becomes clear that the data sets are not appropriate for structure determination. The data appear to belong to the space group P622, but can be processed with reasonable statistics in many different space groups (**Table IV.4**). Analysis via Phenix Xtriage (2) indicate severe data issues such as twinning, which is present in almost all crystals of MFP for which data have been collected (**Table IV.5**). In the rare case that the indicated space group of the data is C222, the lack of an appropriate molecular replacement model has made structure solution impossible to this point.

## THE CRYSTALLIZED FORM OF APO-MFP CONTAINS A DISULFIDE BOND

Given the inability for us to solve the structure of MFP using the conditions we have found, we endeavored to understand the crystals that were forming. Non-reducing SDS-page gels of a typical MFP crystal show a single band of approximately 18 kDa molecular weight (**Figure IV.3**). This result was obtained before the identification of the key disulfide bond between Cys70 and Cys216 discussed in Chapter III, so we initially thought that this lower molecular weight band was likely due to a truncation product. This band was also present in the original protein solution used for crystallization, but was a minor component of the solution (**Figure IV.3**). In order to confirm that the protein crystallizing was our protein of interest, the gel band was analyzed by mass spectrometry. The peptides identified correspond with MFP and cover residues 13-35, 41-105, 164-173, and 175-184 of the expression construct, which consists of 213 total residues (**Figure IV.4**). Based on the molecular weight calculated from the gel (18.5 - 21.5 kDa) and assuming a single truncation site, the protein could have been truncated between residues 24-

52 for an N-terminal truncation or 165-190 for a C-terminal truncation, however residues 24-52 and 165-184 were identified by mass spectrometry, leaving the only potential truncation site for the protein to be between residue 184 and 190. Notably, when the crystals were run on a reducing SDS-page gel, no lower band was observed, which is inconsistent with the presence of a truncated form of MFP. We have now identified the lower band as the disulfide bonded form of MFP (See Chapter III) and not a truncated fragment of MFP.

In an attempt to create a more homogeneous solution of MFP with the disulfide bond present in the crystallized form, two methods were pursued. The first was an oxidation of PfMFPΔ using potassium ferricyanide, and the second was the PfMFP C68A/C93A variant. As discussed in Chapter III, both oxidation of PfMFPΔ and the PfMFP C68A/C93A variant produce a single lower band by non-reducing SDS-page gel electrophoresis, corresponding with the disulfide bond being intact under both conditions. Although both avenues have been pursued, neither of these protein solutions has provided diffraction quality crystals at this time.

## HEAVY ATOM SOAKS OF APO-MFP CRYSTALS HAVE NOT YIELDED PHASE INFORMATION

Attempts at soaking the reproducible apo-MFP crystals have been made with multiple heavy atom solutions (**Table IV.6**). Only two heavy atom soaks yielded crystals that were deemed potentially appropriate for data collection. The first of these soaks was the I3C phasing kit (Hampton Research). Although soaking the crystals with 195 mM I3C dissolved in the well solution, we observed that the crystals began to dissolve back into solution, so crystals were soaked in both the I3C and a cryoprotectant for less than a minute before flash cooling. Although data was collected, no usable data was obtained from these crystals (**Table IV.7**). We also soaked crystals with 10 mM of ethyl mercuric phosphate (Hampton Research). This derivative was added to the drop and allowed to soak for 1 hour before performing cryoprotection in the normal manner for these crystals. These crystals also did not diffract well, with a radial smearing pattern observed, however data sets were collected (**Table IV.7**). Given the propensity for this

protein to bind iron containing cofactors, structures with both an [Fe-S] cluster bound and heme bound have been pursued more than heavy atom soaks.

### [Fe-S] cluster-bound MFP has not yielded a structure

The presence of an [Fe-S] cluster can aid in structure solution by providing the anomalous iron signal needed to phase data. Although many conditions have been tested against a cluster-bound form of MFP (**Table IV.8**), no reproducible conditions have been found. Four conditions have been identified as potentially interesting to pursue, but none yielded data quality crystals. When 5mg/mL MFP with 35% cluster was set against 1.0% (w/v) PEG-3350, 1.0 M ammonium sulfate, and 0.1 M Bis-TRIS, pH 5.5, long needles formed in screen trays (**Figure IV.5A**). These crystals have not been reproducible upon scale up nor have crystals with better morphology been produced from this condition. The same protein also yielded haystack crystals when set against 0.1 M MES, pH 6.0 and 0.8 M ammonium sulfate (**Figure IV.5B**). These crystals were reproducible, but only as haystacks. Attempts at altering the morphology have not been successful.

An intriguing condition was identified while performing sparse matrix screens of 7.5 mg/mL MFP reconstituted with ~90% [4Fe-4S] cluster. The precipitating solution of 0.2 M zinc acetate, 0.1 M sodium acetate, pH 4.5, and 10% (w/v) PEG-3,000 produced a small cluster of needles when incubated in a 1:1 drop ratio at 6 °C (**Figure IV.5C**). Although we have not obtained structural data from this condition as of yet, we were able to reproduce small needles in this condition not yet appropriate for data collection but promising for further optimization. Additionally, crystals were obtained when PfMFPΔ at 7.5 mg/mL was set against 0.1 M di-Ammonium hydrogen phosphate, 50% (v/v) 2-methyl-2,4-pentanediol, and 0.1 M TRIS, pH 8.5 and incubated at room temperature (**Figure IV.5D**). These crystals did not contain color, and were not pursued further due to this lack of color and the phosphate in the condition, which often can yield salt crystals.

### Heme bound MFP has also not yielded a structure

In addition to the heavy metal soaks and the screening of crystallization conditions for [Fe-S] cluster-bound MFP, the heme binding properties of MFP provided an additional avenue to

pursue in order to obtain phase information for crystals of this novel protein. Both aerobic and anaerobic screening of MFP incubated with hemin have been performed, but no crystals have been obtained as of yet. After three months, quasicrystals were observed of 4CA incubated with 5 molar equivalents of hemin set against a precipitating solution of 1.0 ammonium di-hydrogen phosphate and 0.1 M tri-sodium citrate, pH 5.6 (**Figure IV.6A**). Additional quasicrystals were obtained when 4CA incubated with one molar equivalents of hemin was set against the Morpheus II screen (Molecular Dimensions) in many of the conditions using Morpheus Buffer 4, containing both MES and imidazole. The best quasicrystals observed were from the well containing 0.03 M diethylene glycol, 0.03 M triethylene glycol, 0.03 M tetraethylene glycol, 0.03 M pentaethylene glycol, 0.1 M imidazole/MES buffer mix, pH 6.5, 12.5% (v/v) 2-methyl-2,4-propanediol, 12.5% (w/v) PEG-1000, and 12.5% (w/v) PEG-3350 (**Figure IV.6B**). Although these quasicrystals were highly reproducible upon scale up, diffraction quality crystals have not been observed.

## DISCUSSION

Structural biology can provide insights into protein function, even when the biological function of a protein is not well understood. As such, we endeavored to identify conditions that would lead to a structural understanding of MFP from *Plasmodium falciparum*. As of now, we have not been able to obtain a structure of MFP, either aerobically or anaerobically.

Protein crystallization conditions of MFP have been identified, and we were optimistic about our ability to determine the structure of this essential protein. Unfortunately, due to the high alcohol content of the precipitation solutions, the crystallization drops are highly susceptible to evaporation, causing the crystals to crack, dissolve, and move around the drop during harvesting. This volatility greatly complicates the looping process and has led to poor data. Additionally, very few cryo-protectant solutions have been successfully identified, with most common cryoprotectants causing the crystal to dissolve back into solution or reducing or eliminating all diffraction. The two cryoprotectants identified as appropriate for the high alcohol conditions utilized for MFP crystallization both do not appear to cause changes to the crystal by eye, and data has been collected on these crystals (**Figure IV.2**), however no data appropriate for

structure solution has been obtained. The data obtained are twinned, and without an appropriate molecular replacement model, space group identification of the crystals is almost impossible.

Although the crystals have not provided a structure by X-ray diffraction, they have yielded some information on structural conformation of the protein. The crystallized form of MFP appears to be disulfide bonded, by SDS-page gel (**Figure IV.3**). Given the homology model (**See Figure II.4**), biological data obtained through oxidation of the protein (**See Figure III.18**), and the fact that the crystallized form of MFP contains this disulfide bond, it is likely that the C-terminus of MFP is relatively disordered and structurally stabilized by the disulfide bond. In Chapter III, we identified an MFP variant, C68A/C93A, which seems to stabilize the disulfide bond. We anticipate this construct will be beneficial for crystallization of MFP, but given our recent discovery of this particular variant, very few screening experiments have been run on this construct.

Although the [Fe-S] cluster bound form of MFP likely precludes the presence of the relevant disulfide bond due to its potential utilization of all four cysteine residues present in MFP, we were particularly excited by the potential of this form of MFP. The [4Fe-4S] cluster is likely ligated by all four cysteines, including Cys216, stabilizing the potentially disordered C-terminus as well as providing iron atoms for anomalous data collection. As of now, no data have been obtained on these crystals, as most conditions of interest identified in sparse matrix screens have not been reproducible upon scaling up for optimization.

In addition to the [Fe-S] cluster, a heme bound structure would also provide a method for phasing based on the anomalous signal of the iron atom within the heme. Many attempts at identifying crystallization conditions of heme bound MFP were attempted; however, the only identified conditions were that of aerobic 4CA with heme bound. Although the nature of the axial ligands of aerobic 4CA appears to differ from that of PfMFPΔ, we were pleasantly surprised to identify conditions in which quasicrystals formed upon incubation of 4CA with hemin (**Figure IV.6**). Interestingly, these quasicrystals form under conditions containing imidazole, a molecule known to indiscriminately bind to the open coordination site of a 5-coodinate heme moiety. This observation lends more evidence to a potential sixth coordination of ferric heme by one of the four cysteine residues of MFP discussed in Chapter III.

## CONCLUSION

Although no structure of MFP has been obtained after many experiments, the promising new construct C68A/C93A could prove useful for visualizing this important protein. Cluster-bound and heme-bound protein have also not yielded a structure. More work will be needed in order to obtain structural information about MFP.

## EXPERIMENTAL PROCEDURES

### PURIFICATION OF PROTEINS FOR CRYSTALLOGRAPHY

Initial purification of proteins was as described in Chapter III. Additional purification of proteins for crystallization are indicated in **Table IV.2** in the "additional treatments" column and were performed in the following manners.

*Buffer Exchange*

After affinity purification of MFP by either HisTrap (GE life sciences) or Streptactin XT (IBA) column, protein was concentrated in 10 kDa MWCO Amicon spin concentrators. When the volume of protein fell below 1 mL, 14 mL of buffer for buffer exchange were added and the protein was spin concentrated down to less than 1 mL again. This process was repeated 3 times. After the third addition of buffer, protein was concentrated to approximately 40 mg/mL concentration, measured by UV-Vis spectroscopy ($\varepsilon_{280,\ \text{His-MFP}}$ = 19,160 $M^{-1}$ $cm^{-1}$, $\varepsilon_{280,\ \text{PfMFP}}$ = 24.660 $M^{-1}$ $cm^{-1}$, this value and all subsequent were obtained through ExPASy ProtParam (3)). The concentrated protein was stored at 4 °C for further analysis and crystallography.

*Diluted fractions from affinity column*

After a HisTrap affinity column (GE life sciences), pure fractions of His-MFP were concentrated in a 10 kDa MWCO Amicon spin concentrator to 82 mg/mL. The concentrated protein in 50 mM Tris, pH 8.0, 300 mM NaCl and 500 mM imidazole was diluted to the desired concentration (either 10 or 15 mg/mL) using 50 mM Tris, pH 8.0 and 100 mM NaCl buffer. The final concentrations of NaCl and imidazole were calculated based on the ratio of volumes used for dilution. Diluted protein was used immediately for crystallographic screens.

*Preparatory SEC*

After affinity purification of MFP, protein was concentrated to less than 2 mL or less than 5 mL for purification on a Hi Load 16/60 Superdex$^{TM}$ 200 preparatory grade column (GE life sciences) or a Hi Load 26/60 Superdex$^{TM}$ 200 preparatory grade column (GE life sciences), respectively. Protein was injected onto the pre-equilibrated column, and progress through the column was monitored by $A_{280}$. Fractions deemed to have protein by $A_{280}$ were then analyzed by SDS-page gel electrophoresis. Pure fractions of MFP were pooled and concentrated in a 10 kDa MWCO Amicon spin concentrator until protein concentration was approximately 40 mg/mL by UV-Vis spectroscopy ($\varepsilon_{280,\ His\text{-}MFP}$ = 19,160 $M^{-1}$ $cm^{-1}$, $\varepsilon_{280,\ PfMFP}$ = 24.660 $M^{-1}$ $cm^{-1}$ (3)). Concentrated protein was stored at 4 °C for further analysis and crystallography.

*Preparation of aerobic cluster-bound MFP*

After HisTrap affinity purification (GE life sciences) of His-MFP in the presence of 5 mM dithiothreitol (DTT) as a reducing agent, a brown color was observed. The protein was concentrated to less than 2 mL of solution and subjected to additional purification on a Hi Load 16/60 Superdex$^{TM}$ 200 preparatory grade column (GE life sciences) equilibrated with 50 mM HEPES, pH 8.0 and 100 mM NaCl. The individual fractions containing protein by monitoring the absorption of solution at 280 nm were individually concentrated and analyzed by UV-Vis spectroscopy using the pedestal of a NanoDrop 2000c (ThermoFisher). The absorbance was determined at both 280 nm and 410 nm to observe the content of protein and [Fe-S] cluster, respectively. The two fractions deemed to have the highest $A_{410}$ to $A_{280}$ ratios were pooled and diluted to 20 mg/mL protein by $A_{280}$ ($\varepsilon_{280,\ His\text{-}MFP}$ = 19,160 $M^{-1}$ $cm^{-1}$ (3)). Because of the low concentration of cluster, the $A_{410}$ was used to determine cluster content ($\varepsilon_{410,\ [4Fe\text{—}4S]} \sim$ 15,000 $M^{-1}$ $cm^{-1}$). This value was compared to the concentration of protein measured by $A_{280}$ to determine the cluster content.

*Re-purification of MFP after incubation with heme overnight*

After both Streptactin XT affinity purification (IBA) and preparatory SEC on a Hi Load 16/60 Superdex$^{TM}$ 200 preparatory grade column (GE life sciences), PfMFPΔ was titrated with hemin (Sigma Aldrich) before re-purification. A 10 mM hemin solution in 0.5 M sodium hydroxide was used to titrate 50 mL of 6 μM PfMFPΔ by adding 10 μL of hemin to the protein while stirring. The protein was incubated with the heme for 5 minutes. Fifteen heme spikes were performed to

bring the final concentration of hemin to 30 μM (5 molar equivalents). This solution was filtered with a 0.22-micron sterile filter and re-purified by affinity chromatography, preparatory SEC, and concentrated to 5 mg/mL by $A_{280}$ ($\varepsilon_{280,\ PfMFP\Delta}$ = 24.660 $M^{-1}$ $cm^{-1}$ (3)) as previously reported. Concentrated protein was stored at 4 °C for further analysis and crystallography.

*Desalting procedure*

After Streptactin XT affinity purification (IBA), PfMFPΔ was desalted using a HiPrep 26/10 Sephadex G-25 desalting column (GE life sciences) previously equilibrated in 50 mM HEPES, pH 8.0, 100 mM NaCl, and 5 mM DTT.

## PROTEIN PREPARATION FOR CRYSTALLIZATION

All proteins, both aerobic and anerobic, were filtered through a costar Spin-X 22 μm cellulose spin filter by centrifugation immediately before screening. Additional preparations of proteins outlined in **Table IV.2** were performed in the following manner.

*Co-crystallization with compound I and compound 4*

Six compounds were identified as potential binders to MFP through a small molecule microarray not discussed in this work. The proprietary compounds, compound 1 and compound 4 (obtained from the Niles Lab, MIT), were used for co-crystallization experiments. Purified MFP was incubated with 10 mM of the compounds listed for approximately 1 hour. All compounds were dissolved in 100% DMSO to a final concentration of 100 mM. Buffer concentrations and protein concentrations listed are concentrations after dilution to 90% (v/v) by the addition of compound. Solutions were used immediately for crystallographic screens.

*Incubated with hemin*

A stock of 10 mM hemin (Sigma Aldrich) in 0.5 M NaOH was used for heme incubations, both aerobically and anaerobically. Aerobically, solutions were made by diluting concentrated protein in 50 mM HEPES, pH 8.0 and 100 mM NaCl to the desired volume less the volume of 10 mM hemin needed to make a 1x or 5x molar equivalent of hemin. The hemin was then added to the protein and the solution was incubated overnight at 4 °C. Solutions were filtered and screened immediately. Anaerobically, the same procedure was performed with the addition of 5 mM DTT to the buffer.

*Formation of oxidized protein*

After Streptactin XT affinity purification (IBA) ofPfMFPΔ, protein in 50 mM HEPES, pH 8.0, 300 mM NaCl, and 50 mM biotin was incubated with 10 mM potassium ferricyanide overnight. The protein was transferred to 10 kDa MWCO dialysis tubing and 15 mL of oxidized protein was dialyzed against 4 L of 50 mM HEPES, pH 8.0 and 100 mM NaCl. This dialyzed protein was concentrated and stored at 4 °C for further analysis and crystallography.

*Reconstitution of a [4Fe-4S] cluster*

Reconstitution of a [4Fe-4S] cluster in MFP was performed as outlined in chapter III. Cluster content was measured by calculating iron concentration and protein concentration by a ferene assay. Briefly, 100 µL of ~ 80 µM reconstituted protein was removed for the MBraun anaerobic chamber and added to 125 µL of a 0.4% (w/v) ascorbate solution, 7.5 µL concentrated hydrochloric acid, and 12.5 µL of 5 mg/mL ferene. This solution was vortexed for 30 second, after which 500 µL of 8 M guanidinium chloride and 100 µL of 10 M ammonium acetate were added. The solution was incubated for at least 30 minutes, after which the absorbance was read and compared to the extinction coefficient for iron bound ferene of 34,500 $M^{-1}cm^{-1}$ at 593 nm. The calculated concentrations were increased by 8.45-fold to account for dilution during the assay preparation. The concentration of iron was compared to the concentration of protein calculated by a standard Bradford assay with no correction coefficient.

## PROTEIN CRYSTALLIZATION

*Screens utilized for crystallization attempts*

The following screens were used in an attempt to identify crystallization conditions for optimization: AmSO4 (Qiagen), ComPAS (Qiagen), Index (Hampton Research), JCSG+ (Qiagen), LMB (Molecular Dimensions), MCSG1 (Anatrace), MCSG2 (Anatrace), MCSG3 (Anatrace), MCSG4 (Anatrace), MembFac (Hampton Research), MidasPlus (Molecular Dimensions), Morpheus (Molecular Dimensions), Morpheus II (Molecular Dimensions), PACT (Qiagen), PEG/Ion (Hampton Research), PEGRx (Hampton Research), and ProComp (Qiagen). All screens were purchased in 96-well deep block format.

*Aerobic sparse matrix screening*

Sparse matrix screens were set up using the sitting-drop vapor diffusion method with MFP between 1.0 and 40 mg/mL as laid out in **Table IV.2**. Screens were set in a 96-3 low volume Intelli-plate (Art Robbins Instruments) using a Phoenix crystallization robot (Art Robbins Instruments). Sitting drop experiments were set against 70 μL of crystallization solution. Screens were stored in a Rock Imager 1000 (Formulatrix) at the temperature stated in **Table IV.2**. Images were collected at specified time intervals.

*Anaerobic sparse matrix screening*

Sparse matrix screens were set up using the sitting-drop vapor diffusion method with MFP between 1.0 and 40 mg/mL and varying drop concentrations as laid out in **Table IV.8**. In a 96-3 low volume Intelli-plate (Art Robbins Instruments), 71 μL of well solution was set and the tray and solutions were flash frozen in liquid nitrogen and taken into an MBraun anaerobic chamber under $N_2$ environment ($O_2$ < 0.1-0.6 ppm). Trays were immediately covered with sealing film and stored for two days in the anaerobic chamber to allow for thawing and off-gassing of oxygen into the head space of the tray. After two days, the trays were set with protein drops under anaerobic conditions. Protein prepared as in **Table IV.8** was set as sitting drops using a Mosquito pipetting robot (TTP Labtech) at the ratios stated. Trays were resealed and stored in either a Rock Imager 182 (Formulatrix) specially designed to function within the MBraun anaerobic chamber or were removed from the anaerobic chamber and stored in an aerobic Rock Imager 1000 (Formulatrix) when stored at 21 °C or 6 °C, respectively. Images were collected at specific time intervals.

*Crystallization of MFP for data collection*

The conditions of all crystals used for data collection are listed in **Table IV.3**. Sitting-drop vapor diffusion crystallization experiments were performed in a 24-well Intelli-plate (Art Robbins Instruments). Protein used for these experiments varied from 5 to 20 mg/mL concentration measured by UV-Vis spectroscopy at 280 nm ($\varepsilon_{280,\text{ His-MFP}}$ = 19,160 $M^{-1}$ $cm^{-1}$, $\varepsilon_{280,\text{ PfMFP}}$ = 24.660 $M^{-1}$ $cm^{-1}$ (3)). Crystallization experiments were set with 1 μL of protein and 1 μL of well solution set against a 400 μL well. Well solutions were comprised of 0.11 - 0.18 M $MgCl_2$, 25-27% (v/v) 2-propanol, and 0.1 M HEPES or Tris between pH 7.5 and 8.2 in the presence or

absence of 1% 2-methyl-2,4-pentanediol (MPD) or paraffin oil covering the well (**Table IV.3**). The experiments were incubated at 6 °C and crystals appeared within 1 to 7 days of incubation.

Many cryoprotectant solutions were attempted, however only two cryoprotectants, ethylene glycol (Hampton Research) and glucose (Sigma Aldrich), yielded data quality frozen crystals. Crystals were cryo-protected by gradual addition of well solution with either 31.25% (v/v) ethylene glycol, 31.25% (v/v) MPD or 30% (w/v) glucose added. To the 2 µL crystallization drop, 0.6 µL of cryoprotectant was added. The drop was then resealed and allowed to incubate for 10 minutes, after which the crystals were observed to ensure there was no visible cracking. An additional 0.6 µL of cryoprotectant was added, and the drop was resealed and incubated for 10 minutes. This process repeated with the following steps of 1 µL, 2 µL, 2.4 µL, and 2.4 µL and observation of the crystals to ensure no visible cracking. After the last addition, crystals were either harvested or the drop was resealed and left for the amount of time stated in **Table IV.3**. Once harvested, crystals were flash cooled with liquid nitrogen and stored for data collection.

*Seed stock preparation*

For all seed stocks, crystals in the 0.3 µL drop were harvested from screen trays using a pipette, and resuspended in the 70 µL of well solution into a seed bead tube (Hampton Research). The tube was vortexed for 2 minutes to create a seed stock. Seed stocks were made using three different conditions. The first condition was apo-PfMFPΔ at 10 mg/mL in 50 mM HEPES, pH 8.0 and 100 mM NaCl crystallized against 0.1 M magnesium chloride ($MgCl_2$), 25 % (v/v) 2-propanol, and 0.1 M HEPES, pH 7.5 (from ComPAS screen, isopropanol seed stock). Both fresh and frozen isopropanol seed stock was used for screening. The frozen stock was stored at -20 °C. The second condition was 10 mg/mL PfMFPΔ in 50 mM HEPES, pH 8.0 and 100 mM NaCl crystallized against 0.1 M sodium phosphate citrate, pH 4.2, 40% (v/v) ethanol, and 5% (w/v) PEG-1000 (from JCSG+ screen, ethanol seed stock). This seed stock was stored at -20 °C for future use. The third condition was 5 mg/mL 4CA incubated with 5 molar equivalents of hemin crystallized against 1.0 M ammonium di-hydrogen phosphate and 0.1 M tri-sodium citrate, pH 5.6 (from ComPAS screen, 4CA heme seed stock). This seed stock was used immediately after preparation.

*Seeding of sparse matrix screening*

Sparse matrix screens of the protein were set as outlined in **Table IV.2** for aerobic and **Table IV.8** for anaerobic experiments. Aerobic seeding experiments were performed using a NT8® drop setter for protein crystallography (Formulatrix). Briefly, 71 μL of a sparse matrix screen was set in the well of a 96-3 low volume Intelli-plate (Art Robbins Instruments) using a multichannel pipette. Sitting-drop vapor diffusion experiments were performed using 250 nL of well solution, 200 nL of protein, and 50 nL of seed stock. These experiments were stored in the same manner as the original screening experiments. Anaerobic seeding experiments were performed in the same manner as aerobic experiments with the following changes. To obtain anaerobic screens, the Intelli-plate wells were set in the same manner and then flash frozen in liquid $N_2$. The frozen tray was gas exchanged into an MBraun anaerobic chamber ($O_2$ = 0.1-0.6 ppm), resealed, and left to equilibrate for two days. The trays were set using a Mosquito pipetting robot (TTP Labtech) two days after being brought into the anaerobic environment.

*Heavy atom soaking of crystals*

Multiple heavy atom derivatives were attempted (**Table IV.6**), but most did not yield loopable crystals. Two attempts, outlined here, did yield loopable crystals, but the data did not yield a structure. Two μL of a solution of 390 mM 5-amino-2,4,6-triiodoisophthalic acid (I3C phasing kit, Hampton Research) in a stabilizing solution of 0.11 M $MgCl_2$, 25% 2-propanol, 0.1 M HEPES, pH 7.4, and 50% ethylene glycol was added to the 2 μL drop of protein crystals and the drop was resealed for 1 minute. Crystals were looped and flash cooled with liquid nitrogen. Crystals appeared to be dissolving under these conditions, and no usable anomalous signal was obtained from these crystals. A mercury soak was performed by adding 0.2 μL of a 100 mM solution of ethyl mercuric phosphate (Hampton Research) dissolved in water to the 2 μL drop. The drop was covered again and allowed to equilibrate for 30 minutes. These crystals did not dissolve. The crystals were cryoprotected by addition of 2 μL of 100% ethylene glycol and recovered to incubate for four hours. Crystals were harvested and flash cooled in liquid nitrogen before data collection. Data statistics for all derivatized crystals can be found in **Table IV.7**.

## ATTEMPTS AT STRUCTURE SOLUTION

All data sets were collected at the Northeastern Collaborative Access Team beamlines at Argonne National Labs (**Table IV.7**). Data were processed in either HKL2000 (4) or XDS (5). Anomalous data sets were analyzed by Xtriage (2) to determine whether anomalous signal was present and what resolution cutoff to use while finding sites. For data sets with anomalous signal, multiple programs were used to attempt site finding, including Phenix Autosol (6) and SHELX (7). For each data set, four sites were searched for as a starting point. The only appropriate sites obtained were from data set CLD_956_10_I using a 6.0 Å resolution for site finding, which found 5 sites with a correlation coefficient of 0.56. Phenix Phaser-EP (6) was used to generate initial $F_o$ maps. These maps were uninterpretable (FOM 0.19 to 4.0 Å for both hands). Native data was subjected to molecular replacement using Phenix Phaser-MR and poly-alanine models that were created using known PDB coordinates of different lipocalin proteins. All models used as templates for the poly-alanine models for molecular replacement are listed in **Table IV.9**. Results from molecular replacement were also uninterpretable, with all TFZ values below 8.0 and uninterpretable maps, so no further analysis of the data sets were performed.

## IDENTIFICATION OF MODELS FOR MOLECULAR REPLACEMENT

Known coordinate files were identified as potential molecular replacement models by either HHPred (8), TDRTargets (9), or a BLAST (10) search.

### *Model identification by HHPred*

During the generation of the homology model of MFP discussed in Chapter II, HHPred identified 94 PDB files to utilize for the generation of the homology model (8). Models to use for molecular replacement were chosen based on sequence coverage and sequence identity. A lower limit of 90% probability was used for identification of potential models. Models used are listed in **Table IV.9**.

In addition to generating poly-alanine models from the PDB coordinates identified, the homology model generated from PfMFPΔ and discussed in Chapter II was used to attempt molecular replacement.

*Model identification by TDR Targets*

The gene PF3D7_0925900 or *mfp* was identified to be present in the TDR Targets database (9). Four other *mfp* genes in *Plasmodium* species were identified through the ortholog group OG5_166872. The models identified as potential homologs from this database are listed in **Table IV.9**.

*Model identification by BLAST search*

A standard protein BLAST search was performed using the NCBI BLAST tool (10) and the Protein Data Bank proteins (PDB) database. All four wild-type sequences listed in **Figure II.6** were used. No models were identified for *Plasmodium falciparum*. All other coordinates are listed in **Table IV.9**.

## MASS SPECTROMETRY OF MFP CRYSTALS

A non-reducing SDS page gel was run on crystals obtained during screening against 0.1 M $MgCl_2$, 25% (v/v) 2-propanol, and 0.1 M HEPES, pH 7.5. All common supplies were sterilized with ethanol before the gel was run. A single crystal was looped from the screen tray and rinsed through the well solution. The crystal was then dissolved in 50 µL of 50 mM HEPES, pH 8.0 and 100 mM NaCl and 50 µL of 2x Laemmli sample buffer (BioRad) was added. 25 µL of this solution was loaded onto an Any kD pre-cast gel (BioRad) and run at 180V for 30 minutes. The gel was stained with SimplyBlue™ SafeStain (ThermoFisher) and sent for mass spectrometry analysis by QSTAR hybrid Q-TOF analysis in the Biopolymers and Proteomics Facility at the Koch Institute at MIT.

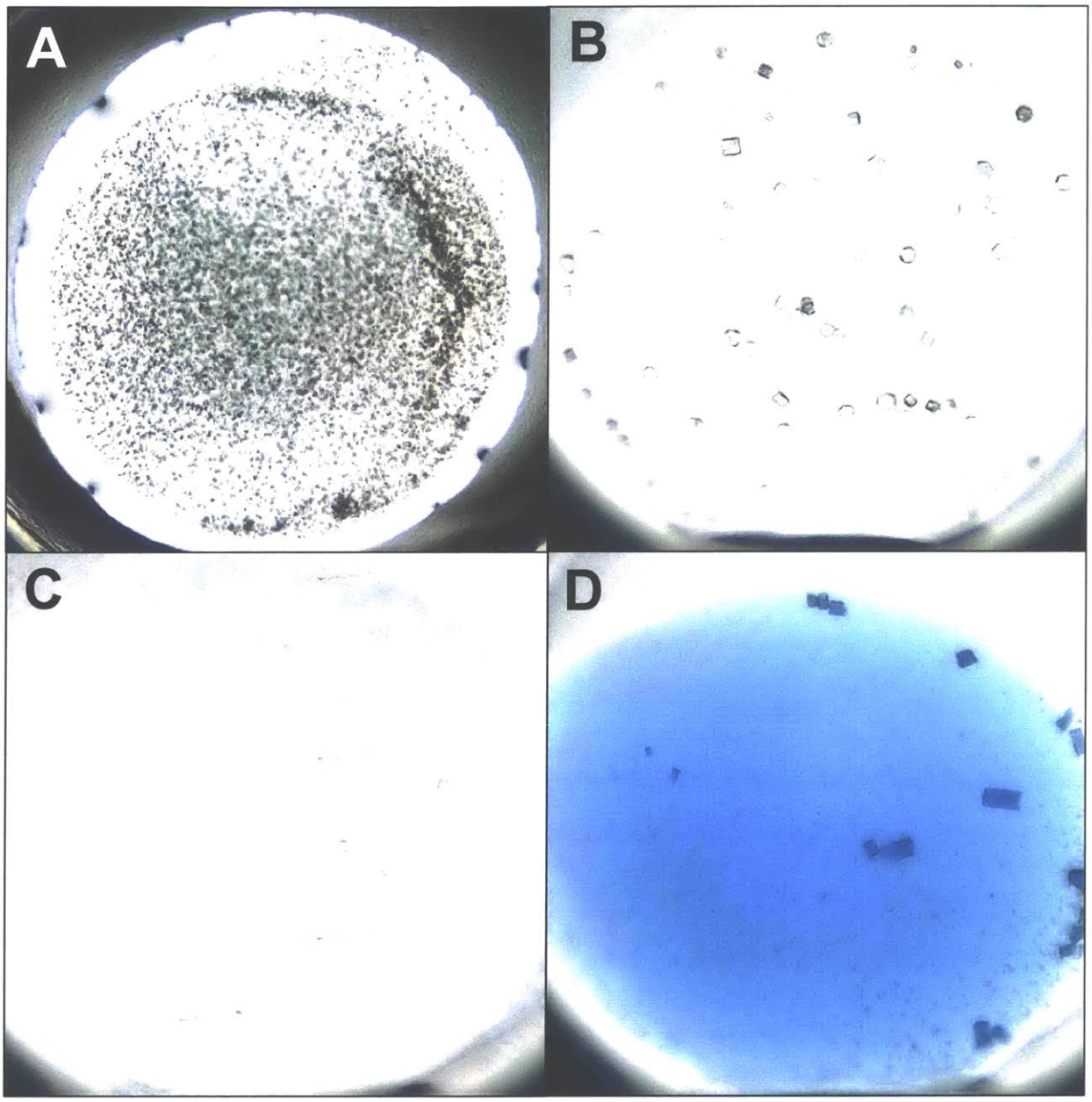



**Figure IV.1. His-MFP crystallizes in high alcohol conditions.** His-MFP forms either microcrystals or hexagonal rods under 20-40% (v/v) alcohol when incubated at 4 °C. **(A)** Microcrystals formed after 5 days during screening against 0.1 M sodium phosphate citrate, pH 4.2, 40% (v/v) ethanol and 5% (w/v) PEG-1000. **(B)** Hexagonal crystals formed after 2 days during screening against 0.2 M calcium chloride, 20% (v/v) 2-propanol, and 0.1 M sodium acetate, pH 4.6. **(C)** Hexagonal crystals formed after 12 hours during screening against 0.1 M HEPES, pH 7.5, 25% (v/v) 2-propanol, and 0.1 M magnesium chloride. **(D)** Crystals from (C) stained with izit dye (Hampton Research), indicating a strong likelihood that the crystals are protein.

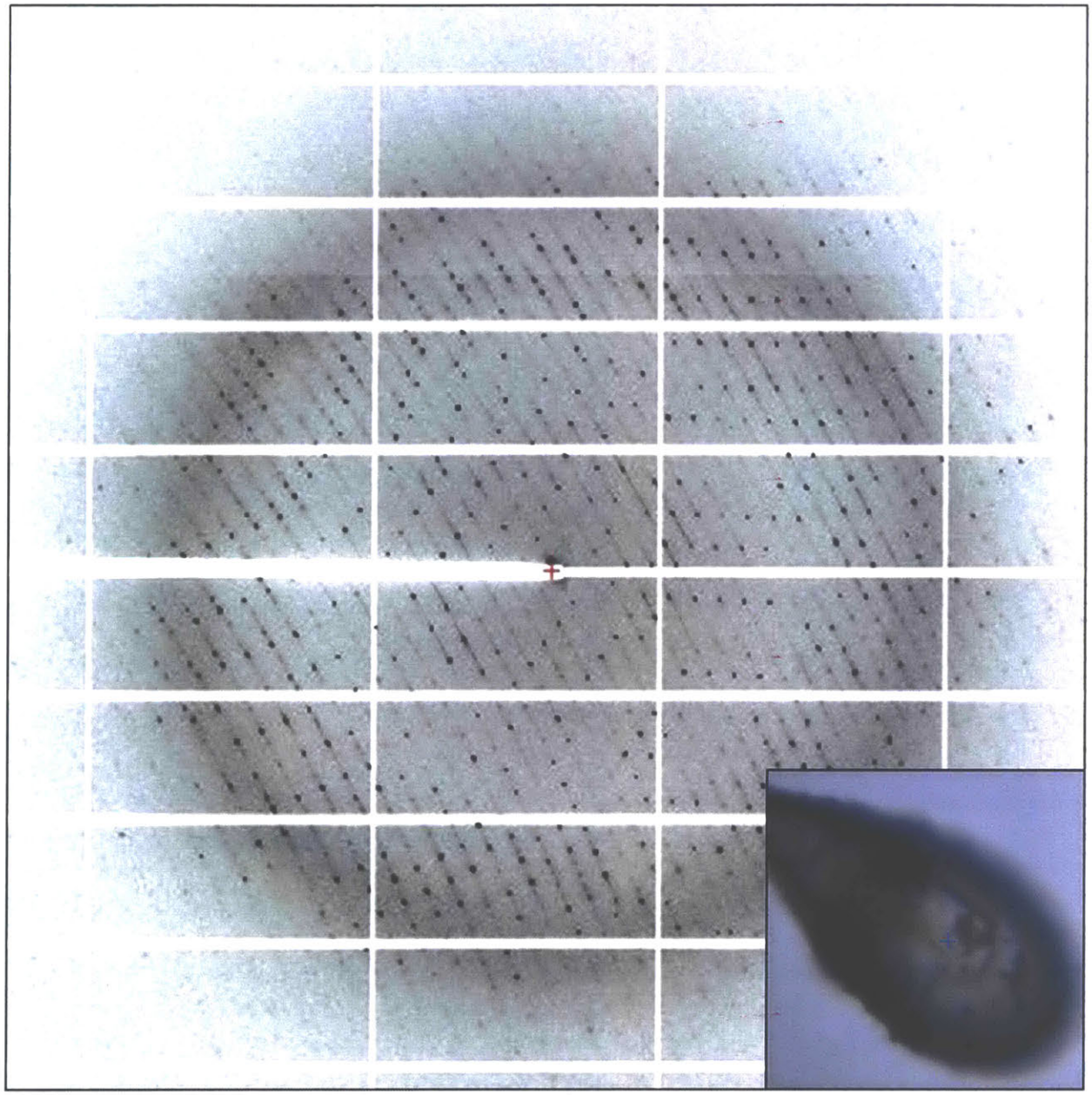


**Figure IV.2. His-MFP crystals diffract to ~2.0 Å resolution.** The diffraction pattern appears to be either twinned or multi-lattice, with multiple loons being observed. **(Inset)** The crystals appear to be singular by eye, with potential cracking occurring.

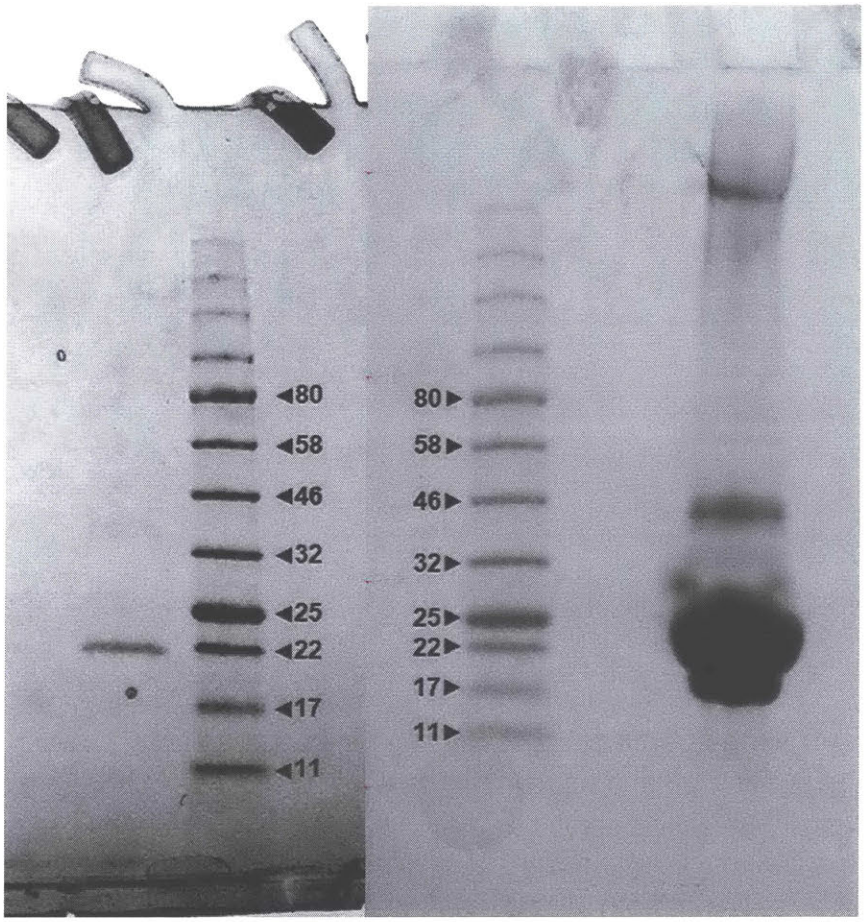



**Figure IV.3. Crystals of His-MFP run lower on a gel than the expected molecular weight of apo-MFP.** A second, lower band was identified in the His-MFP protein solutions. The left panel shows an SDS-page gel run of the protein crystals and shows a band corresponding to a molecular weight of approximately 18.5-21.5 kDa. The right panel is an SDS-page gel run on the highly concentrated solution used for crystallographic screens. The lower molecular weight band corresponding to the protein found to crystallize appears to be present in the original protein solution but is unresolved from the higher band attributed to His-MFP.

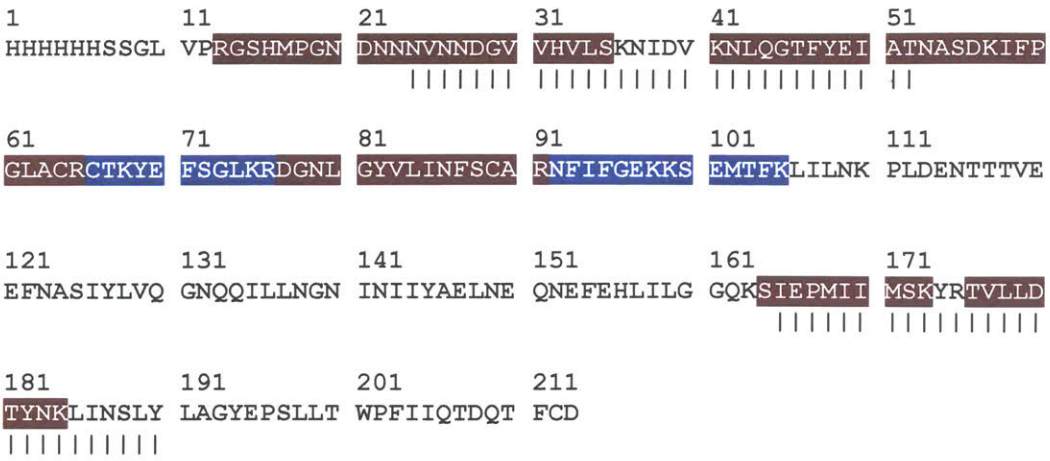



**Figure IV.4. A digest-mass spectrometry analysis confirms that the low molecular weight band is His-MFP.** Reads were identified across most of the protein. Regions of the protein that were contained in a single peptide are highlighted in red. Regions of the protein that were found in multiple peptides are in blue. Regions not identified by mass spectrometry are not highlighted. Vertical bars under the sequence indicate potential cleavage sites based on the approximate molecular weight of the gel band and assuming a single cleavage site.

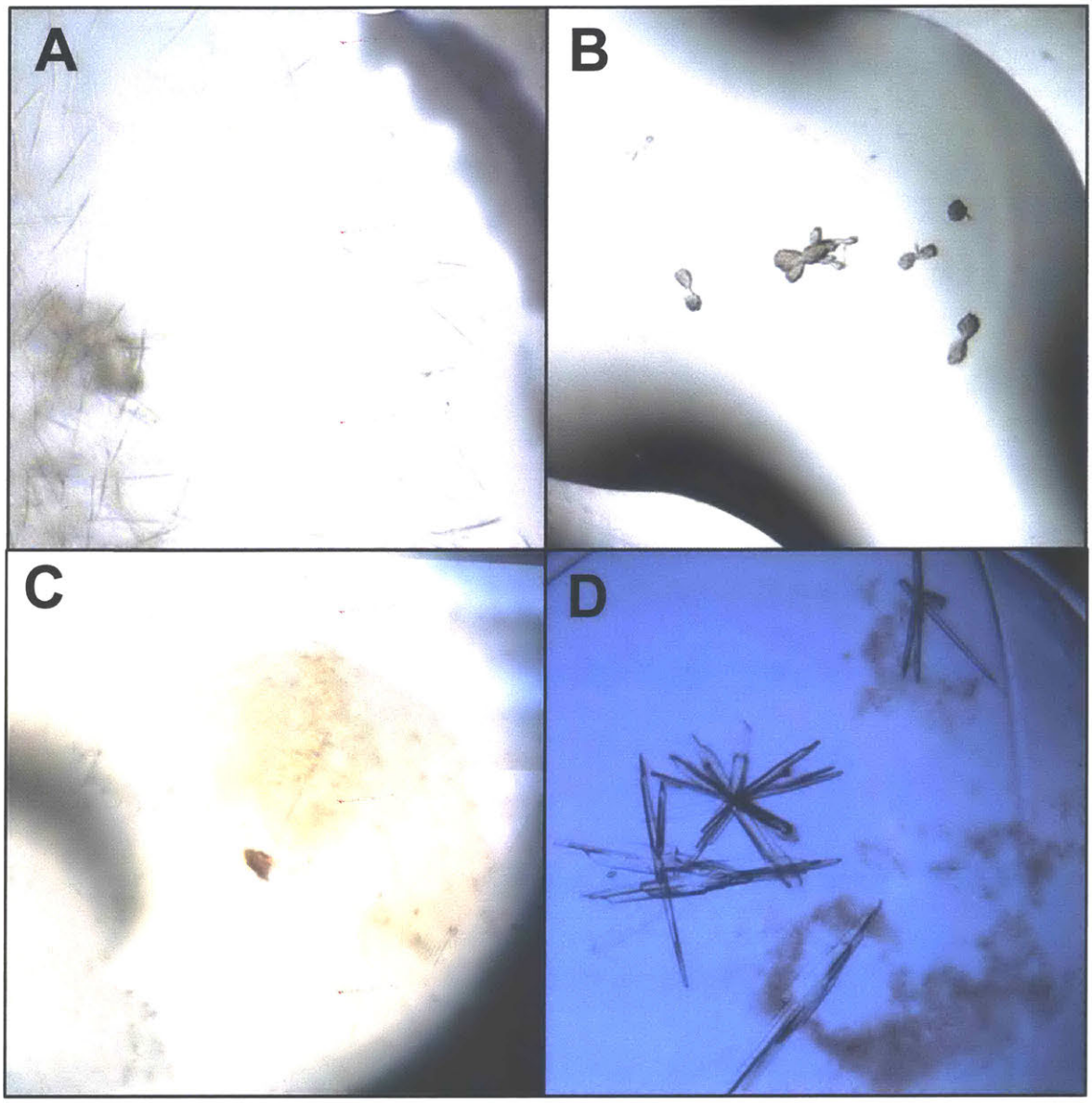



**Figure IV.5. Representative crystals from screens [Fe-S] cluster bound MFP.** All crystals shown were obtained in small-scale, sparse matrix screens. **(A)** Needles formed after about 6 days when anaerobically purified His-MFP at 5 mg/mL was set against 1.0% (w/v) PEG-3350, 1.0 M ammonium sulfate, and 0.1 M Bis-TRIS, pH 5.5. These crystals are not reproducible. **(B)** Haystack crystals formed after about 9 days when anaerobically purified His-MFP at 5 mg/mL was set against 0.1 M MES, pH 6.0 and 0.8 M ammonium sulfate. Haystacks were highly reproducible, even with different constructs, but only as haystacks. **(C)** A cluster of needles formed after about 5 days when reconstituted PfMFPΔ at 7.5 mg/mL was set against 0.2 M zinc acetate, 0.1 M sodium acetate, pH 4.5, and 10% (w/v) PEG 3,000. Crystals have been reproducible, but data has not been collected yet. **(D)** Rod clusters formed after about 1 day when reconstituted PfMFPΔ at 7.5 mg/mL was set against 0.1 M di-ammonium hydrogen phosphate, 50% (v/v) 2-methyl-2,4-pentanediol, and 0.1 M TRIS, pH 8.5. Attempts at reproducing these crystals have not been pursued due to the lack of color and the composition of the condition.

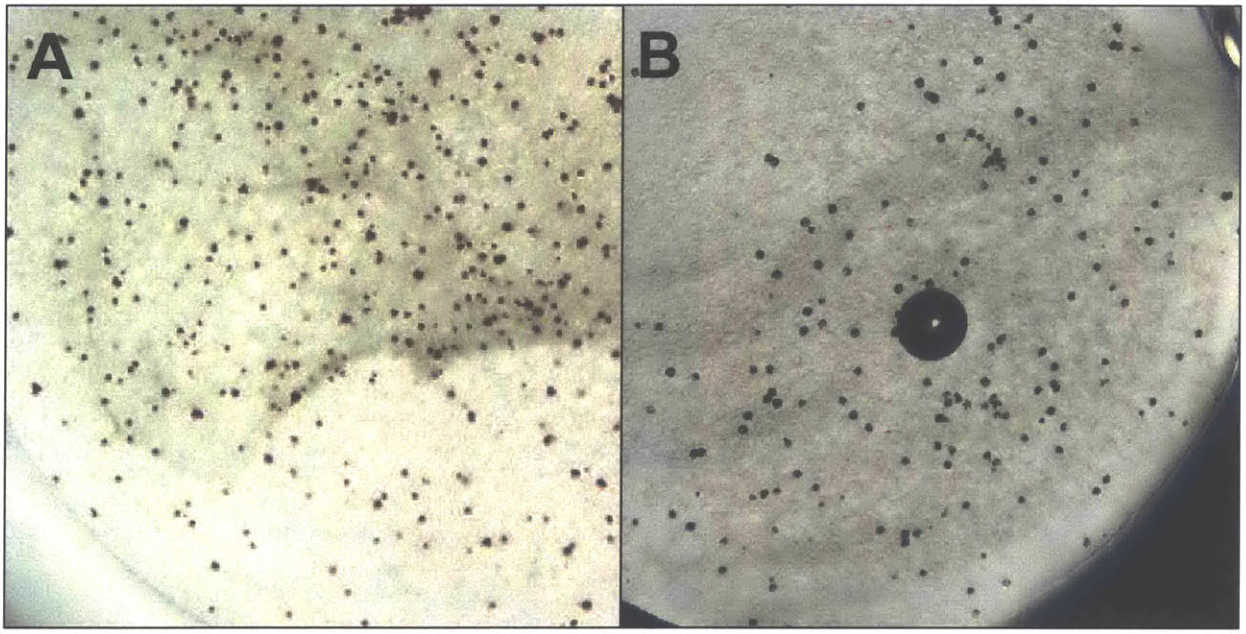



**Figure IV.6. Representative quasicrystals from screens with heme bound MFP.** The only promising conditions obtained are from screens of 4CA incubated with heme. **(A)** Quasicrystals form after about 3 months when 5.0 mg/mL 4CA incubated overnight with 5 molar equivalents of hemin was set against 1.0 M ammonium di-hydrogen phosphate and 0.1 M tri-sodium citrate, pH 5.6. This condition has not been reproducible. **(B)** Quasicrystals form overnight when 5 mg/mL 4CA incubated overnight with 1 molar equivalent of hemin was set against 0.03 M diethylene glycol, 0.03 M triethylene glycol, 0.03 M tetraethylene glycol, 0.03 M pentaethylene glycol, 0.1 M imidazole/MES buffer mix, pH 6.5, 12.5% (v/v) 2-methyl-2,4-propanediol, 12.5% (w/v) PEG-1000, and 12.5% (w/v) PEG-3350. The quasicrystals are reproducible, but no more advantageous morphology has been observed during optimization.

| Variant Name | Tag | Amino Acids of MFP | Residue substitution(s) |
|---|---|---|---|
| 4CA | N-terminal StrepII-tag | 22-217 | C68A, C70A, C93A, C216A |
| His-MFP | N-terminal His-tag | 22-217 | |
| His-MFP-SeMet | N-terminal His-tag | 22-217 | Methionine to selenomethionine |
| PfMFP | N-terminal StrepII-tag | 22-217 | |
| PfMFPΔ | N-terminal StrepII-tag | 33-217 | |
| PfMFPΔ-SeMet | N-terminal StrepII-tag | 33-217 | Methionine to selenomethionine |
| MBP-MFP | N-terminal MBP tag, C-terminal His-tag | 22-217 | |
| C68A/C93A | N-terminal StrepII-tag | 22-217 | C68A, C93A |

**Table IV.1. Many variants of MFP have been used for attempted crystallization.** The most promising variant is the C68A/C93A; however, it has been the least studied for crystallography, as it was obtained very recently.

| Apo-MFP Variant | Buffer | Additional Treatment | [protein] | Screens | Incubation Temperature | Drop Ratio |
|---|---|---|---|---|---|---|
| 4CA | 50 mM HEPES, pH 8.0, 100 mM NaCl | incubated with 1 molar equivalent of heme overnight | 2.5 mg/mL | AmSO4 | 18 °C | 150 nL:150 nL |
| | | | | ComPAS | 18 °C | |
| | | | | Index | 18 °C | |
| | | | | JCSG+ | 18 °C | |
| | | | | Morpheus | 18 °C | |
| | | | | Morpheus II | 18 °C | |
| | | | | PACT | 18 °C | |
| | | | | PEGRx | 18 °C | |
| | | Preparatory SEC | 4.5 mg/mL | ComPAS | 18 °C | |
| | | | | | 6 °C | |
| | | | | JCSG+ | 18 °C | |
| | | | | | 6 °C | |
| | | Preparatory SEC, incubated with equimolar hemin | 4.5 mg/mL | ComPAS | 18 °C | |
| | | | | | 6 °C | |
| | | | | JCSG+ | 18 °C | |
| | | | | | 6 °C | |
| | | Preparatory SEC | 5.0 mg/mL | ComPAS | 18 °C | |
| | | | | | 6 °C | |
| | | | | Index | 18 °C | |
| | | | | | 6 °C | |
| | | | | JCSG+ | 18 °C | |
| | | | | | 6 °C | |
| | | Preparatory SEC, incubated with 1x hemin | 5.0 mg/mL | ComPAS | 18 °C | |
| | | | | | 6 °C | |
| | | | | Index | 18 °C | |
| | | | | | 6 °C | |
| | | | | JCSG+ | 18 °C | |
| | | | | | 6 °C | |
| | | Preparatory SEC, incubated with 1x hemin over night, seeded with 4CA seed stock | 5.0 mg/mL | ComPAS | 18 °C | 250 nL well, 200 nL protein, 50 nL seed stock |
| | | | | Morpheus II | 18 °C | |

| Apo-MFP Variant | Buffer | Additional Treatment | [protein] | Screens | Incubation Temperature | Drop Ratio |
|---|---|---|---|---|---|---|
| 4CA | 50 mM HEPES, pH 8.0, 100 mM NaCl | Preparatory SEC, incubated with 5x hemin | 5.0 mg/mL | ComPAS | 18 °C | 150 nL:150 nL |
| | | | | | 6 °C | |
| | | | | Index | 18 °C | |
| | | | | | 6 °C | |
| | | | | JCSG+ | 18 °C | |
| | | | | | 6 °C | |
| | | incubated with 1 molar equivalent of heme overnight | 5.0 mg/mL | AmSO4 | 18 °C | |
| | | | | ComPAS | 18 °C | |
| | | | | Index | 18 °C | |
| | | | | JCSG+ | 18 °C | |
| | | | | Morpheus | 18 °C | |
| | | | | Morpheus II | 18 °C | |
| | | | | PACT | 18 °C | |
| | | | | PEGRx | 18 °C | |
| | | | 7.5 mg/mL | 0.1 M Morpheus Buffer system 1, pH 6.5, 42.5% (v/v) Morpheus Precipitant Mix 4 + additive screen | 18 °C | |
| | | | | 0.1 M Morpheus Buffer system 1, pH 7.5, 42.5% (v/v) Morpheus Precipitant Mix 4 + additive screen | 18 °C | |

| Apo-MFP Variant | Buffer | Additional Treatment | [protein] | Screens | Incubation Temperature | Drop Ratio |
|---|---|---|---|---|---|---|
| 4CA | 50 mM HEPES, pH 8.0, 100 mM NaCl | incubated with 1 molar equivalent of heme overnight | 7.5 mg/mL | 0.1 M Morpheus Buffer system 1, pH 7.5, 42.5% (v/v) Morpheus Precipitant Mix 4 + detergent screen | 18 °C | 150 nL:150 nL |
| | | | | AmSO4 | 18 °C | |
| | | | | ComPAS | 18 °C | |
| | | | | Index | 18 °C | |
| | | | | JCSG+ | 18 °C | |
| | | | | Morpheus | 18 °C | |
| | | | | Morpheus II | 18 °C | |
| | | | | PACT | 18 °C | |
| | | | | PEGRx | 18 °C | |
| | | | 8.5 mg/L | 0.1 M Morpheus Buffer system 1, pH 6.5, 42.5% (v/v) Morpheus Precipitant Mix 4 + additive screen | 18 °C | |
| | | | | 0.1 M Morpheus Buffer system 1, pH 7.5, 42.5% (v/v) Morpheus Precipitant Mix 4 + additive screen | 18 °C | |

| Apo-MFP Variant | Buffer | Additional Treatment | [protein] | Screens | Incubation Temperature | Drop Ratio |
|---|---|---|---|---|---|---|
| 4CA | 50 mM HEPES, pH 8.0, 100 mM NaCl | incubated with 1 molar equivalent of heme overnight | 8.5 mg/L | 0.1 M Morpheus Buffer system 1, pH 7.5, 42.5% (v/v) Morpheus Precipitant Mix 4 + detergent screen | 18 °C | 150 nL:150 nL |
| His-MFP | 50 mM Tris, pH 8.0 | Buffer Exchanged | 1.0 mg/mL | ComPAS | 18 °C | |
| | | | | Index | 18 °C | |
| | | | | JCSG+ | 18 °C | |
| | | | 1.2 mg/mL | ComPAS | 18 °C | |
| | | | | Index | 18 °C | |
| | | | | JCSG+ | 18 °C | |
| | | | 1.5 mg/mL | ComPAS | 18 °C | |
| | | | | Index | 18 °C | |
| | | | | JCSG+ | 18 °C | |
| | 50 mM HEPES, pH 8.2, 100 mM NaCl | Buffer Exchanged | 10 mg/mL | Index | 6 °C | |
| | | | | JCSG+ | 6 °C | |
| | | | | MCSG1 | 6 °C | |
| | 50 mM Tris, pH 8.0 | Buffer Exchanged | 10 mg/mL | ComPAS | 18 °C | |
| | | | | Index | 18 °C | |
| | | | | JCSG+ | 18 °C | |
| | 50 mM Tris, pH 8.0, 124.4 mM NaCl, 61 mM imidazole | Diluted fraction from affinity column | 10 mg/mL | ComPAS | 18 °C | |
| | | | | Index | 18 °C | |
| | | | | JCSG+ | 18 °C | |
| | 100 mM HEPES, pH 8.0, 500 mM NaCl | Buffer Exchanged | 15 mg/mL | Index | 18 °C | |
| | | | | JCSG+ | 18 °C | |
| | | | | MCSG1 | 18 °C | |
| | 50 mM HEPES, pH 8.2, 100 mM NaCl | Buffer Exchanged | 15 mg/mL | Index | 6 °C | |
| | | | | JCSG+ | 6 °C | |
| | | | | MCSG1 | 6 °C | |

| Apo-MFP Variant | Buffer | Additional Treatment | [protein] | Screens | Incubation Temperature | Drop Ratio |
|---|---|---|---|---|---|---|
| His-MFP | 50 mM Tris, pH 8.0 | Buffer Exchanged | 15 mg/mL | ComPAS | 18 °C | 150 nL:150 nL |
| | | | | Index | 18 °C | |
| | | | | JCSG+ | 18 °C | |
| | 50 mM Tris, pH 8.0, 136.5 mM NaCl, 91.35 mM imidazole | Diluted fraction from affinity column | 15 mg/mL | ComPAS | 18 °C | |
| | | | | Index | 18 °C | |
| | | | | JCSG+ | 18 °C | |
| | 50 mM HEPES, pH 8.0, 100 mM NaCl | Buffer Exchanged | 20 mg/mL | ComPAS | 18 °C | |
| | | | | | 6 °C | |
| | | | | Index | 18 °C | |
| | | | | JCSG+ | 18 °C | |
| | | | | MembFac | 6 °C | |
| | | | | PEGRx | 18 °C | |
| | | | | ProComp | 18 °C | |
| | | | | | 6 °C | |
| | | Preparatory SEC | 20 mg/mL | ComPAS | 18 °C | |
| | | | | | 6 °C | |
| | | | | Index | 18 °C | |
| | | | | | 6 °C | |
| | | | | JCSG+ | 18 °C | |
| | | | | | 6 °C | |
| | | | | MembFac | 6 °C | |
| | | | | PEGRx | 18 °C | |
| | | | | | 6 °C | |
| | | | | ProComp | 6 °C | |
| | | Preparatory SEC, aerobic cluster bound | 20 mg/mL | ComPAS | 18 °C | |
| | | | | JCSG+ | 18 °C | |
| | | | | PEG/Ion | 18 °C | |
| | 50 mM HEPES, pH 8.0, 50 mM NaCl | Buffer Exchanged | 20 mg/mL | ComPAS | 18 °C | |
| | | | | Index | 18 °C | |
| | | | | JCSG+ | 18 °C | |
| | | | | PEGRx | 18 °C | |
| | | | | ProComp | 18 °C | |

| Apo-MFP Variant | Buffer | Additional Treatment | [protein] | Screens | Incubation Temperature | Drop Ratio |
|---|---|---|---|---|---|---|
| His-MFP | 50 mM HEPES, pH 8.0, 100 mM NaCl | Buffer Exchanged | 40 mg/mL | ComPAS | 18 °C | 150 nL:150 nL |
| | | | | Index | 18 °C | |
| | | | | JCSG+ | 18 °C | |
| | 50 mM HEPES, pH 8.0, 100 mM NaCl | Buffer Exchanged | 40 mg/mL | PEGRx | 18 °C | |
| | | | | ProComp | 18 °C | |
| His-MFP-SeMet | 45 mM HEPES, pH 8.0, 90 mM NaCl, 10% (v/v) DMSO | Preparatory SEC, co-crystallized with compound 1 | 20 mg/mL | ComPAS | 18 °C | |
| | | | | JCSG+ | 18 °C | |
| | | | | PEG/Ion | 18 °C | |
| | | Preparatory SEC, co-crystallized with compound 4 | 20 mg/mL | ComPAS | 18 °C | |
| | | | | JCSG+ | 18 °C | |
| | | | | PEG/Ion | 18 °C | |
| | 50 mM HEPES, pH 8.0, 100 mM NaCl | Preparatory SEC | 5 mg/mL | Index | 6 °C | |
| | | | | JCSG+ | 6 °C | |
| | | | | Morph II | 6 °C | |
| | 45 mM HEPES, pH 8.0, 90 mM NaCl, 10% (v/v) DMSO | Preparatory SEC, co-crystallized with compound 1 | 5 mg/mL | Index | 6 °C | |
| | | | | JCSG+ | 6 °C | |
| | | | | Morph II | 6 °C | |
| | | Preparatory SEC, co-crystallized with compound 4 | 5 mg/mL | Index | 6 °C | |
| | | | | JCSG+ | 6 °C | |
| | | | | Morph II | 6 °C | |
| | 50 mM HEPES, pH 8.0, 200 mM NaCl | Preparatory SEC | 5 mg/mL | JCSG+ | 18 °C | |
| | | | | | 6 °C | |
| | | Preparatory SEC, incubated with equimolar hemin | 5 mg/mL | JCSG+ | 18 °C | |
| | | | | | 6 °C | |
| PfMFP | 50 mM HEPES, pH 8.0, 200 mM NaCl | Preparatory SEC, incubated with equimolar hemin | 1.41 mg/mL | JCSG+ | 18 °C | |
| | | | | | 6 °C | |
| | | Preparatory SEC | 1.43 mg/mL | JCSG+ | 18 °C | |
| | | | | | 6 °C | |
| PfMFPΔ | 50 mM HEPES, pH 8.0, 200 mM NaCl | Preparatory SEC, incubated with equimolar hemin | 1.58 mg/mL | JCSG+ | 18 °C | |
| | | | | | 6 °C | |
| | | Preparatory SEC | 1.60 mg/mL | JCSG+ | 18 °C | |
| | | | | | 6 °C | |

| Apo-MFP Variant | Buffer | Additional Treatment | [protein] | Screens | Incubation Temperature | Drop Ratio |
|---|---|---|---|---|---|---|
| PfMFPΔ | 50 mM HEPES, pH 8.0, 100 mM NaCl | oxidized with 10 mM potassium ferricyanide and dialyzed against buffer to remove oxidant | 10 mg/mL | ComPAS | 18 °C | 150 nL:150 nL |
| | | | | Index | 18 °C | |
| | | | | Morpheus II | 18 °C | |
| | | oxidized with 10 mM potassium ferricyanide and dialyzed against buffer to remove oxidant + 1 equivalent of heme | 10 mg/mL | ComPAS | 18 °C | |
| | | | | Index | 18 °C | |
| | | | | Morpheus II | 18 °C | |
| | | Preparatory SEC | 4.5 mg/mL | ComPAS | 18 °C | |
| | | | | | 6 °C | |
| | | | | Index | 18 °C | |
| | | | | | 6 °C | |
| | | | | JCSG+ | 18 °C | |
| | | | | | 6 °C | |
| | | Preparatory SEC, incubated with 1x hemin | 4.5 mg/mL | ComPAS | 18 °C | |
| | | | | | 6 °C | |
| | | | | Index | 18 °C | |
| | | | | | 6 °C | |
| | | | | JCSG+ | 18 °C | |
| | | | | | 6 °C | |
| | | Preparatory SEC, incubated with 4.7x hemin | 4.5 mg/mL | ComPAS | 18 °C | |
| | | | | | 6 °C | |
| | | | | Index | 18 °C | |
| | | | | | 6 °C | |
| | | | | JCSG+ | 18 °C | |
| | | | | | 6 °C | |
| | | Preparatory SEC | 5 mg/mL | AmSO4 | 6 °C | |
| | | | | ComPAS | 6 °C | |
| | | | | JCSG+ | 6 °C | |

| Apo-MFP Variant | Buffer | Additional Treatment | [protein] | Screens | Incubation Temperature | Drop Ratio |
|---|---|---|---|---|---|---|
| PfMFPΔ | 50 mM HEPES, pH 8.0, 100 mM NaCl | Preparatory SEC, incubated with 5x Sypro Orange | 5 mg/mL | AmSO4 | 6 °C | 150 nL:150 nL |
| | | | | ComPAS | 6 °C | |
| | | | | JCSG+ | 6 °C | |
| | | Preparatory SEC, seeded with fresh isopropanol seed stock | 5 mg/mL | JCSG+ | 18 °C | 250 nL well, 200 nL protein, 50 nL seed stock |
| | | | | | 6 °C | |
| | | | | Morpheus | 6 °C | |
| | | | | Morph II | 6 °C | |
| | | Preparatory SEC, seeded with frozen ethanol seed stock | 5 mg/mL | JCSG+ | 18 °C | |
| | | | | | 6 °C | |
| | | | | Morpheus | 6 °C | |
| | | | | Morph II | 6 °C | |
| | | Preparatory SEC, seeded with frozen isopropanol seed stock | 5 mg/mL | JCSG+ | 18 °C | |
| | | | | | 6 °C | |
| | | | | Morpheus | 6 °C | |
| | | | | Morph II | 6 °C | |
| | | re-purified after incubation O/N with heme | 5 mg/mL | JCSG+ | 6 °C | 150 nL:150 nL |
| | 50 mM HEPES, pH 8.0, 200 mM NaCl | Preparatory SEC | 5 mg/mL | AmSO4 | 6 °C | |
| | | | | ComPAS | 6 °C | |
| | | | | JCSG+ | 6 °C | |
| PfMFPΔ-SeMet | 50 mM HEPES, pH 8.0, 100 mM NaCl | Buffer Exchanged | 5 mg/mL | ComPAS | 6 °C | |
| | | | | JCSG+ | 6 °C | |
| MBP-MFP | 50 mM HEPES, pH 8.0, 100 mM NaCl | Preparatory SEC | 15 mg/mL | AmSO4 | 18 °C | |
| | | | | | 6 °C | |
| | | | | ComPAS | 18 °C | |
| | | | | | 6 °C | |
| | | | | Index | 18 °C | |
| | | | | | 6 °C | |
| | | | | PEGRx | 18 °C | |
| | | | | | 6 °C | |
| C68A/ C93A | 50 mM HEPES, pH 8.0, 100 mM NaCl | Preparatory SEC | 5 mg/mL | ComPAS | 18 °C | |
| | | | | | 6 °C | |

| Apo-MFP Variant | Buffer | Additional Treatment | [protein] | Screens | Incubation Temperature | Drop Ratio |
|---|---|---|---|---|---|---|
| C68A/ C93A | 50 mM HEPES, pH 8.0, 100 mM NaCl | Preparatory SEC | 5 mg/mL | JCSG+ | 18 °C | 150 nL:150 nL |
| | | | | | 6 °C | |
| | | | | PACT | 18 °C | |
| | | | | | 6 °C | |
| | | | | PEGRx | 18 °C | |
| | | | | | 6 °C | |

**Table IV.2. Apo-MFP has been screened against many aerobic conditions in an attempt to obtain data-quality crystals.** Over 18,000 conditions were screened with very few leads identified. Variants are described in **Table IV.1**. Note that a single sample (in gray boxes above) was aerobically purified and contained cluster. This sample is the only sample in this table that is not apo-MFP. Compound 1 and compound 4 are proprietary compounds identified in a small molecule microarray screen and were generously gifted by the Niles Lab at MIT.

| Data Set | N5900_748_13 | MFP_S_13 | N5900_748_14 | N5900_748_15 | N5900_748_16A |
|---|---|---|---|---|---|
| **MFP variant** | His-MFP | | His-MFP | His-MFP | His-MFP |
| **Protein concentration** | 20 mg/mL | | 20 mg/mL | 20 mg/mL | 20 mg/mL |
| **Protein Buffer** | 50 mM HEPES, pH 8.0, 100 mM NaCl | | 50 mM HEPES, pH 8.0, 100 mM NaCl | 50 mM HEPES, pH 8.0, 100 mM NaCl | 50 mM HEPES, pH 8.0, 100 mM NaCl |
| **Well Condition** | 0.1 M HEPES, pH 8.0, 0.15 M $MgCl_2$, 25% isopropanol | | 0.1 M HEPES, pH 8.0, 0.15 M $MgCl_2$, 25% isopropanol | 0.1 M HEPES, pH 8.0, 0.15 M $MgCl_2$, 25% isopropanol | 0.1 M HEPES, pH 8.2, 0.15 M $MgCl_2$, 27% isopropanol |
| **Drop Size** | 2 uL | | 2 uL | 2 uL | 2 uL |
| **Seeding** | | | | | |
| **Heavy atom soak** | | | | | |
| **Cryoprotectant** | ethylene glycol, incubated overnight | | ethylene glycol, incubated overnight | ethylene glycol, incubated overnight | MPD, incubated overnight |

| Data Set | MFP_A3 | MFP_A6 | MFP_A10 | MFP_A11 |
|---|---|---|---|---|
| **MFP variant** | His-MFP | His-MFP-SeMet | His-MFP-SeMet | His-MFP |
| **Protein concentration** | 20 mg/mL | 20 mg/mL | 20 mg/mL | 20 mg/mL |
| **Protein Buffer** | 50 mM HEPES, pH 8.0, 100 mM NaCl | 50 mM HEPES, pH 8.0, 100 mM NaCl | 50 mM HEPES, pH 8.0, 100 mM NaCl | 50 mM HEPES, pH 8.0, 100 mM NaCl |
| **Well Condition** | 0.1 M HEPES, pH 8.0, 0.15 M $MgCl_2$, 25% isopropanol | 0.1 M HEPES, pH 8.0, 0.15 M $MgCl_2$, 25% isopropanol | 0.1 M HEPES, pH 8.0, 0.15 M $MgCl_2$, 25% isopropanol | 0.1 M HEPES, pH 8.0, 0.15 M $MgCl_2$, 25% isopropanol |
| **Drop Size** | 2 uL | 2 uL | 2 uL | 2 uL |
| **Seeding** | | | | |
| **Heavy atom soak** | | | | |
| **Cryoprotectant** | ethylene glycol, incubated for 5 minutes | ethylene glycol, incubated for 1 hour | ethylene glycol, incubated for 4 hours | ethylene glycol, incubated for 4 hours |

| Data Set | MFP_A14 | MFP_751_3 | MFP_751_4 | MFP_751_7 | MFP_751_7b |
|---|---|---|---|---|---|
| **MFP variant** | His-MFP-SeMet | PfMFPΔ | PfMFPΔ | PfMFPΔ | |
| **Protein concentration** | 20 mg/mL | 5 mg/mL | 5 mg/mL | 2.5 mg/mL | |
| **Protein Buffer** | 50 mM HEPES, pH 8.0, 100 mM NaCl | 50 mM HEPES, pH 8.0, 100 mM NaCl | 50 mM HEPES, pH 8.0, 100 mM NaCl | 50 mM HEPES, pH 8.0, 100 mM NaCl | |
| **Well Condition** | 0.1 M HEPES, pH 8.0, 0.15 M $MgCl_2$, 25% isopropanol | 0.1 M HEPES, pH 8.0, 0.11 M $MgCl_2$, 25% isopropanol with 100 uL of paraffin oil covering well | 0.1 M HEPES, pH 8.0, 0.11 M $MgCl_2$, 25% isopropanol with 100 uL of paraffin oil covering well | 0.1 M HEPES, pH 8.0, 0.17 M $MgCl_2$, 25% isopropanol | |
| **Drop Size** | 2 uL | 2 uL | 2 uL | 2 uL | |
| **Seeding** | | | | 1 uL protein to 1 uL of 1:10,000 dilution of fresh isopropanol seed stock | |
| **Heavy atom soak** | | | | | |
| **Cryoprotectant** | ethylene glycol, incubated overnight | ethylene glycol, incubated overnight | ethylene glycol, incubated overnight | ethylene glycol, incubated overnight | |

| Data Set | 1811_MFP_3 | 1811_MFP_03Se | 1811_MFP_5 | 1811_MFP_12 | 1811_MFP_13 |
|---|---|---|---|---|---|
| **MFP variant** | His-MFP-SeMet | | His-MFP-SeMet | His-MFP | His-MFP |
| **Protein concentration** | 10 mg/mL | | 10 mg/mL | 25 mg/mL | 25 mg/mL |
| **Protein Buffer** | 50 mM HEPES, pH 8.0, 100 mM NaCl, 1 mM DTT | | 50 mM HEPES, pH 8.0, 100 mM NaCl, 1 mM DTT | 50 mM HEPES, pH 8.0, 100 mM NaCl | 50 mM HEPES, pH 8.0, 100 mM NaCl |
| **Well Condition** | 0.1 M Tris, pH 8.0, 0.18 M $MgCl_2$, 25% isopropanol, 1% MPD | | 0.1 M Tris, pH 8.0, 0.18 M $MgCl_2$, 25% isopropanol, 1% MPD | 0.1 M HEPES, pH 8.2, 0.15 M $MgCl_2$, 25% isopropanol | 0.1 M HEPES, pH 8.2, 0.15 M $MgCl_2$, 25% isopropanol |
| **Drop Size** | 2 uL | | 2 uL | 4 uL | 4 uL |
| **Seeding** | | | | | |
| **Heavy atom soak** | | | | | |
| **Cryoprotectant** | glucose, incubated overnight | | glucose, incubated overnight | glucose, incubated overnight | glucose, incubated overnight |

| Data Set | 1902_MFP_04Fe | 1902_MFP04Se | 1902_MFP_06 | 1902_MFP_06_FeB | 1902_MFP_06_FeC | 1902_MFP_06b |
|---|---|---|---|---|---|---|
| **MFP variant** | His-MFP | | His-MFP | | | His-MFP-SeMet |
| **Protein concentration** | 20 mg/mL | | 20 mg/mL | | | 10 mg/mL |
| **Protein Buffer** | 50 mM HEPES, pH 8.0, 100 mM NaCl | | 50 mM HEPES, pH 8.0, 100 mM NaCl | | | 50 mM HEPES, pH 8.0, 100 mM NaCl, 1 mM DTT |
| **Well Condition** | 0.1 M HEPES, pH 7.5, 0.13 M $MgCl_2$, 25% isopropanol | | 0.1 M HEPES, pH 7.5, 0.13 M $MgCl_2$, 25% isopropanol | | | 0.1 M Tris, pH 8.0, 0.18 M $MgCl_2$, 25% isopropanol, 1% MPD |
| **Drop Size** | 2 uL | | 2 uL | | | 2 uL |
| **Seeding** | | | | | | |
| **Heavy atom soak** | ethyl mercuric phosphate | | ethyl mercuric phosphate | | | |
| **Cryo-protectant** | ethylene glycol, incubated for 2 hours | | ethylene glycol, incubated for 2 hours | | | glucose, incubated overnight |

| Data Set | CLD_956_7 | CLD_956_10 |
|---|---|---|
| **MFP variant** | PfMFPΔ | PfMFPΔ |
| **Protein concentration** | 8 mg/mL | 8 mg/mL |
| **Protein Buffer** | 50 mM HEPES, pH 8.0, 300 mM NaCl | 50 mM HEPES, pH 8.0, 300 mM NaCl |
| **Well Condition** | 0.1 M HEPES, pH 7.4, 0.11 M MgCl2, 25% isopropanol | 0.1 M HEPES, pH 7.4, 0.11 M MgCl2, 25% isopropanol |
| **Drop Size** | 2 uL | 2 uL |
| **Seeding** | | |
| **Heavy atom soak** | I3C | I3C |
| **Cryoprotectant** | ethylene glycol, incubated for 5 minutes | ethylene glycol, incubated for 5 minutes |

**Table IV.3. Crystallization conditions of crystals for which continuous scan data sets have been collected.** Methods for seeding, heavy atom soaks, and cryoprotectants are laid out in the experimental section. Except when noted, all drops are half protein and half well, by volume.

| | P622 | P6 | P312 | P321 | P3 | C222 | P2 | C2 | P1 |
|---|---|---|---|---|---|---|---|---|---|
| Beamline | 24-ID-C | | | | | | | | |
| Wavelength (Å) | 0.979145 Å | | | | | | | | |
| Space group | P622 | P6 | P312 | P321 | P3 | C222 | P2 | C2 | P1 |
| Cell dimensions (Å) | 109.4, 109,4, 64.3, 90, 90,120 | 109.4, 109.4, 64.3, 90, 90, 120 | 109.4, 109.4, 64.3, 90, 90, 120 | 109.4, 109.4, 64.3, 90, 90, 120 | 109.4, 109.4, 64.3, 90, 90, 120 | 109.4, 189.6, 64.3, 90, 90, 90 | 109.4, 64.3, 109.4, 90, 120, 90 | 189.6, 109.3, 64.3, 90, 89.98, 90 | 64.3, 109.4, 109.3, 119.99, 90.11, 89.94 |
| Resolution (Å) | 2.51 | 2.51 | 2.51 | 2.51 | 2.51 | 2.45 | 2.60 | 2.60 | 2.60 |
| Unique reflections | 8166 | 15106 | 15123 | 15438 | 29274 | 24399 | 40216 | 39862 | 72441 |
| Rejcted reflections (%) | 0.03 | 0.03 | 0.03 | 0.04 | 0.04 | 0.07 | 0.11 | 0.05 | 0.08 |
| Completeness | 99.51% | 99.48% | 99.51% | 99.53% | 99.30% | 97.52% | 97.78% | 98.38% | 91.74% |
| Redundancy | 31.9 | 17.4 | 17.4 | 17.0 | 9.0 | 11.1 | 5.9 | 6.2 | 3.4 |
| $R_{sym}$ | 0.185 | 0.184 | 0.184 | 0.183 | 0.180 | 0.186 | 0.166 | 0.117 | 0.088 |
| I/σ(I) | 20.1 | 15.8 | 15.7 | 15.6 | 11.9 | 12.7 | 9.4 | 13.6 | 12.4 |

**Table IV.4. Data statistics from a single His-MFP crystal (N5900_748_13) processed in different space groups.** Data analysis through Phenix Xtriage (2) indicates twinning of the His-MFP crystals, and processing of data in these different space groups does not clearly reveal which space group is correct, as all statistics seem reasonable. Data were processed without screw axes; however, the addition of screw axes did not change the statistics reported here.

| Data Set | N5900_748_13 | MFP_S_13 | N5900_748_14 | N5900_748_15 | N5900_748_16A |
|---|---|---|---|---|---|
| **Beam Line** | APS 24-ID-C | APS 24-ID-C | APS 24-ID-C | APS 24-ID-C | APS 24-ID-C |
| **Frames** | 1-1800 | 1-1440 | 1-1800 | 1-1800 | 1-559 |
| **Oscillation** | 0.20° | 0.25° | 0.20° | 0.20° | 0.20° |
| **Space Group** | P622 | P622 | P622 | P622 | C2 |
| **Cell Dimensions (Å)[1]** | a=109.5, b=109.5, c=64.4 | a=109.66 b=109.66 c=64.38 | a=109.0, b=109.0, c=64.1 | a=109.8, b=109.8, c=64.3 | a=108.3, b=187.6, c=64.3 |
| **Angles[1]** | gamma = 120 | gamma = 120 | gamma = 120 | gamma = 120 | beta = 90.32 |
| **Wavelength (Å)** | 0.9791 | 1.7969 | 0.9791 | 0.9791 | 0.9791 |
| **Resolution (Å)[1]** | 94.78 - 2.20 (2.26 - 2.20) | 94.97 - 2.48 (2.58 - 2.48) | 94.41 - 2.41 (2.50 - 2.41) | 95.04 - 2.46 (2.56 - 2.46) | 93.82 - 2.88 (3.03 - 2.88) |
| **# unique reflections[1]** | 12040 (981) | 8507 (921) | 9033 (815) | 8676 (919) | 22803 (3145) |
| **Completeness (%)[1]** | 99.7 (96.6) | 99.8 (98.5) | 98.7 (87.9) | 99.4 (95.0) | 79.1 (74.8) |
| **Redundancy[1]** | 37.1 (29.6) | 35.6 (33.9) | 36.1 (28.9) | 36.3 (28.6) | 2.6 (2.5) |
| **$\langle I/\sigma I \rangle$[1]** | 11.8 (0.6) | 19.1 (0.8) | 8.5 (0.9) | 17.9 (0.8) | 6.0 (1.6) |
| **$R_{merge}$[1]** | 0.193 (3.705) | 0.131 (4.567) | 0.319 (1.723) | 0.146 (5.052) | 0.138 (0.502) |
| **$R_{merge}$ (anom)[1]** | | 0.130 (4.510) | | | |
| **$CC_{1/2}$[1]** | 0.995 (0.517) | 0.998 (0.705) | 0.986 (0.858) | 0.998 (0.605) | 0.970 (0.745) |
| **Twinning Indicated[2]** | N/A | N/A | N/A | N/A | Yes |
| **Multivariant Z-score L test[2]** | 17.306 | 16.773 | 15.156 | 17.304 | 6.763 |
| **Anomalous Signal[1,2]** | | No | | | |
| **Other errors[2]** | | | anisotropic | | anisotropic, low completeness |

| Data Set | MFP_A3 | MFP_A6 | MFP_A10 | MFP_A11 | MFP_A14 |
|---|---|---|---|---|---|
| **Beam Line** | APS 24-ID-C | APS 24-ID-C | APS 24-ID-C | APS 24-ID-C | APS 24-ID-C |
| **Frames** | 1-1800 | 1-1800 | 1-1215 | 1-1440 | 1-1200 |
| **Oscillation** | 0.20° | 0.20° | 0.20° | 0.25° | 0.30° |
| **Space Group** | P622 | P622 | P622 | P622 | P622 |
| **Cell Dimensions (Å)[1]** | a=108.1, b=108.1, c=64.2 | a=109.6, b=109.6, c=64.1 | a=109.8, b=109.8, c=64.2 | a=108.8, b=108.8, c=64.2 | a=109.1, b=109.1, c=64.1 |
| **Angles[1]** | gamma = 120 | gamma = 120 | gamma = 120 | gamma = 120 | gamma = 120 |
| **Wavelength (Å)** | 0.9791 | 0.9791 | 0.9791 | 0.9791 | 0.9791 |
| **Resolution (Å)[1]** | 93.57 - 2.08 (2.13 - 2.08) | 94.91 - 2.72 (2.86 - 2.72) | 99.70 - 2.93 (3.11 - 2.93) | 99.60 - 1.89 (1.93 - 1.89) | 94.49 - 2.95 (3.12 - 2.95) |
| **# unique reflections[1]** | 13728 (1031) | 6468 (812) | 5246 (811) | 18504 (1082) | 5115 (793) |
| **Completeness (%)[1]** | 99.4 (98.5) | 99.7 (97.8) | 99.7 (98.1) | 99.6 (94.3) | 99.9 (99.4) |
| **Redundancy[1]** | 46.3 (46.0) | 44.6 (43.0) | 30.9 (29.9) | 37.9 (37.6) | 36.1 (37.8) |
| **$\langle I/\sigma I \rangle$[1]** | 20.8 (1.9) | 18.1 (1.5) | 12.2 (1.5) | 22.0 (0.7) | 10.9 (1.1) |
| **$R_{merge}$[1]** | 0.211 (2.300) | 0.285 (2.739) | 0.209 (1.689) | 0.094 (5.707) | 0.279 (3.124) |
| **$R_{merge}$ (anom)[1]** | | | | | |
| **$CC_{1/2}$[1]** | 0.999 (0.830) | 0.999 (0.839) | 0.999 (0.869) | 0.997 (0.643) | 0.999 (0.830) |
| **Twinning Indicated[2]** | N/A | N/A | N/A | N/A | N/A |
| **Multivariant Z-score L test[2]** | 17.261 | 15.859 | 16.752 | 16.273 | 14.729 |
| **Anomalous Signal[1,2]** | | | | | |
| **Other errors[2]** | | | | | |

| Data Set | MFP_751_3 | MFP_751_4 | MFP_751_7 | MFP_751_7b |
|---|---|---|---|---|
| **Beam Line** | APS 24-ID-C | APS 24-ID-C | APS 24-ID-C | APS 24-ID-C |
| **Frames** | 1-1800 | 1-1200 | 1-1200 | 1-1800 |
| **Oscillation** | 0.20° | 0.30° | 0.30° | 0.20° |
| **Space Group** | P622 | P622 | P622 | P622 |
| **Cell Dimensions (Å)[1]** | a=108.7, b=108.7, c=62.7 | a=112.6, b=112.6, 64.6 | a=110.7, b=110.7, c=64.1 | a=110.5, b=110.5, c=63.7 |
| **Angles[1]** | gamma=120 | gamma=120 | gamma=120 | gamma=120 |
| **Wavelength (Å)** | 0.9791 | 0.9791 | 0.9791 | 0.9791 |
| **Resolution (Å)[1]** | 94.16 - 3.58 (3.92 - 3.58) | 97.55 - 3.94 (4.41 - 3.94) | 95.84 - 3.35 (3.61 - 3.35) | 95.70 - 3.58 (3.92 -3.58) |
| **# unique reflections[1]** | 2840 (640) | 2316 (579) | 3590 (660) | 2974 (677) |
| **Completeness (%)[1]** | 99.3 (97.1) | 97.0 (89.7) | 98.1 (91.0) | 99.8 (99.3) |
| **Redundancy[1]** | 52.5 (52.6) | 34.7 (34.6) | 35.8 (35.1) | 33.6 (36.2) |
| **$\langle I/\sigma I \rangle$[1]** | 5.5 (0.0) | 3.8 (0.3) | 4.2 (0.6) | 4.8 (1.5) |
| **$R_{merge}$[1]** | 0.635 (***) | 0.716 (20.325) | 0.726 (5.464) | 0.778 (1.745) |
| **$R_{merge}$ (anom)[1]** | | | | |
| **$CC_{1/2}$[1]** | 0.996 (0.804) | 0.965 (0.845) | 0.990 (0.332) | 0.987 (0.969) |
| **Twinning Indicated[2]** | N/A | N/A | N/A | N/A |
| **Multivariant Z-score L test[2]** | 10.697 | 16.967 | 19.007 | 13.586 |
| **Anomalous Signal[1,2]** | | | | |
| **Other errors[2]** | ice rings | anisotropic, low completeness | ice rings | anisotropic, ice rings |

| Data Set | 1811_MFP_3 | 1811_MFP_03Se | 1811_MFP_5 | 1811_MFP_12 | 1811_MFP_13 |
|---|---|---|---|---|---|
| **Beam Line** | APS 24-ID-C | APS 24-ID-C | APS 24-ID-C | APS 24-ID-E | APS 24-ID-E |
| **Frames** | 1-900 | 1-1800 | 1-1800 | 1-720 | 1-720 |
| **Oscillation** | 0.20° | 0.20° | 0.20° | 0.25° | 0.25° |
| **Space Group** | P622 | P622 | C2 | C222 | C222 |
| **Cell Dimensions (Å)[1]** | a=111.97 b=111.97 c=64.42 | a=111.82 b=111.82 c=64.25 | a=112.50 b=194.07 c=64.36 | a=110.3, b=133.4, c=61.7 | a=110.7, b=133.7, c=61.7 |
| **Angles[1]** | gamma=120 | gamma = 120 | beta = 90.38 | | |
| **Wavelength (Å)** | 0.9791 | 0.9792 | 0.9791 | 0.9791 | 0.9791 |
| **Resolution (Å)[1]** | 96.97 - 3.57 (3.91 - 3.57) | 96.84 - 3.03 (3.22 - 3.03) | 97.33 - 2.77 (2.90 - 2.77) | 85.00 - 2.00 (2.05 - 2.00) | 85.00 - 2.00 (2.05 - 2.00) |
| **# unique reflections[1]** | 3117 (713) | 4915 (742) | 34498 (4301) | 30867 (1909) | 30867 (1909) |
| **Completeness (%)[1]** | 100.0 (99.9) | 99.3 (95.6) | 98.4 (92.4) | 98.7 (84.0) | 98.7 (84.0) |
| **Redundancy[1]** | 16.4 (16.0) | 35.5 (33.7) | 6.7 (6.5) | 12.67 (8.55) | 12.67 (8.55) |
| **$\langle I/\sigma I \rangle$[1]** | 13.6 (5.4) | 11.2 (1.0) | 8.5 (0.7) | 15.0 (1.2) | 15.0 (1.2) |
| **$R_{merge}$[1]** | 0.244 (0.633) | 0.220 (3.341) | 0.171 (2.210) | 0.087 (1.448) | 0.087 (1.448) |
| **$R_{merge}$ (anom)[1]** | | 0.218 (3.392) | | | |
| **$CC_{1/2}$[1]** | 0.994 (0.938) | 0.999 (0.817) | 0.996 (0.718) | 0.998 (0.672) | 0.998 (0.672) |
| **Twinning Indicated[2]** | N/A | N/A | yes, but potential NCS | No | N/A |
| **Multivariant Z-score L test[2]** | 8.303 | 14.624 | 2.249 | 2.011 | 1.768 |
| **Anomalous Signal[1,2]** | | No | | | |
| **Other errors[2]** | anisotropic | anisotropic | | anisotropic | ice rings, anisotropic |

| Data Set | 1902_MFP_04Fe | 1902_MFP04Se | 1902_MFP_06 | 1902_MFP_06_FeB | 1902_MFP_06_FeC |
|---|---|---|---|---|---|
| **Beam Line** | APS 24-ID-C | APS 24-ID-C | APS 24-ID-C | APS 24-ID-C | APS 24-ID-C |
| **Frames** | 1-1800 | 1-1800 | 1-1800 | 1-1800 | 1-1800 |
| **Oscillation** | 0.20° | 0.20° | 0.20° | 0.20° | 0.20° |
| **Space Group** | P622 | P622 | C2 | C2 | C2 |
| **Cell Dimensions (Å)[1]** | a=111.00 b=111.00 c=64.51 | a=110.38 b=110.38 c=64.25 | a=112.4, b=194.9, c=64.6 | a=197.88 b=114.25 c=65.68 | a=194.55 b=112.32 c=64.43 |
| **Angles[1]** | gamma = 120 | gamma = 120 | beta = 90.34 | | |
| **Wavelength (Å)** | 1.7389 | 0.9792 | 0.9791 | 1.7389 | 1.7389 |
| **Resolution (Å)[1]** | 96.13 - 3.88 (4.33 - 3.88) | 95.59 - 3.82 (4.27 - 3.82) | 97.43 - 3.34 (3.61 - 3.34) | 98.94 - 3.47 (3.80 - 3.47) | 97.28 - 3.14 (3.35 - 3.14) |
| **# unique reflections[1]** | 2423 (655) | 2478 (670) | 19788 (3921) | 18467 (4229) | 23588 (3853) |
| **Completeness (%)[1]** | 99.8 (99.7) | 99.6 (98.5) | 98.4 (94.4) | 97.5 (94.6) | 96.8 (87.6) |
| **Redundancy[1]** | 33.9 (33.5) | 34.8 (34.9) | 9.6 (9.7) | 6.7 (6.8) | 6.6 (6.1) |
| **$\langle I/\sigma I \rangle$[1]** | 4.7 (0.9) | 4.6 (0.7) | 4.8 (0.8) | 2.1 (0.1) | 2.8 (0.3) |
| **$R_{merge}$[1]** | 0.497 (3.629) | 0.558 (5.383) | 0.325 (1.865) | 0.224 (3.212) | 0.201 (2.930) |
| **$R_{merge}$ (anom)[1]** | 0.488 (3.558) | 0.546 (5.314) | | 0.214 (2.949) | 0.230 (3.283) |
| **$CC_{1/2}$[1]** | 0.996 (0.828) | 0.998 (0.830) | 0.995 (0.805) | 0.991 (0.463) | 0.996 (0.747) |
| **Twinning Indicated[2]** | N/A | N/A | Yes | Yes | Yes |
| **Multivariant Z-score L test[2]** | 11.486 | 8.984 | 5.99 | 3.973 | 5.788 |
| **Anomalous Signal[1,2]** | to 12.8 Å (weak) | to 8.4 Å (weak) | | No | No |
| **Other errors[2]** | ice rings, anisotropic | ice rings, anisotropic | ice rings, anisotropic | NCS axis, potentially P622, ice rings, low completeness | ice rings, anisotropic, low completeness |

| Data Set | 1902_MFP_06b | CLD_956_7 | CLD_956_10 |
|---|---|---|---|
| **Beam Line** | APS 24-ID-C | APS 24-ID-C | APS 24-ID-C |
| **Frames** | 1-1800 | 1-1800 | 1-1800 |
| **Oscillation** | 0.20° | 0.20° | 0.20° |
| **Space Group** | P622 | P622 | P622 |
| **Cell Dimensions (Å)[1]** | a=112.8, b=112.8, c=64.6 | a=107.91 b=107.91 c=64.30 | a=103.87 b=103.87 c=61.42 |
| **Angles[1]** | gamma = 120 | gamma = 120 | gamma = 120 |
| **Wavelength (Å)** | 0.9791 | 0.9791 | 0.9791 |
| **Resolution (Å)[1]** | 97.70 - 3.18 (3.40 - 3.18) | 93.45 - 2.56 (2.67 - 2.56) | 89.96 - 2.53 (2.64 - 2.53) |
| **# unique reflections[1]** | 4341 (708) | 7515 (843) | 6816 (717) |
| **Completeness (%)[1]** | 98.5 (92.2) | 99.2 (93.4) | 98.4 (87.3) |
| **Redundancy[1]** | 51.9 (49.5) | 36.4 (36.5) | 31.2 (10.5) |
| **<I/σI>[1]** | 7.2 (0.7) | 9.2 (0.6) | 17.1 (0.8) |
| **$R_{merge}$[1]** | 0.544 (4.109) | 0.360 (4.792) | 0.187 (1.814) |
| **$R_{merge}$ (anom)[1]** | | | |
| **$CC_{1/2}$[1]** | 0.995 (0.851) | 0.998 (0.365) | 0.999 (0.513) |
| **Twinning Indicated[2]** | N/A | N/A | N/A |
| **Multivariant Z-score L test[2]** | 14.647 | 18.663 | 1.759 |
| **Anomalous Signal[1,2]** | | | |
| **Other errors[2]** | anisotropic | | |

**Table IV.5. Data statistics for all static collection data sets collected as reported by RAPD automatic processing suite and comments from Xtriage.** [1]Most data statistics were calculated and reported by RAPD. The anomalous signal in RAPD is labeled as either weak or strong and is reported in parentheses. [2]Twinning indication, multivariant Z-score L test, anomalous signal, and other errors were determined through Phenix Xtriage (2). A multivariant Z-score L test of 3.5 or greater indicates that the spread of intensities deviates from the ideal spread. Anomalous values are reported for those data sets collected at the appropriate wavelength for a signal to be observed. Bolded borders indicate the same crystal collected at different wavelengths. Friedel pairs were merged for native data sets and not merged for anomalous data sets. ***Statistics reported do not make physical sense.

| Heavy Atom Compound | Concentration | Soaking protocol | Result |
|---|---|---|---|
| Hemin | 1 mM | N/A | not soluble |
| Cobalt (II) chloride | 10 mM | Added 0.2 uL of 100 mM aqueous CoCl2 to 2 uL drop and soaked overnight | Color not observed in crystals |
| Iron (III) chloride | 10 mM | Added 0.2 uL of 100 mM aqueous FeCl3 to 2 uL drop and soaked overnight | Color not observed in crystals |
| Mersalyl acid | 1 mM | Added 0.2 uL of 10 mM mersalyl acid dissolved in water to 2 uL drop | Crystals dissolved in less than 1 hour |
| Phenylmercury acetate | 10 mM | Added 0.2 uL of 100 mM phenylmercury acetate dissolved in ethanol to 2 uL drop | Crystals dissolved in less than 1 hour |
| p-Chloromercuribenzoic acid | 10 mM | Added 0.2 uL of 100 mM p-chloromercuribenzoic acid dissolved in DMSO to 2 uL drop | Crystals dissolved in less than 1 hour |
| Mercury (II) nitrate | 10 mM | Added 0.2 uL of 100 mM mercury (II) nitrate dissolved in water to 2 uL drop | Crystals dissolved in less than 1 hour |
| Ethyl mercuric phosphate | 10 mM | see experimental | Data collected |
| I3C (Hampton Research) | 195 mM | see experimental | Some dissolving, data collected |

**Table IV.6. Multiple heavy atoms were attempted for derivatizing crystals.** The only two that worked without dissolving the crystals were the I3C derivative and ethyl mercuric phosphate. All heavy atoms were obtained from Hampton Research. The hemin was obtained from Sigma Aldrich.

**1902_MGP_04_Hg**

| | P622 | P6 | P312 | P321 | P3 | C222 |
|---|---|---|---|---|---|---|
| **Beamline** | 24-ID-C | | | | | |
| **Wavelength (Å)** | 1.0069 | | | | | |
| **Space group** | P622 | P6 | P312 | P321 | P3 | C222 |
| **Cell dimensions (Å)** | 109.39. 109.39, 63.74, 90, 90, 120 | 109.40, 109.40, 64.75, 90, 90, 120 | 109.38, 109.38, 63.75, 90, 90, 120 | 109.39, 109.39, 63.75, 90, 90, 120 | 109.42, 109.42, 63.78, 90, 90, 120 | 110.48, 189.77, 64.12, 90, 90, 90 |
| **Resolution (Å)** | 3.87 | 3.87 | 3.87 | 3.87 | 3.87 | 3.87 |
| **Unique reflections** | 2339 | 4162 | 4167 | 4296 | 7943 | 6567 |
| **Completeness** | 99.88% | 99.77% | 99.80% | 99.81% | 99.53% | 98.94% |
| **Redundancy** | 37.5 | 22.0 | 21.6 | 21.2 | 11.6 | 7.7 |
| **$R_{sym}$** | 0.416 | 0.412 | 0.407 | 0.413 | 0.393 | 0.395 |
| **I/σ(I)** | 30.9 | 17.7 | 18.7 | 18.6 | 13.3 | 10.6 |
| **Twinning Indicated** | | yes | yes | yes | yes | yes |
| **Multivariant Z-score L test** | 16.372 | 12.741 | 10.621 | 14.287 | 7.875 | 11.714 |
| **Anomalous Signal** | to 7.4 Å | to 6.7 Å | to 8.3 Å | to 6.7 Å | to 8.3 Å | to 6.7 Å |
| **Other errors** | Ice rings, anisotropic | | | | | |

| | 1902_MFP_04_HgB | CLD_956_7_I | CLD_956_10_I |
|---|---|---|---|
| | **C222** | **P622** | **P622** |
| **Beamline** | 24-ID-C | 24-ID-C | 24-ID-C |
| **Wavelength (Å)** | 1.0069 | 1.7969 | 1.7969 |
| **Space group** | C222 | P622 | P622 |
| **Cell dimensions (Å)** | 110.60, 192.06, 64.42, 90, 90, 90 | 108.06, 108.06, 64.39, 90, 90, 120 | 103.58, 103.58, 61.03, 90, 90, 120 |
| **Resolution (Å)** | 5.00 | 3.00 | 3.00 |
| **Unique reflections** | 3028 | 4776 | 4190 |
| **Completeness** | 97.10% | 99.95% | 99.93% |
| **Redundancy** | 14.5 | 52.2 | 69.7 |
| **$R_{sym}$** | 0.151 | 0.336 | 0.211 |
| **I/σ(I)** | 40.2 | 9.7 | 18.7 |
| **Twinning Indicated** | yes | N/A | N/A |
| **Multivariant Z-score L test** | 13.013 | 20.352 | 0.845 |
| **Anomalous Signal** | to 5.0 Å | none | to 5.0 Å |
| **Other errors** | | Ice rings, merging statistics indicate incorrect space group | Merging statistics indicate incorrect space group |

**Table IV.7. Data statistics for data sets collected by inverse beam in search of anomalous signal.** Twinning indication, multivariant Z-score L test, anomalous signal, and other errors were determined through Xtriage (2). A multivariant Z-score L test of 3.5 or greater indicates that the spread of intensities deviates from the ideal spread. Friedel pairs were not merged for anomalous data sets. Note that no anomalous data was collected on 1902_MFP_06.

| MFP Variant | Buffer | Additional Treatment | [protein] | Screens | Incubation Temperature | Drop Ratio |
|---|---|---|---|---|---|---|
| His-MFP | 50 mM HEPES, pH 8.0, 100 mM NaCl, 5 mM DTT | (~20% cluster content) | 14 mg/mL | 0.8 M ammonium sulfate + 0.1 M citrate buffer, pH 6.0 + Additive Screen | 18 °C | 250 nL:250 nL |
| | | | | 0.8 M ammonium sulfate + 0.1 M citrate buffer, pH 6.0 + Detergent Screen | 18 °C | |
| | | | | AmSO4 | 18 °C | |
| | | | | AmSO4 well E3 + Additive Screen | 18 °C | |
| | | | | AmSO4 well E3 + Detergent Screen | 18 °C | |
| | | | | ComPAS | 18 °C | |
| | | | | Index | 18 °C | |
| | | | | LMB | 18 °C | |
| | | | | MCSG1 | 18 °C | |
| | | | | MCSG2 | 18 °C | |
| | | | | MCSG3 | 18 °C | |
| | | | | MCSG4 | 18 °C | |
| | | | | MidasPlus | 18 °C | |
| | | | | Morpheus | 18 °C | |
| | | | | Morph II | 18 °C | |
| | | | | PEG/Ion | 18 °C | |
| | | | | PEGRx | 18 °C | |

| MFP Variant | Buffer | Additional Treatment | [protein] | Screens | Incubation Temperature | Drop Ratio |
|---|---|---|---|---|---|---|
| His-MFP | 50 mM HEPES, pH 8.0, 100 mM NaCl, 5 mM DTT | (~20% cluster content) | 5 mg/mL | ComPAS | 18 °C | 250 nL well, 200 nL protein, 50 nL seed stock |
| | | | | | 4 °C | |
| | | | | JCSG+ | 18 °C | |
| | | | | | 4 °C | |
| | | | 7 mg/mL | 0.8 M ammonium sulfate + 0.1 M citrate buffer, pH 6.0 + Additive Screen | 18 °C | 250 nL:250 nL |
| | | | | 0.8 M ammonium sulfate + 0.1 M citrate buffer, pH 6.0 + Detergent Screen | 18 °C | |
| | | | | AmSO4 | 18 °C | |
| | | | | AmSO4 well E3 + Additive Screen | 18 °C | |
| | | | | AmSO4 well E3 + Detergent Screen | 18 °C | |
| | | | | ComPAS | 18 °C | |
| | | | | Index | 18 °C | |
| | | | | LMB | 18 °C | |
| | | | | MCSG1 | 18 °C | |
| | | | | MCSG2 | 18 °C | |
| | | | | MCSG3 | 18 °C | |
| | | | | MCSG4 | 18 °C | |
| | | | | MidasPlus | 18 °C | |
| | | | | Morpheus | 18 °C | |

| MFP Variant | Buffer | Additional Treatment | [protein] | Screens | Incubation Temperature | Drop Ratio |
|---|---|---|---|---|---|---|
| His-MFP | 50 mM HEPES, pH 8.0, 100 mM NaCl, 5 mM DTT | (~20% cluster content) | 7 mg/mL | Morph II | 18 °C | 250 nL:250 nL |
| | | | | PEG/Ion | 18 °C | |
| | | | | PEGRx | 18 °C | |
| PfMFPΔ | 50 mM HEPES, pH 8.0, 100 mM NaCl, 2.5 mM DTT | Preparatory SEC, Reconstituted (~90% cluster incorporation) | 5 mg/mL | Index | 18 °C | |
| | | | | JCSG+ | 4 °C | |
| | | | | LMB | 18 °C | |
| | | | | MembFac | 18 °C | |
| | | | | Morph II | 18 °C | |
| | | | | | 4 °C | |
| | | | | PEGRx | 18 °C | |
| | | | | ProComp | 18 °C | |
| | | | 7.5 mg/mL | Index | 18 °C | |
| | | | | JCSG+ | 4 °C | |
| | | | | LMB | 18 °C | |
| | | | | MembFac | 18 °C | |
| | | | | Morph II | 18 °C | |
| | | | | | 4 °C | |
| | | | | PEGRx | 18 °C | |
| | | | | ProComp | 18 °C | |
| | | Desalted | 7.5 mg/mL | AmSO4 | 18 °C | |
| | | | | ComPAS | 18 °C | |
| | | | | Index | 18 °C | |
| | | | | JCSG+ | 18 °C | |
| | | | | Morpheus | 18 °C | |
| | | | | Morph II | 18 °C | |
| | | | | PACT | 18 °C | |
| | | | | PEGRx | 18 °C | |
| | | Desalted, Incubated with 1x hemin | 7.5 mg/mL | AmSO4 | 18 °C | |
| | | | | ComPAS | 18 °C | |
| | | | | Index | 18 °C | |
| | | | | JCSG+ | 18 °C | |
| | | | | Morpheus | 18 °C | |
| | | | | Morph II | 18 °C | |

| MFP Variant | Buffer | Additional Treatment | [protein] | Screens | Incubation Temperature | Drop Ratio |
|---|---|---|---|---|---|---|
| PfMFPΔ | 50 mM HEPES, pH 8.0, 100 mM NaCl, 5 mM DTT | Desalted, Incubated with 1x hemin | 7.5 mg/mL | PACT | 18 °C | 250 nL:250 nL |
| | | | | PEGRx | 18 °C | |
| | | Desalted, Reconstituted (~90% cluster present) | 7.5 mg/mL | 50% (v/v) MPD + Additive Screen | 18 °C | |
| | | | | AmSO4 | 18 °C | |
| | | | | ComPAS | 18 °C | |
| | | | | Index | 18 °C | |
| | | | | JCSG+ | 18 °C | |
| | | | | Morpheus | 18 °C | |
| | | | | Morpheus II | 18 °C | |
| | | | | PACT | 18 °C | |
| | | | | PEGRx | 18 °C | |
| | | | 10.0 mg/mL | 50% (v/v) MPD + Additive Screen | 18 °C | |
| | | | 12.5 mg/mL | 50% (v/v) MPD + Additive Screen | 18 °C | |
| | | Reconstituted (~35% cluster incorporation) | 5 mg/mL | AmSO4 | 18 °C | |
| | | | | ComPAS | 18 °C | |
| | | | | Index | 18 °C | |
| | | | | JCSG+ | 18 °C | |
| | | | | Morph II | 18 °C | |
| | | | | PEGRx | 18 °C | |
| | | Reconstituted (~90% cluster incorporation) | 5 mg/mL | ComPAS | 18 °C | 250 nL well, 200 nL protein, 50 nL seed stock |
| | | | | | 4 °C | |
| | | | | JCSG+ | 18 °C | |
| | | | | | 4 °C | |
| | | | 7.5 mg/mL | ComPAS | 18 °C | |
| | | | | | 4 °C | |
| | | | | JCSG+ | 18 °C | |
| | | | | | 4 °C | |

| MFP Variant | Buffer | Additional Treatment | [protein] | Screens | Incubation Temperature | Drop Ratio |
|---|---|---|---|---|---|---|
| PfMFP | 50 mM HEPES, pH 8.0, 100 mM NaCl, 5 mM DTT | Reconstituted (~60% cluster incorporation) | 2.5 mg/mL | AmSO4 | 18 °C | 250 nL:250 nL |
| | | | | ComPAS | 18 °C | |
| | | | | Index | 18 °C | |
| | | | | JCSG+ | 18 °C | |
| | | | | MembFac | 18 °C | |
| | | | | PEGRx | 18 °C | |
| | | | 5 mg/mL | AmSO4 | 18 °C | |
| | | | | ComPAS | 18 °C | |
| | | | | Index | 18 °C | |
| | | | | JCSG+ | 18 °C | |
| | | | | MembFac | 18 °C | |
| | | | | PEGRx | 18 °C | |

**Table IV.8. Anaerobically purified and reconstituted MFP has been screened against many anaerobic conditions in an attempt to obtain data-quality crystals of a [4Fe-4S] cluster-bound MFP.** All screens were carried out in an MBraun anaerobic chamber ($O_2$ = 0.1 – 2.1 ppm) under nitrogen gas. Cluster content of the protein was determined as described in the experimental section of this chapter. Over 10,000 conditions were screened with very few leads identified. Note that all proteins in this table were anaerobically purified.

| PDB Template | Alignment Score | Query Coverage | E value | Percent Identity | MFP Organism | Tool Used |
|---|---|---|---|---|---|---|
| 1AVG_I | 116.2 | 60% | 1.1E-16 | 8% | PfMFPΔ | HHPred |
| 1DZK_A | 123.6 | 67% | 1.0E-17 | 14% | PfMFPΔ | HHPred |
| 1E5P_B | 113.9 | 64% | 5.2E-16 | 14% | PfMFPΔ | HHPred |
| 1EPA_A | 28.9 | 8% | 3.7E+00 | 55.56% | H. hammondi | BLAST |
| | 28.1 | 8% | 6.0E+00 | 55.56% | N. caninum | BLAST |
| | 127.5 | 66% | 1.7E-18 | 10% | PfMFPΔ | HHPred |
| 1EW3_A | 133.3 | 67% | 1.4E-19 | 13% | PfMFPΔ | HHPred |
| 1EXS_A | 142.7 | 69% | 2.7E-21 | 13% | PfMFPΔ | HHPred |
| 1GKA_B | 31.6 | 74% | 3.8E-01 | 23.23% | T. gondii | BLAST |
| | 144 | 72% | 2.9E-21 | 15% | PfMFPΔ | HHPred |
| 1GM6_A | 134.3 | 70% | 2.5E-19 | 13% | PfMFPΔ | HHPred |
| 1I4U_A | 139.3 | 69% | 4.2E-20 | 15% | PfMFPΔ | HHPred |
| 1IIU_A | 28.8 | 52% | 3.2E+00 | 27.73% | N. caninum | BLAST |
| 1KT6_A | 138.1 | 74% | 5.8E-20 | 11% | PfMFPΔ | HHPred |
| 1KXO_A | 141.7 | 71% | 1.8E-20 | 10% | PfMFPΔ | HHPred |
| 1LF7_A | 135.7 | 71% | 2.1E-19 | 12% | PfMFPΔ | HHPred |
| 1PM1_X | 133.1 | 73% | 6.1E-19 | 9% | PfMFPΔ | HHPred |
| 1QWD_A | 37 | 79% | 6.0E-03 | 27.11% | T. gondii | BLAST |
| | 48.9 | 76% | 6.0E-07 | 33.33% | N. caninum | BLAST |
| | 148.6 | 76% | 1.0E-21 | 19% | PfMFPΔ | HHPred |
| | -- | 86% | 3.2E-07 | 24% | P. vivax | TDR Targets |
| 1X8Q_A | 137.4 | 68% | 1.5E-19 | 11% | PfMFPΔ | HHPred |
| 1YUP_A | 28.5 | 9% | 3.8E+00 | 52.63% | T. gondii | BLAST |
| 1Z24_A | 140.4 | 73% | 4.5E-20 | 14% | PfMFPΔ | HHPred |
| 2A2U_C | 127.5 | 67% | 4.1E-18 | 14% | PfMFPΔ | HHPred |
| 2ACO_A | 37.4 | 79% | 5.0E-03 | 27.11% | T. gondii | BLAST |
| | 48.9 | 69% | 5.0E-07 | 34.42% | N. caninum | BLAST |
| 2CZT_A | -- | 83% | 0.0E+00 | 12% | P. vivax | TDR Targets |
| 2HLV_A | 126 | 67% | 4.5E-18 | 14% | PfMFPΔ | HHPred |
| 2HZQ_A | 37.4 | 55% | 5.0E-03 | 31.58% | T. gondii | BLAST |
| | 38.9 | 53% | 1.0E-03 | 32.17% | H. hammondi | BLAST |
| | 43.1 | 52% | 5.0E-05 | 31.58% | N. caninum | BLAST |
| | -- | 76% | 0.0E+00 | 14% | P. falciparum | TDR Targets |
| | -- | 80% | 0.0E+00 | 22% | P. berghei | TDR Targets |
| | -- | 84% | 0.0E+00 | 22% | P. knowlesi | TDR Targets |
| | -- | 80% | 0.0E+00 | 22% | P. yoelii | TDR Targets |

| PDB Template | Alignment Score | Query Coverage | E value | Percent Identity | MFP Organism | Tool Used |
|---|---|---|---|---|---|---|
| 2HZR_A | 147.4 | 71% | 1.1E-21 | 17% | PfMFPΔ | HHPred |
| 2L5P_A | 137.1 | 74% | 1.2E-19 | 11% | PfMFPΔ | HHPred |
| 2P56_A | -- | 43% | 7.8E-01 | 38% | P. knowlesi | TDR Targets |
| 2RA6_B | 129.4 | 68% | 1.1E-18 | 8% | PfMFPΔ | HHPred |
| 2XST_A | 134.3 | 68% | 7.5E-20 | 8% | PfMFPΔ | HHPred |
| 2YG2_A | 72.9 | 71% | 6.5E-08 | 9% | PfMFPΔ | HHPred |
| 2YXD_A | -- | 37% | 4.8E-01 | 33% | P. yoelii | TDR Targets |
| 3CMG_A | -- | 16% | 5.8E-01 | 32% | P. berghei | TDR Targets |
| 3CQR_B | 147.9 | 74% | 1.8E-21 | 12% | PfMFPΔ | HHPred |
| 3EBW_A | 142.5 | 71% | 2.9E-21 | 15% | PfMFPΔ | HHPred |
| 3FIQ_A | 118.9 | 67% | 7.5E-17 | 8% | PfMFPΔ | HHPred |
| 3KFF_A | 134.2 | 69% | 1.0E-19 | 11% | PfMFPΔ | HHPred |
| 3KQ0_A | 121.6 | 67% | 1.7E-16 | 11% | PfMFPΔ | HHPred |
| 3L4R_A | 131.1 | 65% | 7.8E-19 | 13% | PfMFPΔ | HHPred |
| 3MBT_A | 37 | 75% | 6.0E-03 | 26.62% | T. gondii | BLAST |
| | 51.2 | 73% | 7.0E-08 | 34.16% | N. caninum | BLAST |
| | 147.5 | 73% | 6.4E-22 | 18% | PfMFPΔ | HHPred |
| | -- | 79% | 0.0E+00 | 24% | P. knowlesi | TDR Targets |
| | -- | 75% | 0.0E+00 | 25% | P. yoelii | TDR Targets |
| 3O22_A | 139.5 | 69% | 7.4E-21 | 11% | PfMFPΔ | HHPred |
| 3QKG_A | 134.3 | 69% | 5.5E-19 | 9% | PfMFPΔ | HHPred |
| 3S26_A | 138.5 | 72% | 8.4E-20 | 6% | PfMFPΔ | HHPred |
| 3SAO_A | 124.3 | 66% | 6.5E-18 | 11% | PfMFPΔ | HHPred |
| 3WE2_A | -- | 24% | 3.8E-01 | 36% | P. berghei | TDR Targets |
| | -- | 26% | 8.7E-01 | 33% | P. vivax | TDR Targets |
| 4A3T_A | -- | 48% | 0.0E+00 | 20% | P. yoelii | TDR Targets |
| 4CK4_A | 139.6 | 67% | 1.2E-20 | 18% | PfMFPΔ | HHPred |
| 4ES7_A | 140.2 | 77% | 8.3E-20 | 8% | PfMFPΔ | HHPred |
| 4GE1_B | 145 | 75% | 9.5E-21 | 14% | PfMFPΔ | HHPred |
| 4N7C_A | 131.6 | 67% | 1.1E-18 | 12% | PfMFPΔ | HHPred |
| 4NYQ_A | 46.3 | 68% | 1.3E-03 | 10% | PfMFPΔ | HHPred |
| 4O3M_A | -- | 36% | 7.7E-01 | 34% | P. falciparum | TDR Targets |
| 4O9S_A | 147.1 | 81% | 1.1E-20 | 10% | PfMFPΔ | HHPred |
| 4ODD_C | 125.8 | 66% | 3.6E-18 | 13% | PfMFPΔ | HHPred |
| 4OS3_A | 135.9 | 71% | 1.5E-19 | 8% | PfMFPΔ | HHPred |
| 4QAF_A | 123.3 | 66% | 5.5E-18 | 11% | PfMFPΔ | HHPred |

| PDB Template | Alignment Score | Query Coverage | E value | Percent Identity | MFP Organism | Tool Used |
|---|---|---|---|---|---|---|
| 4R0B_A | 142.4 | 69% | 6.4E-21 | 11% | PfMFPΔ | HHPred |
| 4RUN_A | 133 | 64% | 2.3E-19 | 16% | PfMFPΔ | HHPred |
| 4WFV_A | 127.2 | 67% | 1.9E-18 | 12% | PfMFPΔ | HHPred |
| 4XMH_A | 145.1 | 73% | 5.5E-21 | 13% | PfMFPΔ | HHPred |
| 4ZHH_E | 127.8 | 71% | 4.2E-18 | 10% | PfMFPΔ | HHPred |
| 5EZ2_A | 42.7 | 52% | 7.0E-05 | 37.17% | N. caninum | BLAST |
| 5EZ2_B | 140.6 | 70% | 3.1E-20 | 14% | PfMFPΔ | HHPred |
| 5F6Z_A | 42.7 | 52% | 7.0E-05 | 37.17% | N. caninum | BLAST |
| 5H9N_A | 120.6 | 65% | 4.0E-17 | 11% | PfMFPΔ | HHPred |
| 5NGH_A | 132.5 | 67% | 3.3E-19 | 11% | PfMFPΔ | HHPred |
| 5U1T_A | -- | 94% | 0.0E+00 | 14% | P. knowlesi | TDR Targets |
| 5X7Y_C | 140.6 | 71% | 2.0E-20 | 14% | PfMFPΔ | HHPred |
| 6GQZ_A | 128.2 | 71% | 2.4E-18 | 9% | PfMFPΔ | HHPred |

**Table IV.9. Many known structures of lipocalins have been used in an attempt to solve the structure of MFP by molecular replacement.** All models were used for molecular replacement against the data set N5900_748_13 in all processed space groups (see figure IV.4). For other data sets, the molecular replacements were performed with the homology model discussed in Chapter 2 and 3 as well as 3QKG ($\alpha_1$-microglobulin) and 4XMH (nitrophorin 7), due to their heme-binding ability (11,12), as well as 1EPA, 1QWD, 2HZQ, 3MBT, and 3WE2 due to their identification in multiple databases. Models used were identified based on the predictions from HHPred (8) when making the homology model of MFP as well as NCBI BLASTs against the PDB (10) and the TDR Targets database (9).

# REFERENCES


1. Matz, J. M., and Matuschewski, K. (2018) An in silico down-scaling approach uncovers novel constituents of the Plasmodium-containing vacuole. *Scientific Reports* **8**
2. Zwart, P. H., Grosse-Kunstleve, R. W., and Adams, P. D. (2005) Xtriage and Fest: automatic assessment of X-ray data and substructure structure factor estimation. *CCP4 Newsletter **Winter**, Contribution 7*
3. Gateiger, E., Hoogland, C., Gattiker, A., Devaud, S., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005) Protein identification and analysis tools on the ExPASy server. in *The Proteomics Protocols Handbook* (Walker, J. M. ed.), Humana Press. pp 571-607
4. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. in *Methods in Enzymology*, Academic Press. pp 307-326
5. Kabsch, W. (2010) XDS. *Acta Crystallographica Section D Biological Crystallography* **66**, 125-132
6. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography* **66**, 213-221
7. Sheldrick, G. M. (2008) A short history ofSHELX. *Acta Crystallographica Section A Foundations of Crystallography* **64**, 112-122
8. Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A. N., and Alva, V. (2018) A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *Journal of Molecular Biology* **430**, 2237-2243
9. Agüero, F., Al-Lazikani, B., Aslett, M., Berriman, M., Buckner, F. S., Campbell, R. K., Carmona, S., Carruthers, I. M., Chan, A. W. E., Chen, F., Crowther, G. J., Doyle, M. A., Hertz-Fowler, C., Hopkins, A. L., McAllister, G., Nwaka, S., Overington, J. P., Pain, A., Paolini, G. V., Pieper, U., Ralph, S. A., Riechers, A., Roos, D. S., Sali, A., Shanmugam, D., Suzuki, T., Van Voorhis, W. C., and Verlinde, C. L. M. J. (2008) Genomic-scale prioritization of drug targets: the TDR Targets database. *Nature Reviews Drug Discovery* **7**, 900-907
10. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-410
11. Karnaukhova, E., Rutardottir, S., Rajabi, M., Wester Rosenlöf, L., Alayash, A. I., and Åkerström, B. (2014) Characterization of heme binding to recombinant $\alpha_1$-microglobulin. **5**
12. Champagne, D. E., Nussenzveig, R. H., and Ribeiro, J. M. C. (1995) Purification, Partial Characterization, and Cloning of Nitric Oxide-carrying Heme Proteins (Nitrophorins) from Salivary Glands of the Blood-sucking InsectRhodnius prolixus. *Journal of Biological Chemistry* **270**, 8691-8695

# CHAPTER V: CHARACTERIZATION OF MFP HOMOLOGS

## CHAPTER SUMMARY

MFP, a putative heme binding and iron-sulfur containing protein, has been found to be conserved across all species of *Plasmodium*, and homologs have been identified across a wide range of parasitic organisms including apicomplexa. In this chapter, we report investigations of three non-*Plasmodium* homologs, from *Hammondia hammondi*, *Neospora caninum*, and *Toxoplasma gondii* and one *Plasmodium* homolog from *Plasmodium berghei*. The three non-*Plasmodium* homologs studied here are found to be insoluble under the conditions used for purification. In contrast, MFP from *Plasmodium berghei* (PbMFP) behaves similarly to that of MFP from *Plasmodium falciparum* (PfMFP), binding both heme and an iron sulfur cluster. PbMFP, like PfMFP, contains a [4Fe-4S] cluster when anaerobically purified or upon *in vitro* reconstitution. PbMFP also will bind a [2Fe-2S] cluster upon aerobic purification. Under both aerobic and anaerobic conditions, PbMFP without cluster binds a heme moiety; however, when full reconstitution of a [4Fe-4S] cluster is obtained, no heme binding is observed. Taken together, these results indicate that PbMFP is functioning in a similar fashion to PfMFP. Although much is left unanswered about the non-*Plasmodium* homologs, our work identifies strong similarities between the *in vitro* biochemical properties of two MFP proteins found in relatively distant species of *Plasmodium*. This finding implies that MFP will perform a similar role in all species of *Plasmodium*.

## CONTRIBUTIONS

Khan Osman provided the expression constructs for all homologs (Niles Lab, MIT). Allena Mistral Goren and Hannah Grupe performed the recombinant expression and purification (Drennan Lab, MIT). Hannah Grupe and Dante Avalos performed the aerobic heme binding assays (Drennan Lab, MIT). Mary C. Andorfer helped with collection of EPR spectra (Drennan Lab, MIT). Allena Mistral Goren performed the anaerobic work (Drennan Lab, MIT). Mary C. Andorfer, Christopher D Dawson, Francesca A. Vacarro, Phoebe Li, Gerardo Perez, and Hannah Grupe helped with crystallographic work (Drennan Lab, MIT). EPR experiments were performed in the Department of Chemistry Instrumentation facility under the supervision of John Grimes and Walt Massefski (MIT). We would like to thank Professor Daniel Suess (MIT), Constantin Voll (The 3M Company), and Alex Taguchi (RubrYc Therapeutics) for their help in learning EPR spectroscopy techniques and theory.

## INTRODUCTION

Through our work outlined thus far, MFP has been identified as an essential protein in *Plasmodium falciparum*, but crystallographic studies have not yielded structural insight into the protein's function. Many times, when crystallization of a protein of interest does not yield results, we turn to homologs to obtain insight into the structure of the given protein. During our bioinformatic investigation of MFP in Chapter II, we identified three potential homologs from other apicomplexa. The putative proteins from *Hammondia hammondi*, *Neospora caninum*, and *Toxoplasma gondii* offer another avenue for study and could lead to insight into the conservation of MFP throughout apicomplexa and other parasitic organisms.

In addition to the three apicomplexa, MFP from *Plasmodium berghei* was recently identified as Parasitophorous Vacuole Protein 5 (PV5) (1). PV5 was found to be essential for *Plasmodium* viability, and was named due to its localization to the parasitophorous vacuole (1). Given that the vertebrate host for *P. berghei* and *P. falciparum* differ (2), these findings suggest at least some conservation of function between MFP in different species of *Plasmodium*. In order to bolster this hypothesis, we undertook biochemical characterization of this protein, which we have termed PbMFP for continuity, to verify similar *in vitro* functions.

Given the anticipated structural similarities and the difficulty of crystallizing PfMFP, we pursued the four homologous proteins identified here for crystallographic work. To verify that these proteins are biologically relevant homologs, we endeavored to biochemically characterize these homologs in a similar fashion to that of PfMFP. All three non-*Plasmodium* homologs were insoluble under the conditions tested; however, the MFP homolog from *Plasmodium berghei* appeared well behaved under our purification conditions and was used for additional crystallographic and biochemical experiments. We identified conserved *in vitro* protein properties, indicating that this protein likely has conserved functions across all *Plasmodium* species.

# RESULTS

## MFP HOMOLOGS EXPRESS WELL IN *E. COLI*

Previously, we had identified three different organisms containing MFP-like proteins (**See Figure II.6**). Given that all three of these organisms have similar modes of infection to *Plasmodium* (3-5), we decided to investigate these proteins in addition to PfMFP. The three proteins were termed HhMFP, NcMFP and TgMFP, corresponding to the organisms *Hammondia hammondi*, *Neospora caninum*, and *Toxoplasma gondii*, respectively. We recombinantly expressed the three native genes using Rosetta-gami™ 2 cells due to the cell line's ability to express proteins from genes containing rare codons (6). Each of the three homologs expressed in high relative abundance compared to the native *E. coli* proteins; however, the proteins were found to be only slightly soluble, with the majority of the protein being found in the insoluble fraction of the lysate (**Figure V.1**).

In order to investigate the oligomeric homogeneity of these homologs, the soluble fractions were concentrated for analytical size exclusion chromatography (**Figure V.2**). TgMFP was the most well behaved of the three homologs with approximately half being monomeric with the other appreciable fraction being higher order aggregates. HhMFP and NcMFP each show appreciable amounts of three different oligomeric states in addition to the aggregates. These proteins were not pursued further.

In contrast, PbMFP expressed in large quantities and was highly soluble under the purification conditions used for the PfMFP constructs (**Figure V.1C**). The oligomeric profile of the initial PbMFP purification was analyzed by size exclusion chromatography (SEC) (**Figure V.2**), and was found to have a similar oligomeric profile to that of PfMFP (**Figure V.5**), showing a mix of three main oligomeric states, a monomeric, a dimeric, and a trimeric or multimeric peak. Given the high sequence identity (71.68%) between PfMFP and PbMFP (**Figure V.3**), this finding is not surprising. Interestingly, purifications of PbMFP containing no reducing agent under aerobic conditions appear reddish-brown by eye, and were investigated for similar cofactor incorporation as that of PfMFP.

## PbMFP CONTAINS AN IRON-SULFUR CLUSTER

Given the visible color of the aerobically prepared PbMFP protein, we probed the nature of the chromophore. Under aerobic purification, UV-Visible spectroscopy suggests a [2Fe-2S] cluster with strong absorbances around 420 and 450 nm (**Figure V.4A**). This finding was confirmed using electron paramagnetic resonance (EPR) spectroscopy (**Figure V.4B**). The aerobically purified PbMFP was reconstituted *in vitro* to determine whether the [2Fe-2S] cluster was the only relevant cluster identity. The anaerobic reconstitution of PbMFP produced a UV-Visible spectrum more consistent with a [4Fe-4S] cluster, specifically a small peak at 412 nm (**Figure V.5A**). The reconstituted protein was found to contain 3.3 irons per protein and was also found to change the oligomeric profile of the protein (**Figure V.5B**). It is worth noting that without a strong understanding of the physiological activity of MFP and its homologs, speculating on whether either of the clusters is physiologically relevant would be premature.

## PbMFP BINDS A HEME MOIETY

Although no physiological activity has been identified as of yet, heme binding to MFP seems to be an important aspect of its function. Like PfMFP, we tested PbMFP for its ability to bind heme both with and without iron-sulfur cluster present. Like PfMFPΔ, a soret at 412 nm is observed under aerobic binding conditions (**Figure V.6A**). Interestingly, the soret at 370 nm observed for the C70A, C216A, and C68A/C93A variants of PfMFP is also observed under aerobic conditions for PbMFP (**See Figures III.17 and III.20**). Under anaerobic conditions, the binding of heme appears to be altered to that of PfMFPΔ, with a much larger peak at 386 nm corresponding with free heme (**Figure V.6B, See Figure III.12B**). Both time courses of apo- and holo-PbMFP appear to have very little shift in soret over time, implying very little binding of heme to PbMFP under anoxic conditions (**Figure V.6C-D**).

## ATTEMPTS AT CRYSTALLIZATION OF PbMFP

In order to try to obtain structural insight about the function of MFP, we attempted crystallization of the PbMFP protein. We initially hypothesized that the dimer, being the oligomeric state thought to be of biological relevance for many of the lipocalin-like family of proteins (7), would be the state that would crystallize. As such, we individually screened each of the oligomeric

states that were separated out by size exclusion chromatography (SEC). It is worth noting that all three fractions contained visible color, indicating some amount of [Fe-S] cluster was present in all three oligomeric profiles. The clear delineation between these states in the SEC traces proved useful for this endeavor. We first tried reproducing the most well-behaved condition identified in the crystallization trials of PfMFP. Interestingly, the multimeric fraction yielded crystals similar to those of the PfMFP (**Figure V.7A**) whereas the monomeric and dimeric fractions solely yielded precipitation (**Figure V.7B**). To find additional crystallization conditions, each of the oligomeric states was screened against 400 different conditions, which yielded two promising conditions for the multimeric fraction and a single condition for both the monomeric and dimeric fractions (**Figure V.7C, D**). We are currently working on optimizing the conditions in order to obtain structural data.

## DISCUSSION

Very often in crystallography, when a protein of interest does not readily crystallize, we turn to homologs in the hopes that they will facilitate structure determination. In our efforts to structurally characterize MFP, we turned to a homolog from *Plasmodium berghei* as well as the closest three non-*Plasmodium* homologs found through our bioinformatics work reported in Chapter II. Here we detail the work completed so far on the putative MFP homologs.

The insolubility of the non-*Plasmodium* homologs studied herein makes it difficult to make any conclusions regarding their biochemical or biophysical properties. More studies of how to solubilize these proteins will be needed to determine whether these proteins are, indeed, functional homologs of MFP.

The MFP homolog from *Plasmodium berghei* appears to behave similarly to MFP from *P. falciparum*. PbMFP, like PfMFP, binds both a [4Fe-4S] cluster and a heme moiety, but not at the same time. Interestingly, PbMFP purifies aerobically under non-reducing conditions with a small amount of [2Fe-2S] cluster intact. The protein fold of an [Fe-S] cluster-containing protein can have a large effect on cluster stability, so it is likely that the residues that differ between PfMFP and PbMFP are driving the altered *in vitro* incorporation. We hypothesize that the residues directly adjacent to Cys68 and Cys93 in PfMFP could play a role in this difference. The aliphatic residue directly adjacent to the first cysteine in the $C_{68}XC_{70}$ motif and the two flanking residues

of Cys93 in PfMFP are all significantly altered in PbMFP. Ala67 and Ala94 in PfMFP are a leucine and an aspartic acid in PbMFP, respectively, and Ser92 of PfMFP is an alanine in PbMFP. These altered residues could provide more conformational rigidity, increasing the stability of an [Fe-S] cluster under the conditions tested. This could be investigated in other species of *Plasmodium* or by mutational analysis of the constructs already obtained.

Interestingly, PbMFP appears to bind heme in a slightly altered fashion from PfMFPΔ. Although the main soret at 412 nm is still present, there is an additional peak observed at 370 nm upon titration of PbMFP with heme (**Figure V.6A**). This peak is believed to correspond to a 5-coordinate heme species (8,9). This finding, especially paired with the altered cluster binding of PbMFP, could indicate that the cysteines in PbMFP have slightly altered chemical properties from those of PfMFPΔ. Similar to PfMFP, though, the [4Fe-4S] cluster of PbMFP does appear to significantly hinder heme binding under anoxic conditions. The reduced heme binding to apo-PbMFP under anoxic conditions is perplexing. Nevertheless, it is possible, given the extended time to strip the cluster under anoxic conditions as compared with PfMFPΔ (more than 14 days for PbMFP vs. 5 days for PfMFPΔ), that the apo form of PbMFP is unstable after cluster removal. While no visible aggregates were observed, we did not check the oligomeric state after incubation with EDTA. More investigation will be needed to determine whether the reduced binding of heme to apo-MFP under anoxic conditions was an artifact of experimentation or an altered property of the PbMFP protein.

Crystallographic studies of PbMFP appear to be a relevant avenue to pursue, with two promising conditions being obtained during the screening process. When trying to reproduce the crystals observed with PfMFP, only the multimeric state as identified by SEC formed crystals. Although this finding on its own was puzzling, screening of PbMFP yielded two promising conditions for the multimeric protein and only a single condition for the monomeric and dimeric form, indicating that there may be some relevance to the higher order oligomers. Although the relevance of this oligomer may be an artifact of *in vitro* purification, it may prove useful for crystallization and structure identification. It is also worth noting that the trimeric state is the same oligomeric state of the purified heme bound PfMFPΔ, adding to the probability that there could be some *in vivo* utility for this oligomeric state.

## CONCLUSION

MFP homologs are a yet unharnessed tool to add to our understanding of MFP and its necessity for *Plasmodium* viability during the blood stage of infection. PbMFP is a well-behaved homolog that could, in the near future, provide necessary insight into the structure and function of MFP. The similarity in both primary structure of PfMFP and PbMFP, as well as the cofactor binding behavior, suggest that the structure of PbMFP will yield valuable insight into the elusive structure of PfMFP. Additionally, the more stable iron-sulfur cluster in PbMFP could help in teasing out the biological relevance of the cluster.

## EXPERIMENTAL PROCEDURES

### EXPRESSION OF PBMFP

PbMFP in a pET28b vector was transformed into NiCo21(DE3) competent *Escherichia coli* cells (New England Biolabs) and plated on an LB (Lennox) agar plate with 50 μg/mL kanamycin (Gold Biotechnologies) selection antibiotic. A single colony was used to inoculate a 5 mL culture of LB containing 50 μg/mL kanamycin. The culture was grown overnight at 37 °C shaking at 220 rpm.

One milliliter of the overnight culture was used to inoculate a 50 mL start culture of LB containing 50 μg/mL kanamycin. This culture was grown for 3 hours at 37 °C shaking at 220 rpm.

The full starter culture was used to inoculate 2 L of LB containing 50 μg/mL kanamycin. This material was split into two 2.8 L Fernbach flasks, each containing 1 L of the culture. The cultures were grown at 37 °C shaking at 220 rpm until $OD_{600} \sim 0.6$. At that time, protein expression was induced by the addition of β-D-1-thiogalactopyranoside (IPTG, Gold Biotechnologies) to a final concentration of 400 μM, and the culture was allowed to continue growing at 18 °C and 220 rpm.

After 18 hours, cells were harvested by centrifugation at 4,000 *g* for 30 minutes at 4 °C before being frozen in liquid nitrogen for storage. Cell pellets were stored at -70 °C until purification.

## EXPRESSION OF NON-*PLASMODIUM* MFP HOMOLOGS

The three non-*Plasmodium* MFP homologs in pET28b vectors were expressed as above for PbMFP with the following changes: all expression constructs were transformed into Rosetta-gami™ 2 (DE3)pLysS Competent cells (Novagen), and the selection antibiotic of chloramphenicol was used at 34 μg/mL in addition to kanamycin.

## AEROBIC PURIFICATION OF MFP HOMOLOGS

Approximately 10 g of cells were resuspended in 50 mL of lysis buffer containing 50 mM HEPES, pH 8.0 and 300 mM sodium chloride (NaCl). Approximately 40 mg of lysozyme (Gold Biotechnologies), 2 μL of benzonase (Millipore), and one cOmplete protease cocktail tablet (Roche) were added to the solution. The cells were allowed to lyse at 4 °C on a rotator for 2 hours. The cell lysate was clarified by centrifugation at 25,000 *g* and 4 °C for 30 minutes. The clarified lysate was filtered with a 22-micron sterile syringe filter. The lysate was loaded onto 10 mL of Streptactin-XT Superflow resin (IBA) in a gravity column. The column was then washed with 50 mL of lysis buffer. Protein was subsequently eluted in three fractions using 6, 16, and 8 mL of elution buffer (lysis buffer with 50 mM biotin added). Fractions were collected and analyzed by gel electrophoresis.

## ANAEROBIC PURIFICATION OF PBMFP

PbMFP was purified anaerobically in addition to the aerobic purification. In an MBraun anaerobic chamber at 4 °C under $N_2$ atmosphere, approximately 10 mg frozen cell pellet with 40 mg of solid lysozyme (Gold Biotechnologies), 2 μL of benzonase (Millipore) and one cOmplete protease cocktail tablet (Roche) were resuspended in 50 mL of anaerobic lysis buffer containing 50 mM HEPES, pH 8.0, 300 mM NaCl, and 10 mM dithiothreitol (DTT). The cells were chemically lysed for 3 hours and clarified by centrifugation. The oxygen-free centrifuge tubes were sealed in the anaerobic chamber before being removed for centrifugation at 25,000 g and 4 °C for 30 minutes. The sealed centrifuge bottles were brought back into the anaerobic chamber. The clarified lysate was filtered with a 22-micron sterile syringe filter and loaded onto 10 mL of Streptactin-XT superflow resin (IBA) in a gravity column. The protein was eluted from the

column as previously stated with anaerobic elution buffer (anaerobic lysis buffer with 50 mM biotin added). All fractions were collected for analysis by aerobic gel electrophoresis.

## Oligomeric analysis of MFP homologs

The elution fractions of the individual non-*Plasmodium* MFP homologs were concentrated to 300 µL of protein. These solutions were filtered through a costar Spin-X 22 µm cellulose spin filter by centrifugation, and subsequently loaded onto a Superdex$^{TM}$ 200 increase 10/300 GL equilibrated in protein storage buffer (50 mM HEPES, pH 8.0 and 100 mM NaCl) using a 100 µL protein loop on a BioRad FPLC instrument. The sample progress through the column was observed at 280 and 410 nm by an inline multi-wavelength detector to monitor protein absorbance and iron-sulfur cluster absorbance, respectively.

## Reconstitution of a [4Fe-4S] cluster in PbMFP and metal analysis

In an MBraun anaerobic chamber at 4 °C under $N_2$ atmosphere, PbMFP in 50 mM HEPES, pH 8.0, 300 mM NaCl, and 10 mM DTT was incubated with 2 molar equivalents of $FeCl_3$ for approximately 5 minutes before the addition of 2 molar equivalents of $Na_2S$. The solution was incubated overnight, approximately 16 hours, and was desalted to remove excess iron using a HiPrep 26/10 Sephadex G-25 desalting column (GE). UV-Vis spectra were collected using a Thermo Fisher NanoDrop$^{TM}$ 2000c. A small aliquot of protein was removed from the anaerobic chamber for iron analysis. Briefly, 100 µL of reconstituted protein was added to 125 µL of a 0.4% (w/v) ascorbate solution, 7.5 µL concentrated hydrochloric acid, and 12.5 µL of 5 mg/mL ferene. This solution was vortexed for 30 second, after which 500 µL of 8 M guanidinium chloride and 100 µL of 10 M ammonium acetate were added. The solution was incubated for at least 30 minutes after which the absorbance was read and compared to the extinction coefficient for iron bound ferene of 34,500 $M^{-1}cm^{-1}$ at 593 nm. The calculated concentrations were increased by 8.45-fold to account for dilution during the assay preparation.

## Electron paramagnetic resonance (EPR) spectroscopy of PbMFP

Aerobic, as isolated PbMFP was reduced for EPR spectroscopy using 10 mM sodium dithionite. Samples were flash frozen in liquid nitrogen and measured by X-band (~9.4 GHz) EPR

spectroscopy in the Department of Chemistry Instrumentation Facility at The Massachusetts Institute of Technology. Spectra were collected using a Bruker EMX spectrometer equipped with an ER 4119HS high sensitivity resonator, a Gunn diode microwave source, and a Bruker/CoolEdge 4K waveguide cryogen-free cryostat. Spectra were measured at 9.37 GHz microwave frequency, 200 μW power, 8 G modulation amplitude, and the temperatures indicated.

## HEME BINDING ASSAYS

To investigate the stoichiometry of heme binding, heme was titrated against apo-PbMFP, both aerobically and anaerobically. To obtain both aerobic and anaerobic apo-PbMFP, 250 mM EDTA was added to a 100 μM sample of PbMFP and incubated until no [Fe-S] cluster peak was observed by UV-Vis spectroscopy, approximately 5 days or 2 weeks under aerobic and anaerobic conditions, respectively. The protein was desalted on a PD-10 desalting column (Amersham Biosciences) and protein concentration was measured by the absorbance at 280 nm ($\varepsilon_{280,\ PbMFP}$ = 27,390 $M^{-1}$ $cm^{-1}$, this value and all subsequent were obtained through ExPASy ProtParam (10)). For the aerobic samples, an initial spectrum of 5 or 10 μM apo-PbMFP was taken using a 1 cm path length cuvette on a Thermo Fisher NanoDrop 2000c. Subsequently, 1 mL of 5 or 10 μM apo-PbMFP was spiked with 1 or 2 μL of 0.59 mM hemin (Sigma) in 0.5 M NaOH, respectively, and allowed to incubate for 5 minutes. At this time, a UV-Visible spectrum was taken of the solution. In parallel, the same amount of hemin was added to protein buffer with no protein to obtain a control spectrum. The process of spiking the solution with heme was continued until 17 or 34 μL of heme for 5 or 10 μM samples of protein, respectively, approximately 2 equivalents, had been added. Final spectra were prepared by subtracting the heme control at the appropriate concentration from the spectrum of PbMFP incubated with heme. The anaerobic titration was performed in the same manner with the following changes: the experiment was conducted in an MBraun anaerobic chamber under a nitrogen atmosphere with 10 mM DTT as an additional component in the buffer. Protein used was prepared anaerobically as described above and the heme solution was allowed to gas exchange in the anaerobic chamber for 24 hours before the experiment. No control spectra of heme alone were obtained under anaerobic conditions.

To investigate the time dependence of apo-PbMFP heme binding, 10 μM of anaerobically purified apo- or reconstituted-PbMFP was incubated with 10 μM hemin in 50 mM HEPES, pH 8.0, 100 mM NaCl, and 10 mM DTT in an MBraun anaerobic chamber. UV-Vis spectra blanked against 10 μM heme in 50 mM HEPES, pH 8.0, 100 mM NaCl and 10 mM DTT were collected using a cuvette on a Thermo Fisher NanoDrop 2000c at 30 seconds, 5 minutes, 10 minutes, 30 minutes, 1 hour, 2 hours, and 4 hours.

## CRYSTALLIZATION OF PBMFP

Purified PbMFP was additionally purified by size exclusion chromatography on a Hi Load 16/60 Superdex$^{TM}$ 200 preparatory grade column (GE life sciences). Three peaks were observed: a monomer peak, a dimer peak, and a trimer/tetramer peak, which was termed "multimer." The fractions for these peaks were collected individually and concentrated to approximately 20 mg/mL by $A_{280}$ absorbance. A well solution of 0.2 M $MgCl_2$, 0.1 M HEPES, pH 7.4, and 25% (v/v) isopropanol was set against sitting drops of 1 μL of each oligomeric state and 1 μL of well solution and incubated at 4 °C in a 24-well intelli-plate (Art Robbins Instruments). Screen trays were set with all three oligomeric states at 10 mg/mL protein by $A_{280}$ in 50 mM HEPES, pH 8.0 and 100 mM NaCl. Sparse matrix screens of PbMFP were performed aerobically using a NT8® drop setter for protein crystallography (Formulatrix) with drops containing 250 μL protein and 250 μL well solution set against a 70 μL well. Screens were stored aerobically at 20 °C.

FIGURES

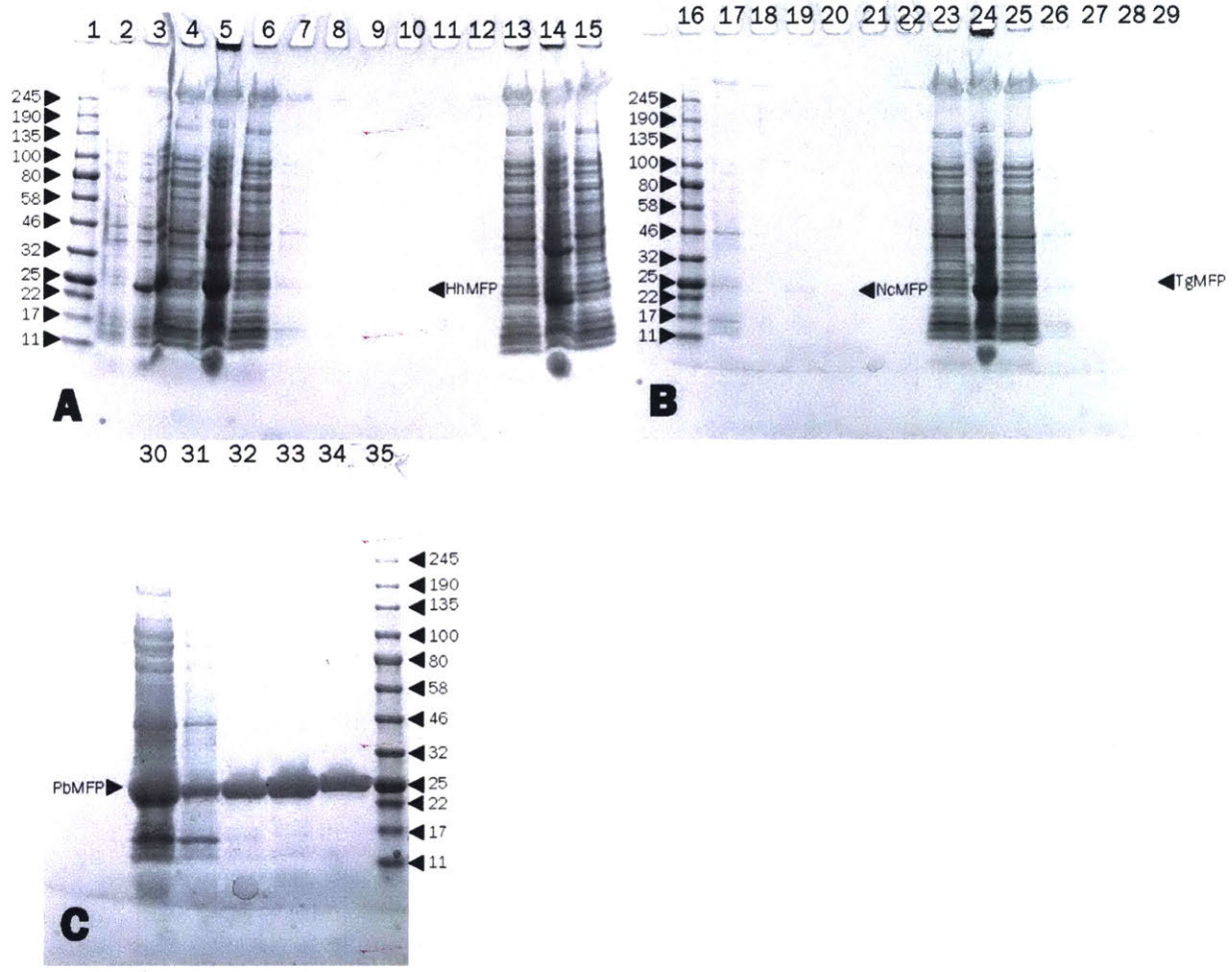



**Figure V.1. Non-*Plasmodium* homologs are insoluble under our purification conditions. (A&B)** Purifications of non-*Plasmodium* MFP homologs (HhMFP, lanes 4-10, NcMFP, lanes 13-20, TgMFP, lanes 23-29) show the protein of interest in the insoluble fraction after lysate clarification (lanes 5, 14, and 24) rather than the soluble fractions (lanes 6-10, 15-20, and 25-29). As contrast, **(C)** PbMFP is highly soluble under the purification conditions used. Lanes 6, 15, 25, and 30 correspond to the flow through of protein binding to the column. Lanes 7, 17, 26, and 31 correspond to the wash step. Lanes 8-10, 18-20, 27-29, and 32-34 correspond to elution steps.

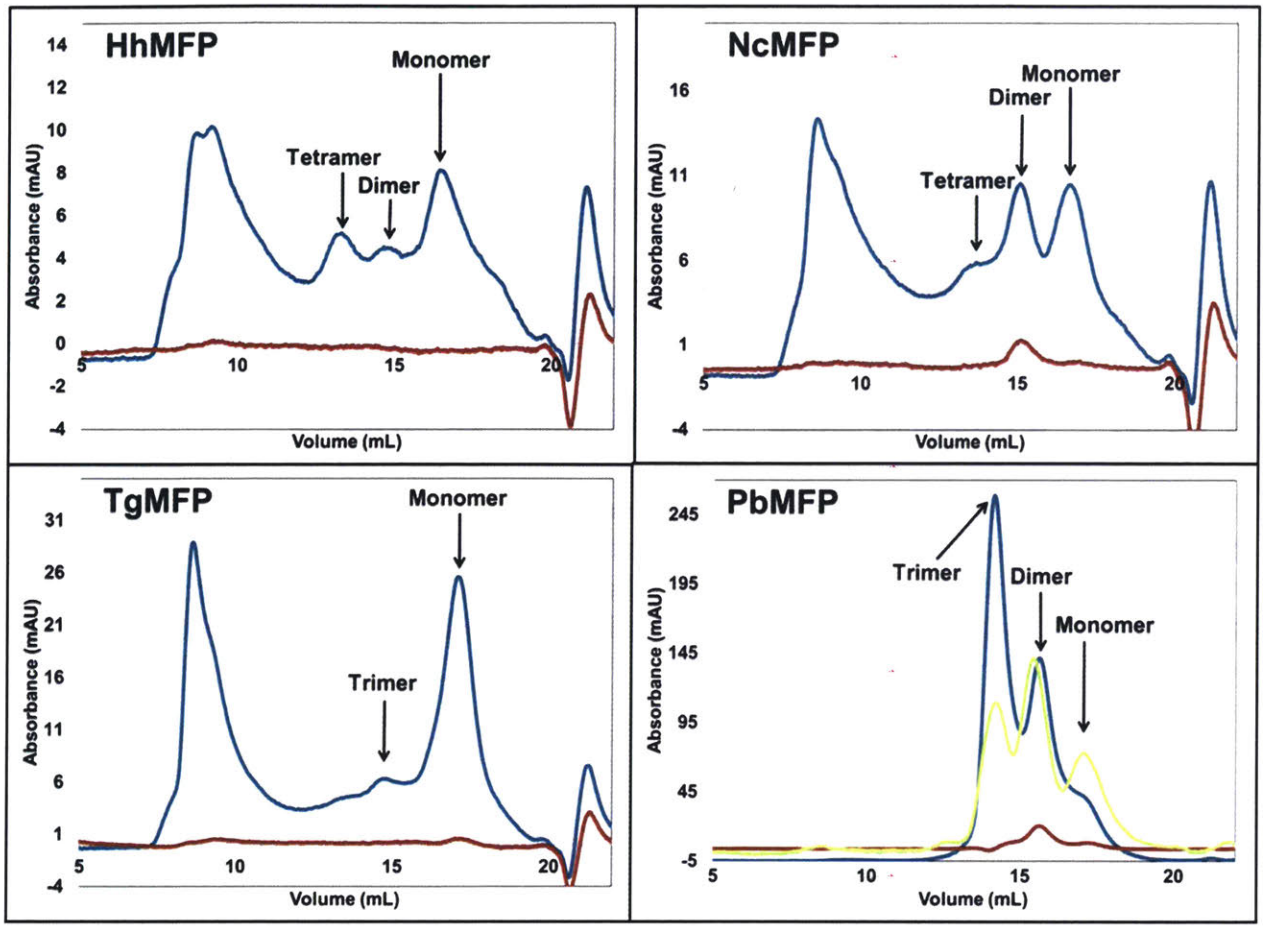



**Figure V.2. Oligomeric profiles of MFP homologs by size exclusion chromatography are vastly different.** The three non-*Plasmodium* homologs (HhMFP, NcMFP, and TgMFP) contain a large amount of higher order oligomers that elute at the void volume (~9 mL). PbMFP shows an oligomeric profile similar to that of MFP from *Plasmodium falciparum* (Green line, His-MFP). For all charts, blue and red traces correspond to the absorbance at 280 nm (protein) and absorbance at 410 nm (iron-sulfur cluster) of the sample, respectively.

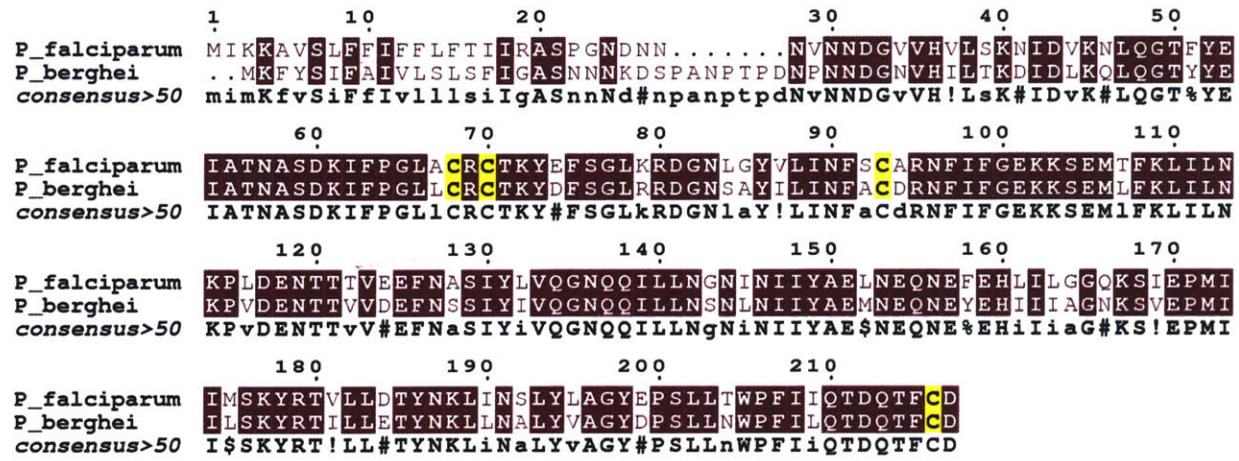



**Figure V.3. Sequence alignment of PfMFP and PbMFP.** Conserved residues are highlighted in red. Similar or unaligned residues are in red text. The four conserved cysteines that potentially bind the iron-sulfur cluster are highlighted in yellow. Residue numbers correspond to those of *P. falciparum*. The sequences are 71.5% identical and 85.0% similar. Figure generated with MultAlin (11) and ESPript (12).

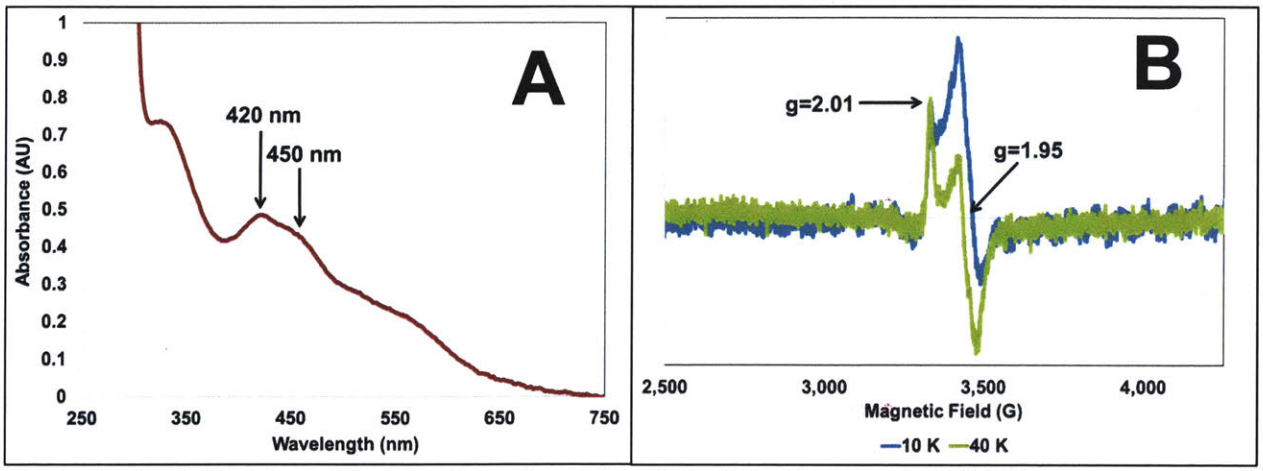



**Figure V.4. PbMFP purifies aerobically with a [2Fe-2S] cluster. (A)** UV-Vis spectroscopy of PbMFP after aerobic purification without the presence of reductant shows a chromophore with strong absorbance bands near 420 and 450 nm as well as a reddish-brown color (not pictured). **(B)** EPR spectroscopy confirms the presence of the [2Fe-2S] cluster. G-values are reported for the 40 K spectrum, as the 10 K spectrum was saturated at the power used.

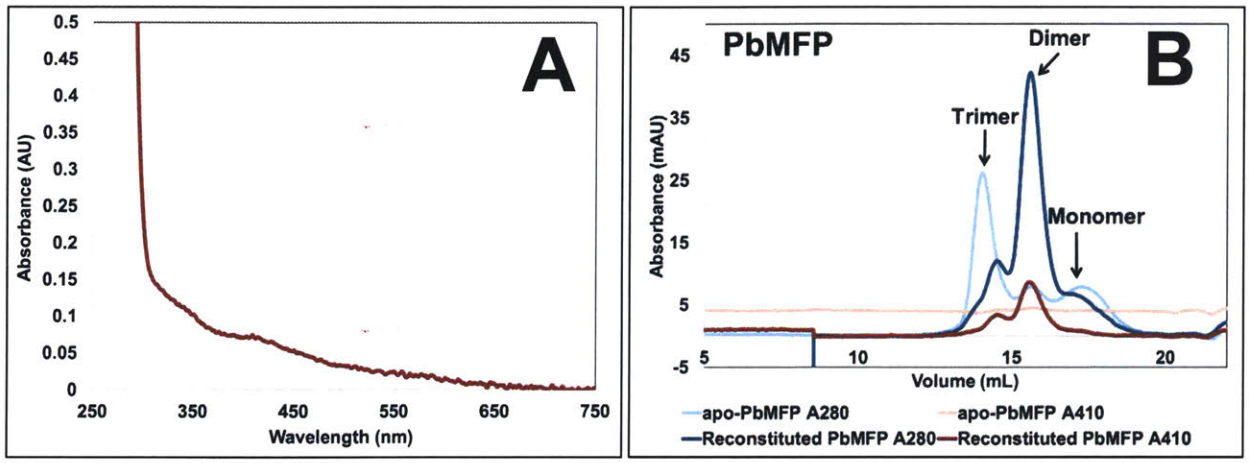



**Figure V.5. PbMFP will fully reconstitute to a [4Fe-4S] cluster under anaerobic conditions.** **(A)** The UV-Vis spectrum of PbMFP after reconstitution indicates a [4Fe-4S] cluster. **(B)** The reconstituted protein appears to change oligomeric state from a trimer to a dimer by SEC.

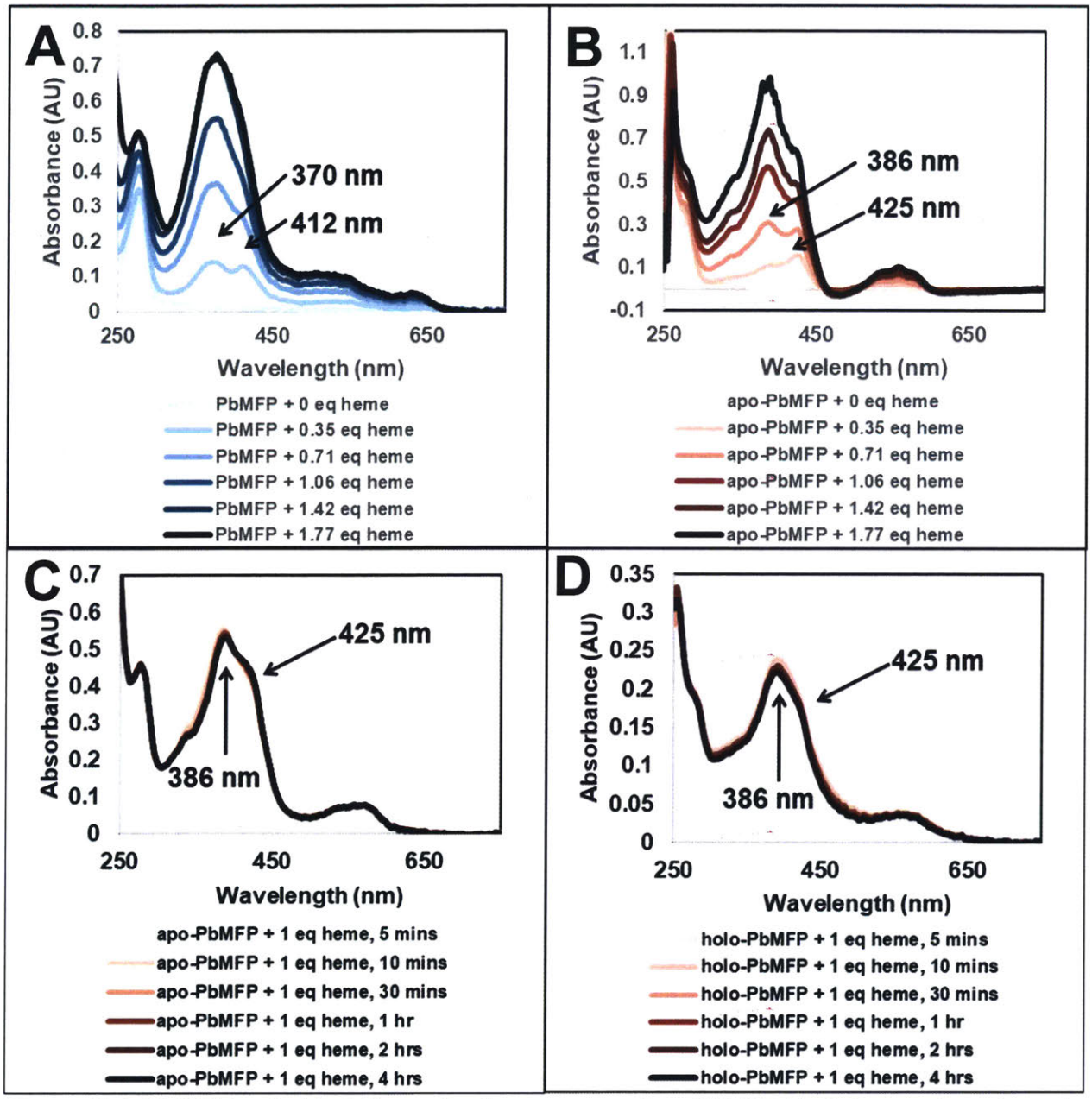



**Figure V.6. PbMFP only binds heme in the apo state.** Apo-PbMFP binds heme when cluster is not present both **(A)** aerobically and **(B)** anaerobically. Under anaerobic conditions **(C)** a larger soret indicative of heme binding is observed after 4 hours than **(D)** the fully reconstituted MFP, in which almost no soret shift is seen after 4 hours of incubation with equimolar heme. (A-C) were all performed at 10 μM protein and (D) was performed at 5 μM protein.

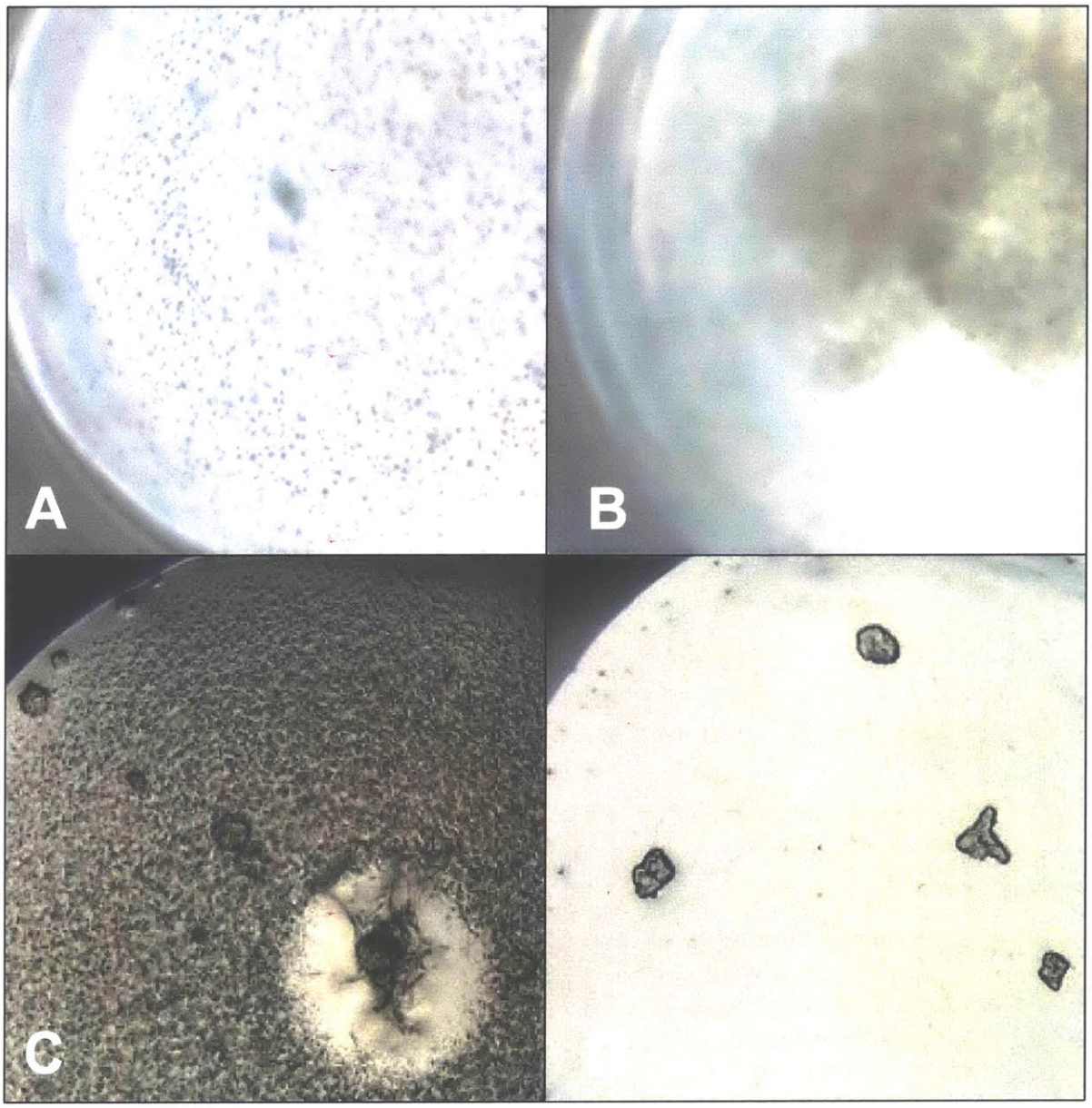



**Figure V.7. Crystallization trials of PbMFP.** The multimer of PbMFP is thus far the most promising species for crystallization. All three oligomeric states of PbMFP were screened using the crystallization condition for PfMFP (0.2 M $MgCl_2$, 0.1 M HEPES, pH 7.4, and 25% (v/v) isopropanol) as well as sparse matrix screens. **(A)** Crystals observed using the multimeric fraction off of the SEC obtained in the original PfMFP isopropanol condition. **(B)** Non-crystalline precipitation was observed when the monomer or dimer was used for crystallization attempts. **(C)** Quasicrystals obtained from screen tray with 0.2 M calcium chloride, 0.1 M sodium acetate, pH 4.6, and 20% (v/v) isopropanol **(D)** Possible crystals obtained from screen tray with 30 mM lithium sulfate, 30 mM sodium sulfate, 30 mM potassium sulfate, 0.1 M MOPSO/BisTris, pH 6.5, 5% (w/v) PEG 20,000, 25% (w/v) trimethylpropane, and 1% (w/v) NDSB 195. The last two conditions were obtained with the multimer. The potential crystals pictured in D were also obtained in the dimer and monomer condition (not pictured).

# REFERENCES


1. Matz, J. M., and Matuschewski, K. (2018) An in silico down-scaling approach uncovers novel constituents of the Plasmodium-containing vacuole. *Scientific Reports* **8**
2. Perkins, S. L., and Schall, J. J. (2002) A molecular phylogeny of malarial parasites recovered from cytochrome *b* gene sequences. *Journal of Parasitology* **88**, 972-978
3. Frénal, K., Dubremetz, J.-F., Lebrun, M., and Soldati-Favre, D. (2017) Gliding motility powers invasion and egress in Apicomplexa. *Nature Reviews Microbiology* **15**, 645-660
4. Walzer, K. A., Adomako-Ankomah, Y., Dam, R. A., Herrmann, D. C., Schares, G., Dubey, J. P., and Boyle, J. P. (2013) Hammondia hammondi, an avirulent relative of Toxoplasma gondii, has functional orthologs of known T. gondii virulence genes. **110**, 7446-7451
5. Nolan, S. J., Romano, J. D., Luechtefeld, T., and Coppens, I. (2015) Neospora caninum Recruits Host Cell Structures to Its Parasitophorous Vacuole and Salvages Lipids from Organelles. **14**, 454-473
6. Novagen. (2006) pET System Manual, 11th Edition.
7. Flower, D. R. (1996) The lipocalin protein family: structure and function. *Biochemical Journal* **318**, 1-14
8. Brewitz, H. H., Kühl, T., Goradia, N., Galler, K., Popp, J., Neugebauer, U., Ohlenschläger, O., and Imhof, D. (2015) Role of the Chemical Environment beyond the Coordination Site: Structural Insight into Fe (IIIP) rotoporphyrin Binding to Cysteine-Based Heme-Regulatory Protein Motifs. *ChemBioChem* **16**, 2216-2224
9. Wißbrock, A., Paul George, A. A., Hans, Kühl, T., and Imhof, D. (2019) The molecular basis of transient heme-protein interactions: analysis, concept and implementation. *Bioscience Reports* **39**, BSR20181940
10. Gateiger, E., Hoogland, C., Gattiker, A., Devaud, S., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005) Protein identification and analysis tools on the ExPASy server. in *The Proteomics Protocols Handbook* (Walker, J. M. ed.), Humana Press. pp 571-607
11. Corpet, F. (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research* **16**, 10881-10890
12. Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research* **42**, W320-W324

# CHAPTER VI: STRUCTURAL STUDIES OF MRUBY2 TOWARD THE DESIGN OF A BIOLOGICALLY ENGINEERED HEME SENSOR

## CHAPTER SUMMARY

mRuby2, a fluorescent protein, is utilized during *in vivo* studies to track the localization of tagged proteins. We have identified a novel function of mRuby2 which both provides a new use and a limitation of mRuby2 for localization studies. Herein, we discuss the binding of heme and simultaneous fluorescent quenching of mRuby2 *in vitro*. We have identified that mRuby2 binds one heme molecule per protein and that the quenching of fluorescence is dependent upon the ratio of heme to protein. To try to probe the molecular interaction causing this fluorescence quenching, we have determined the crystal structure of mRuby2 to 1.95 Å resolution and have obtained crystals of mRuby2 which likely contains heme bound. *In crystallo* fluorescence is observed for mRuby2 when no heme is present, providing a tool for identifying crystals of mRuby2 with and without heme bound.

## CONTRIBUTIONS

Khan Osman provided the expression constructs for all homologs (Niles Lab, MIT). Allena Mistral Goren performed the recombinant expression and purification and heme binding assays (Drennan Lab, MIT). Allena Mistral Goren, Francesca A. Vacarro, Mary C. Andorfer, Hannah Grupe, and Dante Avalos performed the crystallographic work (Drennan Lab, MIT).

## INTRODUCTION

Heme regulation is important for many different organisms, as free heme is cytotoxic (1,2). The ability to probe cellular heme pools could lead to a stronger understanding of the regulation and utilization of heme in different organisms but probing the localization of free heme *in vivo* is challenging. Many studies have used heme-binding proteins in tandem with fluorescent probes to look at free heme in the cell (3-7). One strategy widely utilized is the investigation of labile heme by pairing GFP and its analogs with Cyt $b_{562}$ (3,4). The FRET quenching of this pair upon Cyt $b_{562}$ heme binding acts as a read out for heme localization (3). The utility of these sensors is hindered by the indirect readout of heme binding, as the observed signal is related to the ligand binding to a partner protein rather than heme binding to the fluorescent probe itself. The development of a fluorescent probe for the direct readout of labile heme pools *in vivo* by binding of heme to the probe would provide for a more direct readout of concentrations and localization.

During our search for a fluorescent protein with heme binding properties, we identified mRuby2 as a candidate for this work (Personal communication, Khan Osman, Niles Lab). mRuby2 is a bright fluorescent protein from the RFP family originally designed to improve the FRET dynamic range when paired with a GFP analog, Clover (8). A substitution of just four amino acids (**Figure VI.1**) reduced photobleaching effects over its parent RFP, mRuby, making it a more appropriate probe for some biological work, whether for FRET or localization studies (8). Here, we discuss our novel findings that mRuby2 binds heme *in vitro* and the implications of this phenomenon toward the development of a heme probe that could be used to track localization of heme *in vivo*.

## RESULTS

### BINDING OF HEME TO MRUBY2 QUENCHES FLUORESCENCE

During *in vitro* biochemical work, our collaborator Khan Osman (Niles Lab, MIT) noticed fluorescence quenching of mRuby2 upon incubation of the protein with heme. This serendipitous observation prompted our studies of this fluorescent probe.

In order to perform our analyses of mRuby2, the protein was recombinantly expressed using a Strep-II tag. The presence of this tag did not prevent fluorescence of mRuby2 (**Figure VI.2**).

Upon incubation of StrepII-tagged mRuby2 with heme dissolved in sodium hydroxide (NaOH), we observed a quenching of fluorescence (**Figure VI.3A**). This quenching does not occur when the same amount of 0.5 M NaOH is added to the protein solution, eliminating the possibility of a pH dependent quenching or a protein degradation event by NaOH (**Figure VI.3B**). The ability of heme to bind mRuby2 was assessed measuring absorption spectra (**Figure VI.3C, VI.4A**). Two observable heme-binding modalities are observed; the first is a peak at 370 nm and the second at ~408 nm (**Figure VI.3C, VI.4A**). These peaks are likely indicative of a 5-coordinate and 6-coordinate binding mode, respectively (9-11). We find that the addition of ~ one equivalent of heme is sufficient to quench visible fluorescence (**Figure VI.4C**). As such, we were interested in identifying the structural changes that occur upon heme-binding to mRuby2.

## MRUBY2 FLUORESCES *IN CRYSTALLO*

In order to compare a heme-bound mRuby2 structure to a structure of mRuby2 in a fluorescent state, we endeavored to solve the first structure of mRuby2. Crystallization screens of mRuby2 were performed using sparse matrix screens to identify ideal conditions for protein crystallization. This protein readily crystallized in many different conditions during sparse matrix screening. To verify that our crystallographic studies of mRuby2 were of a relevant fluorescent conformation, we exposed the crystals to UV-light. Interestingly, many of the crystals of mRuby2 fluoresced under UV-light, indicating that a relevant conformation of the protein was present *in crystallo* (**Figure VI.5**).

## MRUBY AND MRUBY2 ARE SIMILAR IN STRUCTURE

The structure of mRuby2 was solved to 1.95 Å resolution using mRuby (PDB 3U0M) as a molecular replacement model (**Figure VI.6A, Table VI.1**) (12). This model was chosen over that of mRuby under acidic conditions (PDB 3U0L), as the pH of our protein buffer is 8.0, there was no buffer present in the precipitating solution, and sodium formate is a weak base. Thus our structure is likely to be at a basic pH. The structure of mRuby2 overlays very well with mRuby, having a relative RMSD of 0.737 Å for all atoms (**Figure VI.6B**). No distinct differences were observed between the two structures other than the change of the four amino acids that define the

difference between mRuby and mRuby2 and the split conformation of the fluorophore in mRuby2, which can be visualized in our mRuby2 structure (**Figure VI.6C**). Although the density for the fluorophore is non-ideal (**Figure VI.6C**), when either conformation is modeled alone, positive difference density indicating the presence of the second conformation appears after refinement (**Figure VI.6D**). Interestingly, one conformation of the fluorophore overlays well with the single conformation of the fluorophore visualized in the mRuby structure and accounts for 54% occupancy of the fluorophore in the mRuby2 structure. The four altered residues are not in direct proximity to the fluorophore (**Figure VI.7A**). The four amino acid substitutions between mRuby and mRuby2 have strong density in our structure and overlay well in position with the mRuby structure (**Figure VI.7B-D**).

### ATTEMPTS TO STRUCTURALLY CHARACTERIZE HEME-BOUND MRUBY2 ARE ON-GOING

Attempts to perform structural analysis on heme-bound mRuby2 are on-going. We recently obtained two precipitating conditions upon screening of mRuby2 incubated with heme that are promising (**Figure VI.8A**). Upon optimization of the initial conditions, some crystals and quasicrystals have been obtained (**Figure VI.8B-F**). Interestingly, some of these crystals fluoresce (**Figure VI.8B-C**) whereas others do not (**Figure VI.8D-F**). This difference in fluorescence likely indicates that some crystals have heme present while others do not. Interestingly, the morphology and time of growth on the two sets of crystals are different. The non-fluorescent crystals grow almost instantaneously but are small and not conductive to data collection yet. They do, however, appear more crystalline in nature that the fluorescent quasicrystals. The fluorescent quasicrystals also grow more slowly and are much larger.

## DISCUSSION

In our efforts to obtain a novel probe for the investigation of heme localization *in vivo*, we identified mRuby2 as a fluorescent protein whose fluorescence is quenched upon incubation with heme. As such, we endeavored to determine the structural properties of mRuby2 that underlie this interesting phenomenon.

Heme binding by mRuby2 appears to be concentration dependent by UV-Vis spectroscopy (**Figure IV.4A**). Fluoresence quenching also appears concentration dependent (**Figure IV.4C**). The titratable nature of the quenching of mRuby2 fluorescence by heme could provide a method for investigating heme concentrations *in vivo*, however more work is necessary to investigate this behavior *in vivo* and *in vitro*.

It is worth noting that while we are reporting this heme-based fluorescence quenching as a novel function of mRuby2 for studies of heme localization, we are also voicing a limitation of this probe for *in vivo* assays where free heme is present. Reports of mRuby2 behaving poorly compared to predictions (13) could be attributed to this heme fluorescence quenching and, as such, care should be taken when using mRuby2 and related fluorophores for imaging when free heme is present.

In order to investigate the structural underpinnings of this phenomenon, we turned to protein crystallography. We first needed to solve the structure of mRuby2 in a fluorescent state to identify differences leading to fluorescence quenching by the presence of heme. mRuby2 readily crystallizes in many conditions. To determine whether any of these conditions were relevant, we attempted to find a quick indication of the fluorescent state of mRuby2 *in crystallo*. The identification of fluorescence of mRuby2 crystals under UV-light provides a visual probe to quickly screen whether the crystal of interest is in the desired fluorescent state. Although we have not utilized this method for other fluorescent proteins, we believe that other bright fluorophores would likely exhibit similar behavior *in crystallo*, providing a quick method of screening for crystals with relevant fluorescence activity.

The structure of mRuby2 solved from crystals that fluoresce under UV light aligns well with the structure of mRuby, its parent fluorophore, indicating the potential for many shared properties. Interestingly, the currently reported chemical differences between mRuby and mRuby2 are minimal. The differences include a slight increase in the molar absorptivity and a slight increase in quantum yield as well as an improved FRET response when paired with the Clover fluorescent protein (8). These differences are small and the changes in the structure are small. The four amino acids that are different between mRuby and mRuby2 are not within close proximity of the fluorophore (**Figure VI.6D**). The most interesting of the four altered side chains is Asp214. This

residue is highly solvent exposed and not interacting with the fluorophore, but the residue adjacent to it, His215, appears to interact with the fluorophore through a pi-stacking interaction (**Figure VI.9**). This pi-stacking interaction is conserved between the two proteins. It is possible that the substitution of a serine to an aspartic acid alters the stability or the electronics of the amino acids around it, mainly His215, altering the interaction and potentially allowing for more flexibility of the fluorophore at higher pH. The structure of mRuby2 does not contain a clear open coordination site for heme binding, so structural investigations of the heme bound form are still warranted. Based on the similarity of the two structures, we predict that mRuby is likely also quenched within the presence of heme. More studies are needed to determine whether this is indeed the case.

As for the crystals obtained upon incubation of mRuby2 with heme, the morphology difference of crystals grown in the presence of heme could aid in picking crystals for data collection. As the non-fluorescent crystals are very small, we have not yet verified that they are, indeed, protein crystals and not salt or free heme. More work is needed.

## CONCLUSION

We have identified a novel utility of mRuby2 as a tool for the investigation of *in vivo* localization of heme. mRuby2 appears to bind heme *in vitro* as indicated by a change in its absorption features. The fluorescence intensity of mRuby2 also decreases upon heme binding. The novel utility of this protein also adds some limitations for cellular localization investigations when free heme is present.

## EXPERIMENTAL PROCEDURES

### EXPRESSION OF MRUBY2

mRuby2 in a pET28b vector was transformed into NiCo21(DE3) competent *Escherichia coli* cells (New England Biolabs) and plated on an LB (Lennox) agar plate with 50 μg/mL kanamycin (Gold Biotechnologies) selection antibiotic. A single colony was used to inoculate a 5 mL culture of LB containing 50 μg/mL kanamycin. The culture was grown overnight at 37 °C shaking at 220 rpm.

One milliliter of the overnight culture was used to inoculate a 50 mL start culture of LB containing 50 μg/mL kanamycin. This culture was grown for 3 hours at 37 °C shaking at 220 rpm.

The full start culture was used to inoculate 2 L of LB containing 50 μg/mL kanamycin. This material was split into two 2.8 L Fernbach flasks, each containing 1 L of the culture. The cultures were grown at 37 °C shaking at 220 rpm until $OD_{600} \sim 0.6$. At that time, protein expression was induced by the addition of β-D-1-thiogalactopyranoside (IPTG, Gold Biotechnologies) to a final concentration of 400 μM, and the culture was allowed to continue growing at 18 °C and 220 rpm.

After 18 hours, cells had an observable pink color and were harvested by centrifugation at 4,000 *g* for 30 minutes at 4 °C before being frozen in liquid nitrogen for storage. Cell pellets were stored at -70 °C until purification.

### PURIFICATION OF MRUBY2

Approximately 10 g of cells were resuspended in 50 mL of lysis buffer containing 50 mM HEPES, pH 8.0 and 300 mM sodium chloride (NaCl). Approximately 40 mg of lysozyme (Gold Biotechnologies), 2 μL of benzonase (Millipore), and one cOmplete protease cocktail tablet (Roche) were added to the solution. The cells were allowed to lyse at 4 °C on a rotator for 2 hours. The cell lysate was clarified by centrifugation at 25,000 *g* and 4 °C for 30 minutes. The clarified lysate was filtered with a 22-micron sterile syringe filter. The lysate was loaded onto 10 mL of Streptactin-XT Superflow resin (IBA) in a gravity column. The column was then washed with 50 mL of lysis buffer. Protein was subsequently eluted in three fractions using 6, 16, and 8 mL of elution buffer (lysis buffer with 50 mM biotin added). Fractions were collected and analyzed by gel electrophoresis. Fractions containing pure protein were pooled and diluted to 45 mL with 50 mM HEPES, pH 8.0 and 300 mM NaCl (storage buffer). This solution buffer exchanged by passage through a 50 mL Amicon stirred cell concentrator. When the volume of protein was less than 5 mL, the protein was diluted to 45 mL again and concentrated to less than 5 mL. This step was repeated two more times. The resulting buffer exchanged solution was approximately 5 mg/mL by UV-Vis spectroscopy ($\varepsilon_{280}$ = 30,035 $M^{-1}$ $cm^{-1}$, this value and all

subsequent were obtained through ExPASy ProtParam (14)). This concentrated protein was stored at 4 °C for further analysis and crystallography.

## HEME BINDING ASSAYS

To investigate the heme binding properties of mRuby2, hemin was titrated against recombinantly expressed protein. An initial spectrum of 10 μM mRuby2 was collected using a 1 cm path length quartz cuvette on a NanoDrop 2000c (ThermoFisher). Subsequently, 2 μL of 0.59 mM hemin (Sigma) dissolved in 0.5 M NaOH was added to 1 mL of 10 μM mRuby2 and allowed to incubate for 5 minutes. After the incubation, a UV-Visible spectrum was taken of the protein solution. The process of spiking the solution with 2 μL of hemin was continued until 34 μL of hemin, approximately 2 equivalents, had been added. A heme standard dissolved in protein storage buffer was run in the same manner in parallel. The final spectra were obtained by subtracting the hemin in buffer spectrum from the mRuby2 with hemin spectrum at specific concentrations of hemin.

## CRYSTALLIZATION OF MRUBY2

Sparse matrix screens were performed with mRuby2 between 3.4 and 4.3 mg/mL measured by $A_{280}$ ($\varepsilon$ = 30,035 $M^{-1}$ $cm^{-1}$ (14)) in 50 mM HEPES, pH 8.0 and 100 mM NaCl in 96-well Intelli-plate sitting-drop vapor diffusion trays with 150 nL of protein and 150 nL of well solution in the drop and 70 μL of well solution using a Phoenix pipetting robot (Art Robbins Instruments). Screens were stored at 18 °C and crystals formed almost instantaneously. Diffraction quality crystals were obtained from these screening experiments. Small, cubic, pink crystals were observed upon incubation against a well solution of 3.5 M sodium formate (from Index screen, Hampton Research) with the same well solution added to the drop. Crystals were looped through paraffin oil (Hampton research) as cryoprotectant before being flash frozen in liquid nitrogen.

## DETERMINATION OF CRYSTAL FLUORESCENCE

To determine whether crystals fluoresce, the crystallization trays were placed under a microscope. An image of the crystal under bright-field was taken with an iPhone 8 camera (Apple). The bright-field light was turned off and a UV lamp was shone on the crystals. Before

looking at the crystals again under the microscope, a UV-filter for a typical camera lens was used to avoid eye and camera damage. Images were taken through this filter of mRuby2 crystals. No images were taken if fluorescence was not observed.

## DATA COLLECTION AND STRUCTURE DETERMINATION OF MRUBY2

A native dataset was collected on this crystal at the 24-ID-C beam line at the Advanced Photon Source (Argonne, IL) using the Pilatus-6M pixel array detector at 12,662 eV or 0.9792 Å. The crystal was kept at a temperature of 100 K during data collection. Data were indexed, integrated, and scaled using the XDS processing suite (15) in the space group $P2_12_12_1$ to 1.954 Å resolution (**Table IV.1**).

The structure of mRuby2 was solved to 1.954 Å resolution by molecular replacement using the previously reported mRuby structure (PDB 3U0M) with waters removed. A single solution was identified using the program Phaser (16) with an LLG score of 4169 and a TFZ score of 49.1 and a single molecule in the asymmetric unit. The NRQ cif restraints file was downloaded from the PDB and used for refinement. A rigid body refinement was performed on the initial model, but no change was observed. Ten rounds of simulated annealing were run in Phenix Refine to reduce model bias before beginning refinement. Residues and side chains without electron density were removed before iterative building in Coot (17) refinement in Phenix (18,19) were performed. During later stages of refinement, waters were added manually to the structure using positive $F_o$-$F_c$ electron density.

During building, residues without density and areas of side chains without density were removed from the working model. The resulting structure includes residues 19-182 and 186-241 of 251 residues, where residue 1 is the first residue of the StrepII tag (**Figure VI.1**). 99.05% of residues are in the most favorable area of a Ramachandran plot, 0.95% are in the additionally favored region, and 0% are in the Ramachandran outlier region. The current working model has an $R_{work}$ and an $R_{free}$ are 0.1956 and 0.2307, respectively. Verification of the final structure was performed using a simulated annealing composite omit map. All data and refinement statistics are in **Table VI.1**.

## CRYSTALLIZATION OF MRUBY2 WITH HEMIN

Sparse matrix screens of mRuby2 at 4.0 mg/mL measured by $A_{280}$ ($\varepsilon$ = 30,035 $M^{-1}$ $cm^{-1}$ (14)) in 50 mM HEPES, pH 8.0 and 100 mM NaCl and incubated with 1 equivalent, 2.5 equivalents, and 5 equivalents of hemin (Sigma Aldrich) dissolved in NaOH were performed in 96-well Intelli-plate sitting-drop vapor diffusion trays (Art Robbins Instruments) with 150 nL of protein and 150 nL of well solution in the drop and 70 µL of well solution using a Phoenix pipetting robot (Art Robbins Instruments). The screens were stored at 18 °C in a Rock Imager 1000 (Formulatrix). Two promising conditions of (1) 4.0 M potassium formate, 0.1 M Bis-TRIS propane, pH 9.0, and 2% polyethylene glycol monomethyl ether 2000 (PEG-MME 2000) (from PEGRx screen, Hampton Research) and 0.1 M Bis-TRIS, pH 6.5 and 25% polyethylene glycol 3350 (PEG-3350) (from Index screen, Hampton) were identified and used for optimization of mRuby2 incubated with hemin bound. These initial conditions were used to set optimization trays. In these optimizations, fluorescent crystals were obtained as well as non-fluorescent crystals.

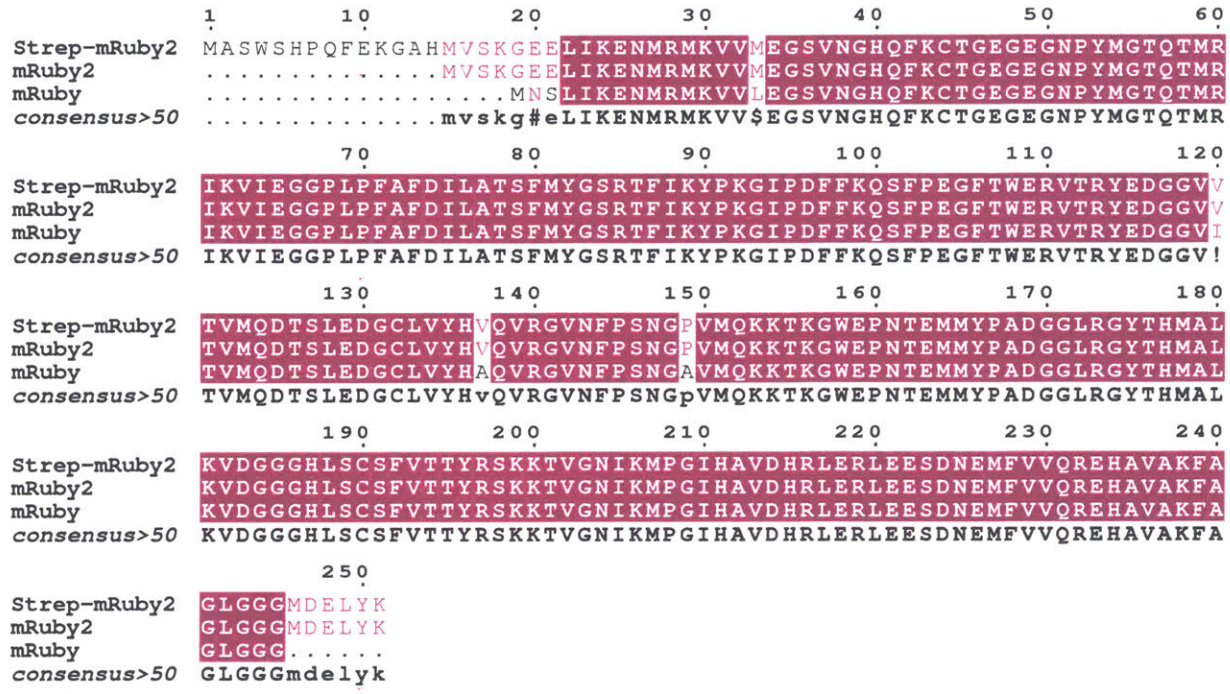



**Figure VI.1. mRuby2 differs from mRuby by 4 amino acids.** The expression construct, labeled here as StrepII-mRuby2, is referred to as mRuby2 in the text of this chapter.

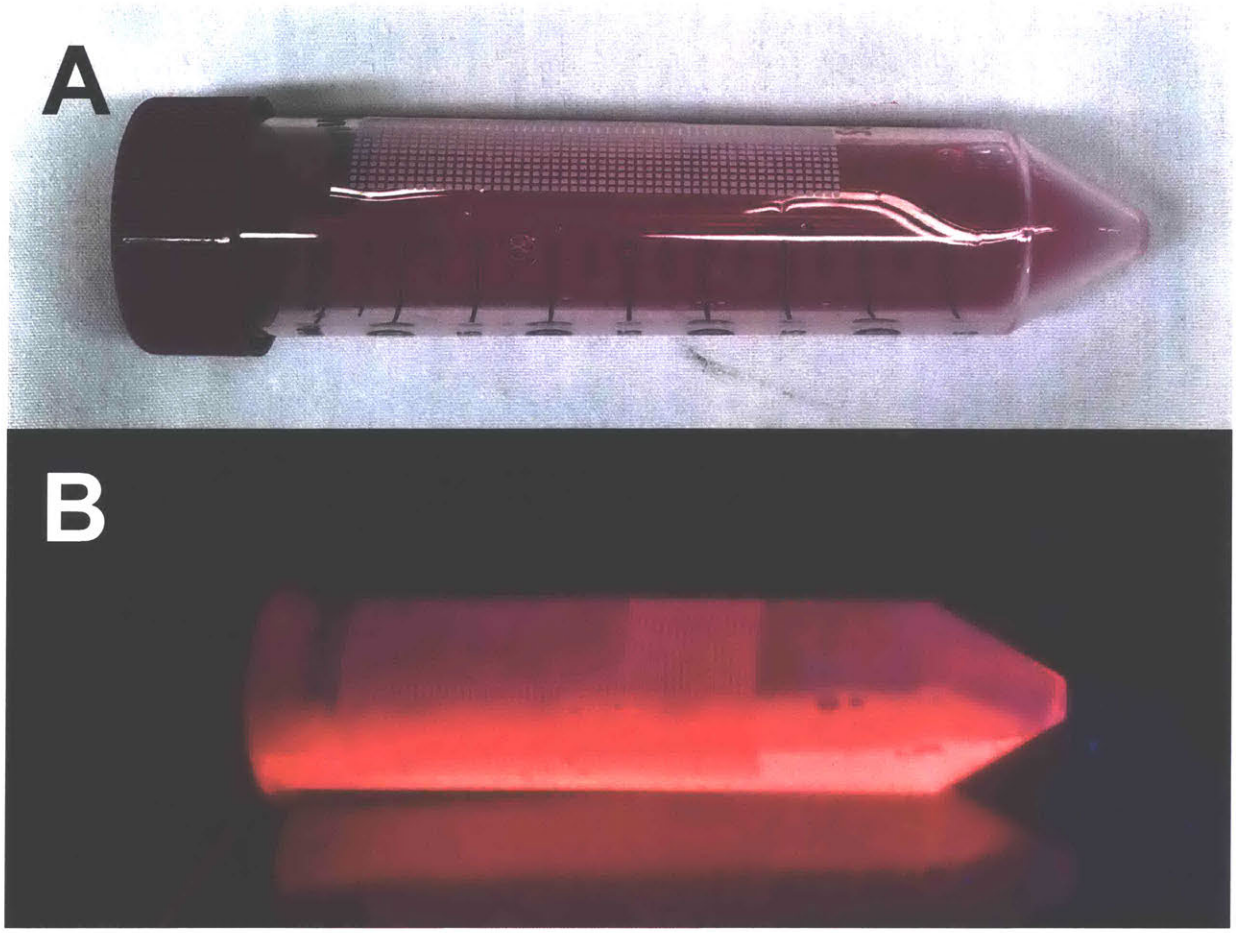



**Figure VI.2. Strep-II tagged mRuby2 fluoresces under UV-light.** Images of recombinantly expressed mRuby2 under **(A)** bright field and **(B)** UV-light.

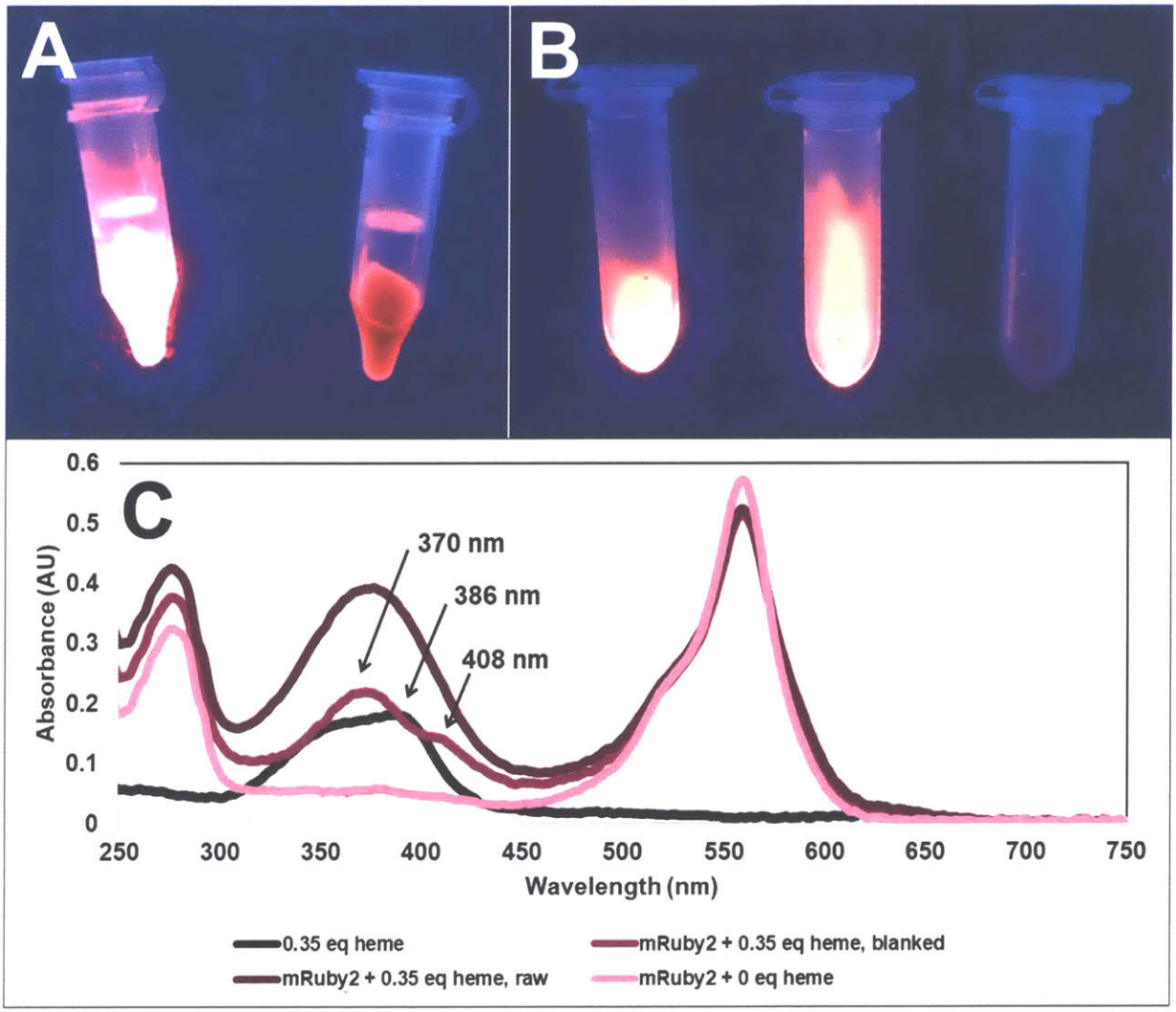



**Figure VI.3. mRuby2 fluorescence is quenched by the presence of heme. (A)** mRuby2 strongly fluoresces under UV light (left), however in the presence of equimolar heme, the fluorescence is quenched (right). **(B)** The same quenching of fluorescence is not observed upon incubation with the sodium hydroxide vehicle (center) as compared with protein alone (left) and protein incubated with equimolar heme (right). The color discrepancy observed is due to images being captured on two different days. **(C)** A UV-Vis spectral shift from the unbound form with a peak at 386 nm (black trace) to the bound form (dark pink) is observed upon heme binding Additionally, the shift in the spectrum once heme has been added is clearly not additive, as the peak at 370 nm in the heme-added sample is larger than just an additive effect from the free heme (black) to the mRuby2 spectrum with no heme present (light pink). Free-heme in solution masks the protein-bound heme signal, so a heme blank is necessary to observe the soret due to heme binding to protein (fuchsia curve).

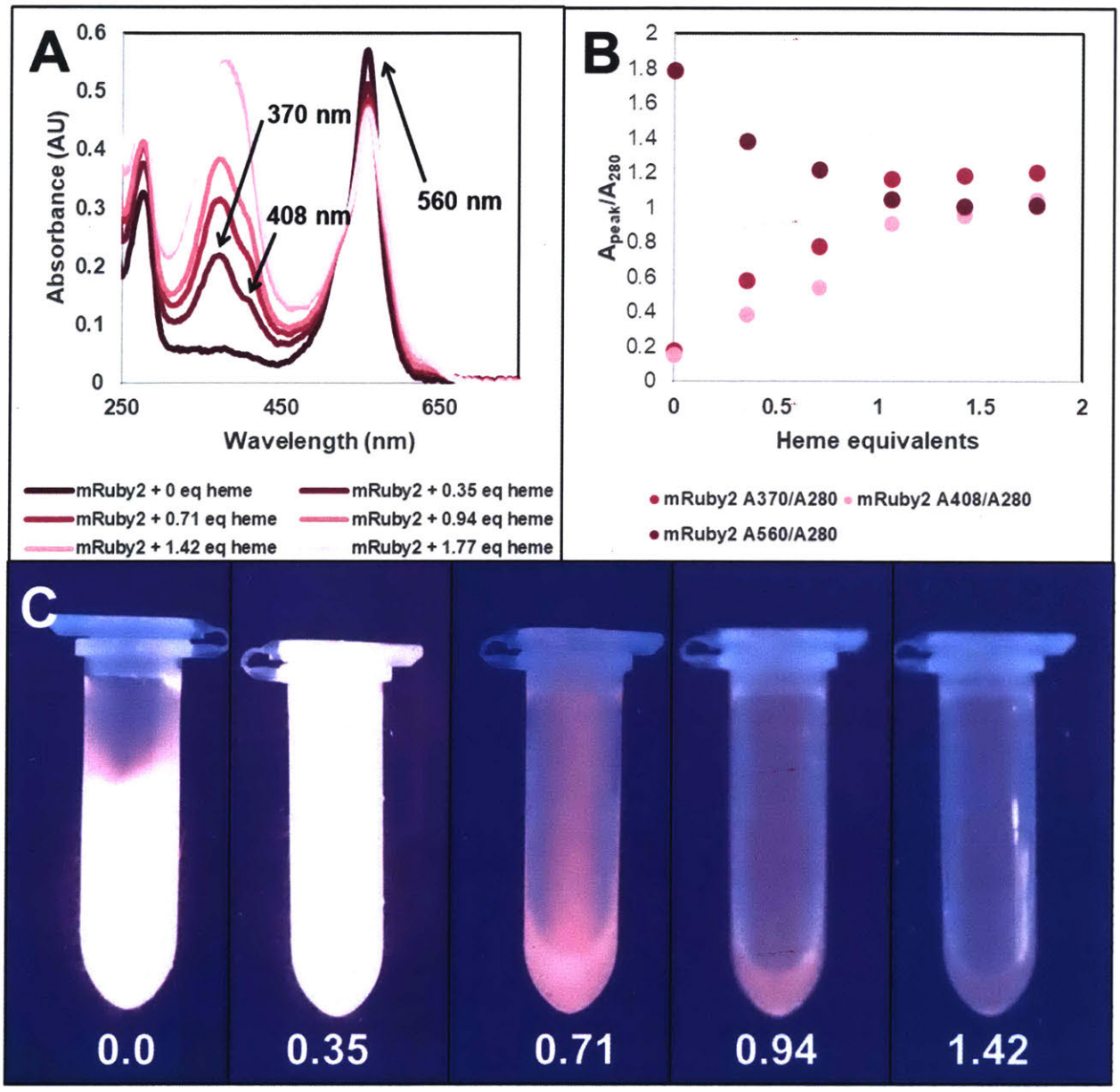



**Figure VI.4. Titration of mRuby2 with heme shows that heme binds to mRuby2. (A)** Two distinct binding modalities for heme are observed, with maximum absorbances of 370 and 408 nm. **(B)** The increase in both $A_{370}/A_{280}$ and $A_{408}/A_{280}$ ratio observed upon addition of heme is paired with a reduction in $A_{560}/A_{280}$ ratio. **(C)** The decrease in fluorescence intensity is visible under blacklight as heme is added, from 0.0 eq to 1.42 eq.

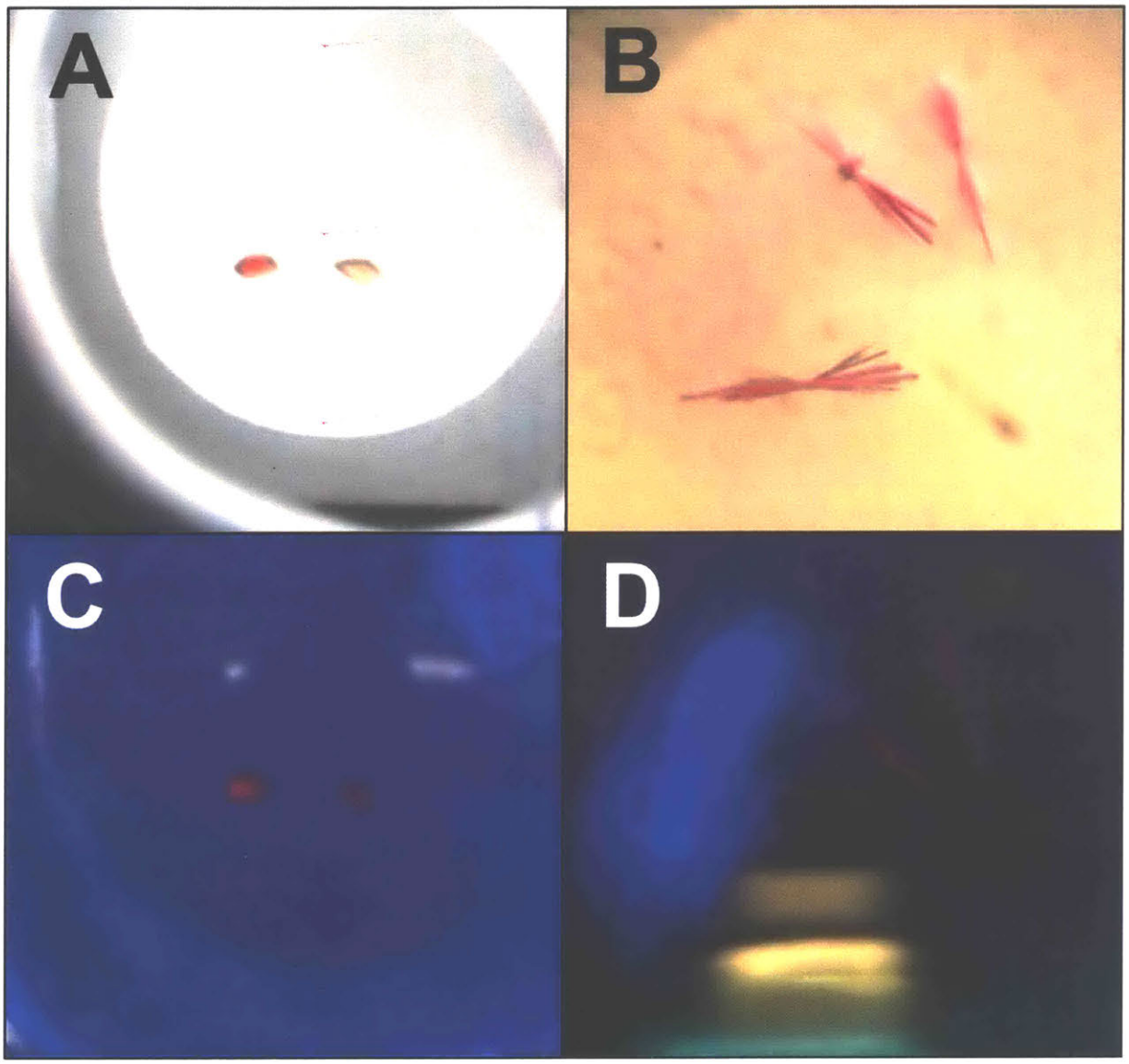



**Figure VI.5. Crystals of mRuby2 fluoresce under UV light.** Photos of mRuby2 under **(A & B)** bright field and **(C & D)** UV light.

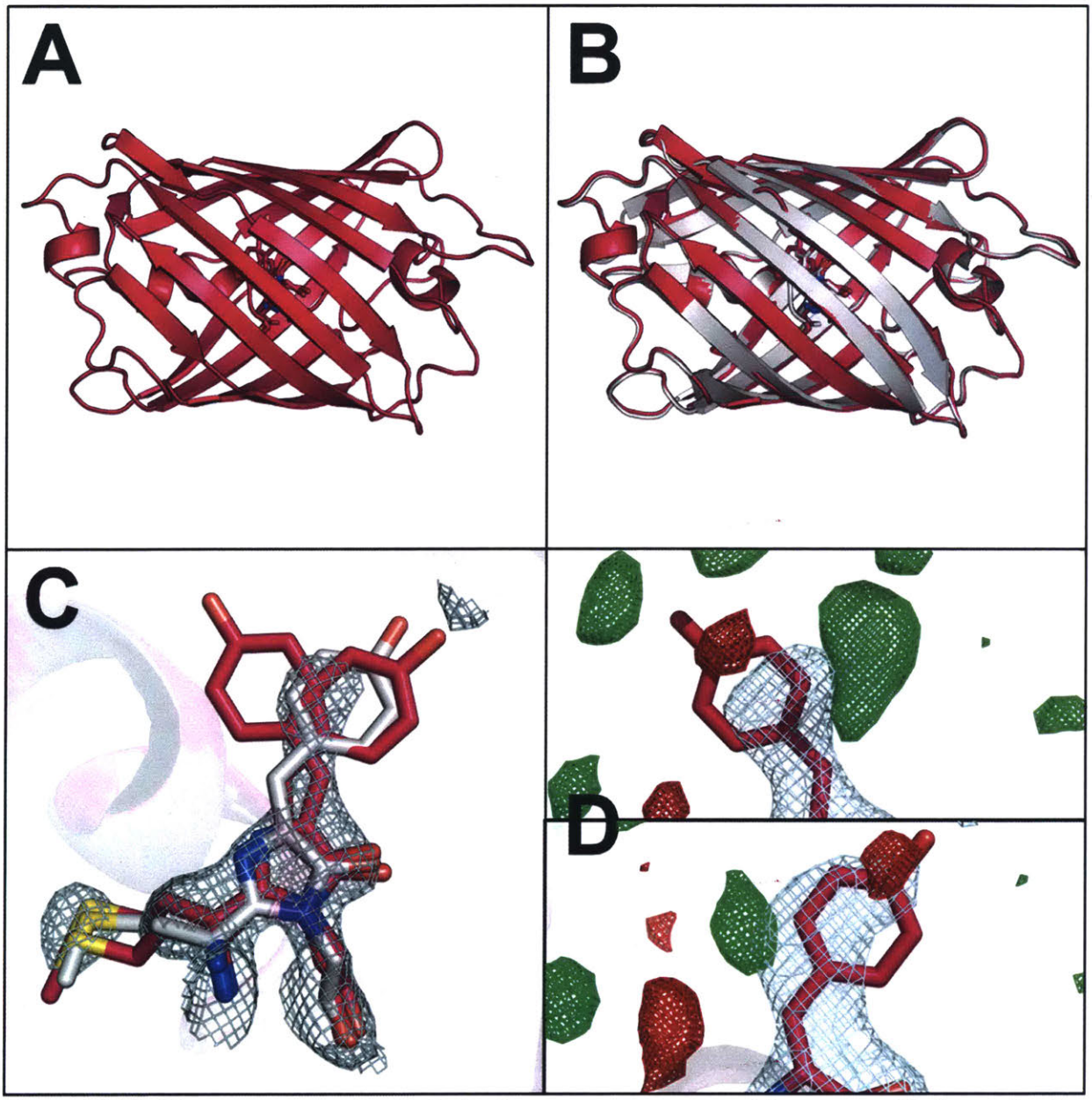



**Figure VI.6. The structure of mRuby2 shows a strong resemblance to mRuby and other fluorescent proteins.** The large β-barrel with an internal fluorophore is typical of these proteins. **(A)** The overall structure of mRuby2 (hot pink) looks similar to other RFPs. **(B)** Very minimal differences can be observed in the overall structure of mRuby2 when compared to its parent protein, mRuby (gray, PDB 3U0M) (12). **(C)** The fluorophore of mRuby2 at high pH is in a dual conformation as opposed to the high pH structure of mRuby (gray). **(D)** When a single conformation is refined, positive difference density appears for the other. The top image displays the *cis*-isoform and the bottom displays the *trans*-isoform. The $2F_o$-$F_c$ density is contoured to 1 σ in light green. The positive and negative $F_o$-$F_c$ density is displayed in green and red, respectively.

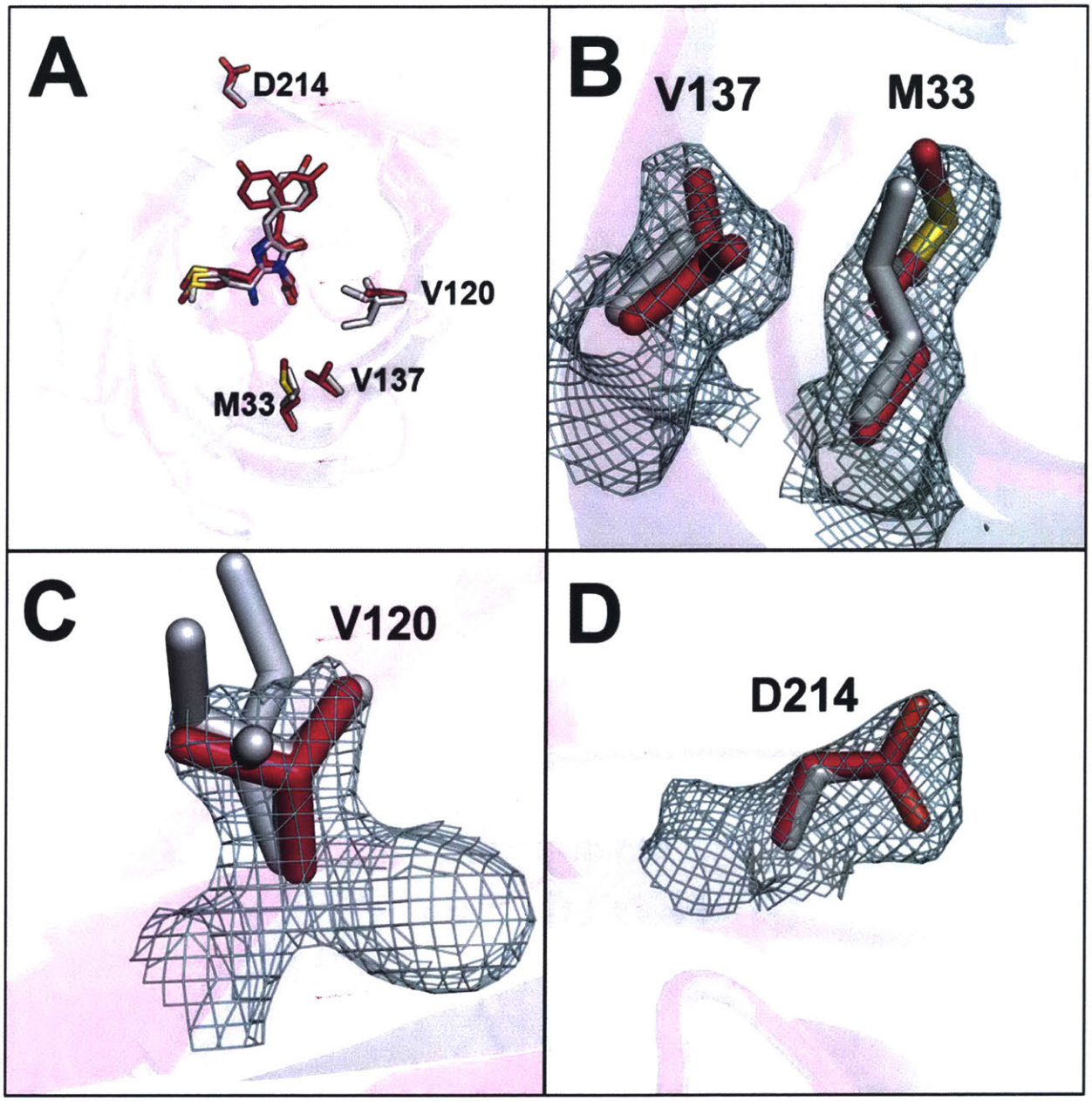



**Figure VI.7. The altered amino acids of mRuby2 are present in the structure of mRuby2.** The density for each residue confirms position and orientation. **(A)** None of these residues are close to the fluorophore. $2F_o$-$F_c$ maps were contoured to 1.0 σ in light green around **(B)** Met33 and Val137, **(C)** Val120, and **(D)** Asp214.

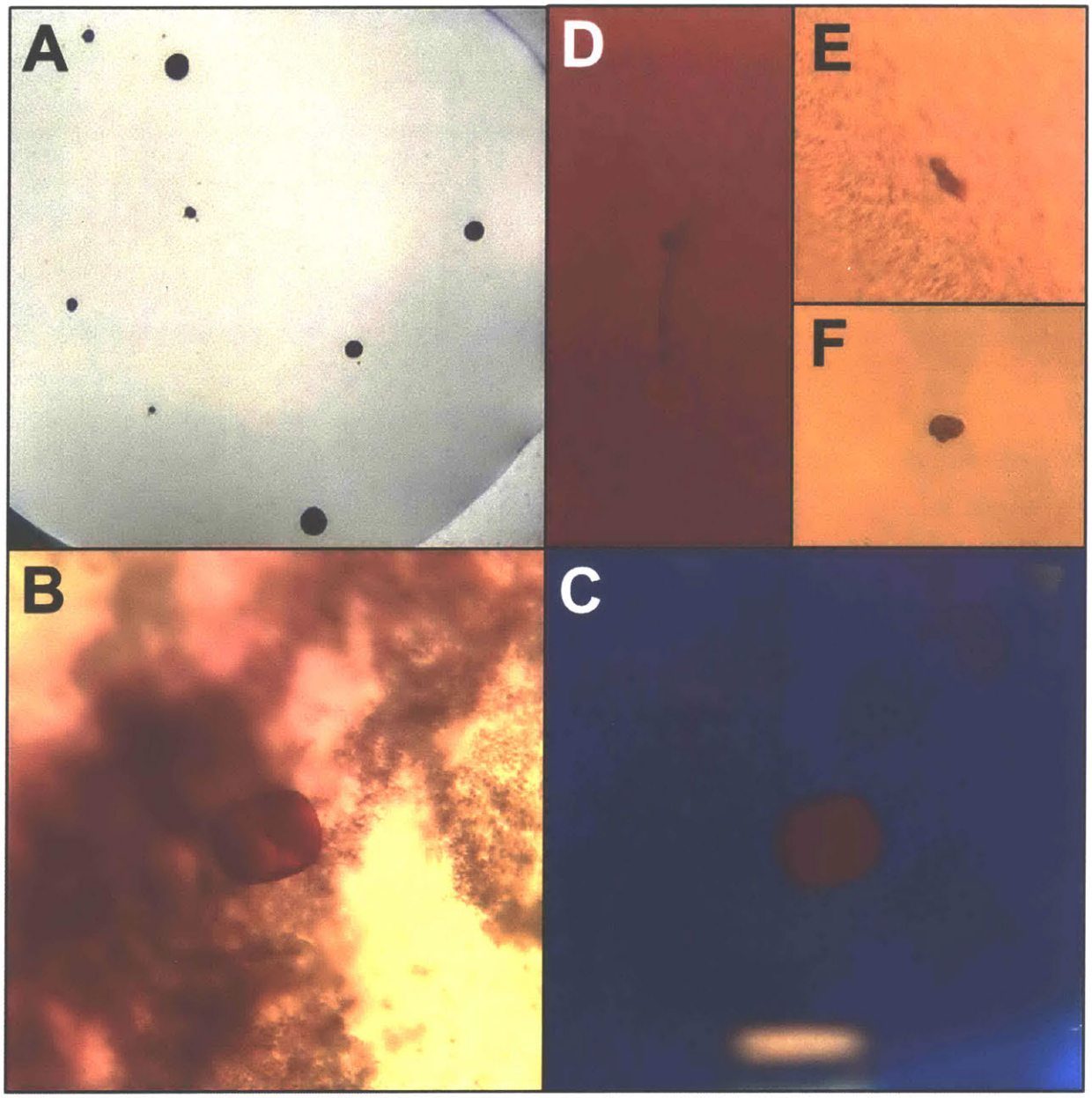



**Figure IV.8. Some of the quasi-crystals that form when mRuby2 is incubated with heme fluoresce whereas others do not. (A)** One of two initial quasi-crystalline conditions identified for further optimization of mRuby2 in the presence of one equivalent of heme. **(B-F)** Crystals and quasi-crystals obtained after optimization of the screening condition. Some quasi-crystals still fluoresce under UV-light (B-C) and likely do not have heme bound, but some of the crystals observed do not fluoresce under UV-light (D-F) and could have heme bound. Crystallization conditions are as follows: (A) 4.0 M potassium formate, 0.1 M Bis-TRIS propane, pH 9.0, and 2% PEG-MME 2000, (B-C) 0.1 M Bis-TRIS, pH 6.5, 30% PEG-3350, (D) 2.5 M potassium formate, 0.1 M Bis-TRIS propane, pH 9.0, 1.5% PEG-MME 2000, (E) 3.0 M potassium formate, 0.1 M Bis-TRIS propane, pH 9.0, 1.0% PEG-MME 2000, and (F) 3.0 M potassium formate, 0.1 M Bis-TRIS propane, pH 9.0, 2.0% PEG-MME 2000.

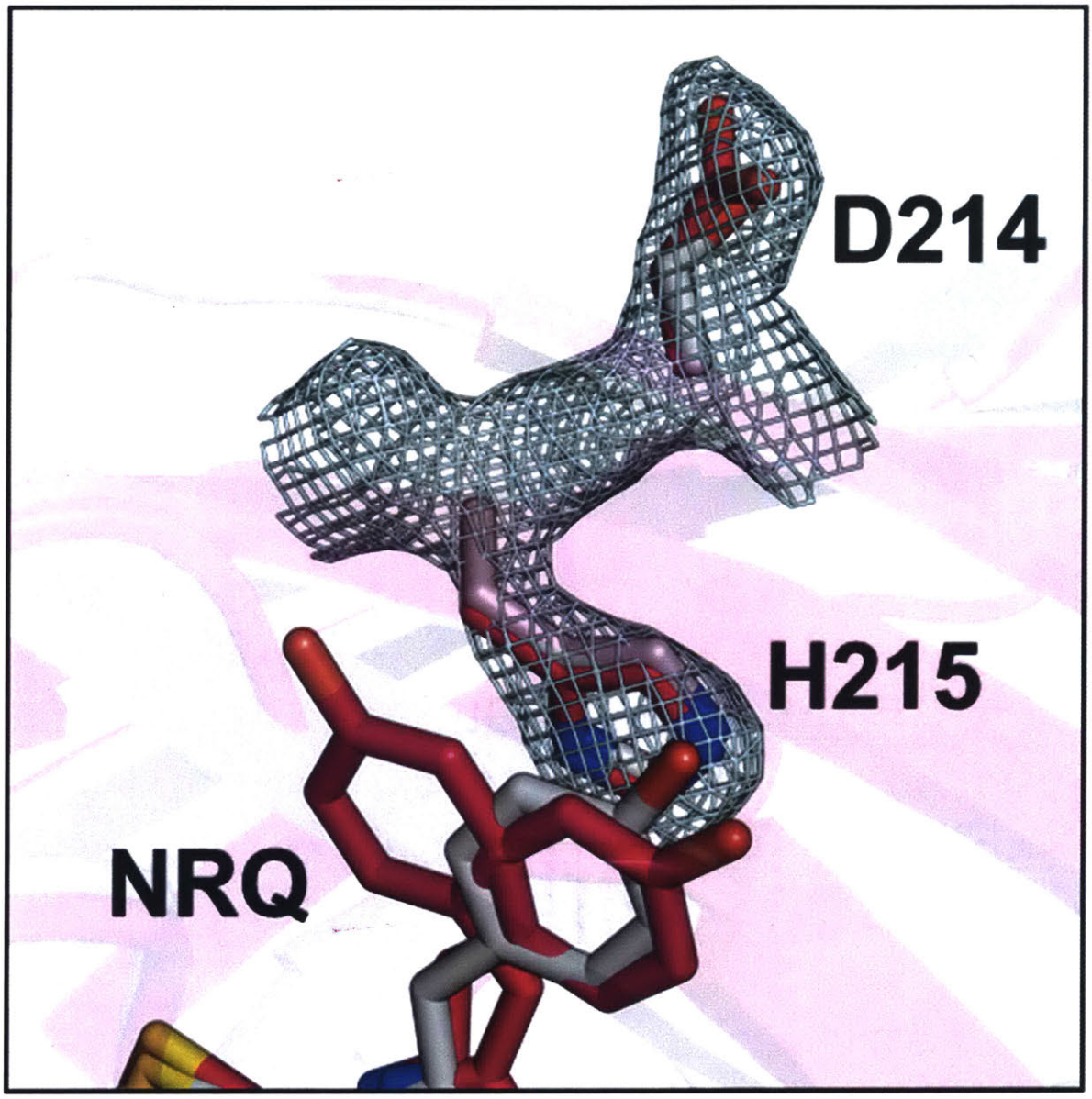



**Figure VI.9. Asp214 could be affecting the fluorophore through its adjacent residue, His215.** Although none of the substituted amino acids between mRuby and mRuby2 are closer than 6.7 Å to the fluorophore, the residue directly adjacent to Asp214, His215, is participating in a pi-stacking interaction with the *trans*-isomer of the fluorophore (12).

| | |
|---|---|
| Beamline | 24-ID-C |
| Wavelength (Å) | 0.9792 |
| Space group | $P\ 2_12_12_1$ |
| Cell dimensions (Å) | 31.27, 67.06, 95.87 |
| Resolution (Å) | 54.95 - 1.954 (2.024 - 1.954) |
| Unique reflections | 15163 (1418) |
| Completeness (%) | 99.31 (96.21) |
| Redundancy | 15.7 (15.7) |
| $R_{sym}$ | 0.146 (1.92) |
| $CC_{1/2}$ | 0.999 (0.609) |
| $I/\sigma(I)$ | 13.77 (1.31) |
| | |
| **Model refinement** | |
| Resolution limits (Å) | 54.95 - 1.954 (2.024 - 1.954) |
| $R_{work}/R_{free}$[&] | 0.1956/0.2307 |
| Reflections | 15134 |
| No molecules in asu | 1 |
| Number of non-hydrogen atoms | 1813 |
| Protein | 1660 |
| Fluorophore (NRQ)[$] | 23 |
| Solvent | 130 |
| Average B-factor ($Å^2$) | |
| Overall | 39.86 |
| Protein | 39.24 |
| Fluorophore (NRQ)[$] | 46.22 |
| Solvent | 47.04 |
| R.M.S. deviations | |
| Bond Lengths (Å) | 0.007 |
| Bond Angles (°) | 1.30 |
| Rotamer outliers (%) | 1.19 |
| Ramachandran Plot (%) | |
| Most Favored | 99.05 |
| Additionally Allowed | 0.95 |
| Disallowed | 0.00 |

**Table VI.1. Data and refinement statistics for the current model of mRuby2.**

[$]NRQ, the fluorophore, accounts for residues 81-83.

[&]5% of reflections used for $R_{free}$

Parentheses indicate statistics for high resolution bin

# REFERENCES


1. Jeney, V. (2002) Pro-oxidant and cytotoxic effects of circulating heme. **100**, 879-887
2. Sigala, P. A., and Goldberg, D. E. (2014) The Peculiarities and Paradoxes of Plasmodium Heme Metabolism. *Annual Review of Microbiology* **68**, 259-278
3. Arpino, J. A. J., Czapinska, H., Piasecka, A., Edwards, W. R., Barker, P., Gajda, M. J., Bochtler, M., and Jones, D. D. (2012) Structural Basis for Efficient Chromophore Communication and Energy Transfer in a Constructed Didomain Protein Scaffold. *Journal of the American Chemical Society* **134**, 13632-13640
4. Hanna, D. A., Harvey, R. M., Martinez-Guzman, O., Yuan, X., Chandrasekharan, B., Raju, G., Outten, F. W., Hamza, I., and Reddi, A. R. (2016) Heme dynamics and trafficking factors revealed by genetically encoded fluorescent heme sensors. **113**, 7539-7544
5. Yuan, X., Rietzschel, N., Kwon, H., Walter Nuno, A. B., Hanna, D. A., Phillips, J. D., Raven, E. L., Reddi, A. R., and Hamza, I. (2016) Regulation of intracellular heme trafficking revealed by subcellular reporters. *Proceedings of the National Academy of Sciences* **113**, E5144-E5152
6. Abshire, J. R., Rowlands, C. J., Ganesan, S. M., So, P. T. C., and Niles, J. C. (2017) Quantification of labile heme in live malaria parasites using a genetically encoded biosensor. 201615195
7. Song, Y., Yang, M., Wegner, S. V., Zhao, J., Zhu, R., Wu, Y., He, C., and Chen, P. R. (2015) A Genetically Encoded FRET Sensor for Intracellular Heme. **10**, 1610-1615
8. Lam, A. J., St-Pierre, F., Gong, Y., Marshall, J. D., Cranfill, P. J., Baird, M. A., McKeown, M. R., Wiedenmann, J., Davidson, M. W., Schnitzer, M. J., Tsien, R. Y., and Lin, M. Z. (2012) Improving FRET dynamic range with bright green and red fluorescent proteins. **9**, 1005-1012
9. Brewitz, H. H., Kühl, T., Goradia, N., Galler, K., Popp, J., Neugebauer, U., Ohlenschläger, O., and Imhof, D. (2015) Role of the Chemical Environment beyond the Coordination Site: Structural Insight into Fe (IIIP) rotoporphyrin Binding to Cysteine-Based Heme-Regulatory Protein Motifs. *ChemBioChem* **16**, 2216-2224
10. Wißbrock, A., Paul George, A. A., Hans, Kühl, T., and Imhof, D. (2019) The molecular basis of transient heme-protein interactions: analysis, concept and implementation. *Bioscience Reports* **39**, BSR20181940
11. Lan, W., Wang, Z., Yang, Z., Zhu, J., Ying, T., Jiang, X., Zhang, X., Wu, H., Liu, M., Tan, X., Cao, C., and Huang, Z.-X. (2011) Conformational Toggling of Yeast Iso-1-Cytochrome c in the Oxidized and Reduced States. *PLoS ONE* **6**, e27219
12. Akerboom, J., Carreras Calderón, N., Tian, L., Wabnig, S., Prigge, M., Tolö, J., Gordus, A., Orger, M. B., Severi, K. E., Macklin, J. J., Patel, R., Pulver, S. R., Wardill, T. J., Fischer, E., Schüler, C., Chen, T.-W., Sarkisyan, K. S., Marvin, J. S., Bargmann, C. I., Kim, D. S., Kügler, S., Lagnado, L., Hegemann, P., Gottschalk, A., Schreiter, E. R., and Looger, L. L. (2013) Genetically encoded calcium indicators for multi-color neural activity imaging and combination with optogenetics. *Frontiers in Molecular Neuroscience* **6**

13. Heppert, J. K., Dickinson, D. J., Pani, A. M., Higgins, C. D., Steward, A., Ahringer, J., Kuhn, J. R., and Goldstein, B. (2016) Comparative assessment of fluorescent proteins for in vivo imaging in an animal model system. *Mol Biol Cell* **27**, 3385-3394
14. Gateiger, E., Hoogland, C., Gattiker, A., Devaud, S., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005) Protein identification and analysis tools on the ExPASy server. in *The Proteomics Protocols Handbook* (Walker, J. M. ed.), Humana Press. pp 571-607
15. Kabsch, W. (2010) XDS. *Acta Crystallographica Section D Biological Crystallography* **66**, 125-132
16. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *Journal of Applied Crystallography* **40**, 658-674
17. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and Development of Coot. *Acta Crystallographica Section D Biological Crystallography* **66**, 486-501
18. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012) Towards automated crystallographic structure refinement withphenix.refine. *Acta Crystallographica Section D Biological Crystallography* **68**, 352-367
19. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography* **66**, 213-221

# CHAPTER VII: THE UTILIZATION OF IRON BINDING PROTEINS TOWARD THE INVESTIGATION AND TREATMENT OF *PLASMODIUM*

Malaria is a potentially fatal disease caused by *Plasmodium*, a eukaryotic parasite (1). In order to identify novel methods of treating and preventing malaria, we need to first better understand the biology of *Plasmodium*. To add to the understanding of this deadly parasite, we have undertaken studies of two metal-binding proteins. The first, MFP, is a novel *Plasmodium* protein of unknown function which is a potential drug target. The second, mRuby2, is a fluorescent probe already utilized in biological studies for which we have identified a novel use as a tool for testing heme localization *in vivo*.

## MFP AS A NOVEL POTENTIAL THERAPEUTIC TARGET

The increase in drug resistance of *Plasmodium* is prompting studies into novel ways of controlling the spread of malaria as well as a desire for novel therapeutics (2,3). In the process of searching for a target *Plasmodium* protein, we identified the gene PF3D7_0925900 and the protein it expresses, MFP, for further study. The essential and conserved nature of the protein coupled with its interesting metal binding properties makes it an interesting avenue of study for novel therapeutic development (4,5).

The targeting of essential proteins within bacteria and parasites by therapeutics is a common strategy for treating infections, reducing the cellular viability of the organism, either killing it or leaving it highly susceptible to the host immune system. Although bacterial systems utilize vastly different biology than their human hosts, eukaryotic parasites contain much more overlap in biological function, making identification of therapeutic targets less straightforward (6). In order to specifically target *Plasmodium*, identification of proteins without an overlapping function in the vertebrate host is crucial. Thus, MFP, lacking any identified homologs in humans, is promising as a target.

Additionally, due to the conserved nature of this protein across all *Plasmodium* species, it is possible that a single therapeutic could be utilized to treat infections from multiple species of the parasite. The utility of this in areas where malaria is common cannot be understated, as the testing to determine which strain of *Plasmodium* is causing a malaria infection could be bypassed, reducing the time to treatment as well as the cost. This protein is also found in other non-*Plasmodium* parasites and could provide insight into the creation of a broad-spectrum therapeutic. Additionally, because we have not found a similar protein within primates, the likelihood of off-target effects is low if the therapeutic could be made specific for MFP. The therapeutic specificity could be aided by the interesting metal binding properties of this protein.

Through our work, we have determined that MFP binds both heme and an iron-sulfur cluster, but not at the same time. The oxygen-sensitive [4Fe-4S] cluster inhibits heme binding to MFP under our experimental conditions and provides potential insight into the way these two cofactors bind. Given the small size of the protein, our data on the cysteine variants of MFP, and the inability to bind heme and cluster at the same time, it is likely that the binding sites for both cluster and heme are in close proximity to each other, increasing the likelihood of the development of a small molecule therapeutic that prevents the binding of either one or both cofactors. While the *in vivo* utility of these two binding partners have yet to be determined, the dual metal-binding properties of MFP could help in the drug development process. Assuming that both cofactors are necessary for full function at different times, we could target either one or both metal-bound states to decrease the likelihood of the development of resistance.

## MUCH IS YET TO BE DISCOVERED ABOUT MFP

The previously uncharacterized protein MFP has now been identified to bind both heme and iron-sulfur cluster, two iron-containing prosthetic groups with many diverse functions. Although the presence of these two specific binding partners lends some insight into the potential functions of this protein, the wide array of functions performed by both heme and iron-sulfur clusters leave many options open for the biological relevance of MFP (7,8). Investigation of the *in vivo* function of MFP is ongoing and will likely lend insight into the reason for this dual metal binding.

In the future, structural characterization of MFP with either heme or iron-sulfur cluster bound will likely yield insight into both the function of the protein when either cofactor is present as well as the mechanism of inhibition with respect to cofactor binding. The discovery of the transient nature of the disulfide bond and the new variant C68A/C93A that stabilizes this bond between Cys70 and Cys216 will aide in the solution of an apo-structure. In addition to adding to our understanding of cofactor binding of this essential *Plasmodium* protein, the experiments outlined in this thesis have laid the ground work for identifying conditions optimal for MFP *in vitro*, greatly increasing the chance of structural determination by crystallography.

In addition to the fact that we have not as of yet discovered the biological function of MFP, the fact that *Plasmodium* is a eukaryotic parasite complicates our ability to study this protein *in vitro*. Eukaryotic proteins often have post-translational modifications rarely accessible through bacterial recombinant protein expression (9). Recently, it has been reported that MFP from *Plasmodium falciparum* expressed in eukaryotic cell lines contains both glycosylation (10) and lysine acetylation (11) sites. These types of modifications are non-trivial when protein is expressed in bacterial cell lines and, therefore, the metal-binding properties of MFP have not been tested when these modifications are present. It is possible that these modifications will alter the ability of MFP to bind either heme or an iron-sulfur cluster and the relevance of that discrepancy should be investigated.

## MRUBY2 COULD BE USED AS A HEME SENSOR DURING THE STUDY OF *PLASMODIUM*

In addition to our identification of the novel protein MFP, we serendipitously identified a novel use of mRuby2 as a heme sensor. This fluorescent probe was developed for FRET assays when paired with Clover, but can also be utilized in *in vivo* cellular studies (12). The identification of the fluorescence quenching of mRuby2 by heme provides a potential tool for metal localization studies both in *Plasmodium* as well as in other organisms. With this tool in hand, we can add to our understanding of the location and, potentially, concentrations of free heme in the cell, helping to further our understanding of iron and heme regulation within *Plasmodium*.

# CONCLUSION

This thesis has characterized two proteins which will help to further research of *Plasmodium* and development of novel treatments for malaria. The first, MFP, is an essential metal binding protein from *Plasmodium* that could be a novel drug target for malaria therapeutics. The second, mRuby2, can be used as a tool in understanding metal acquisition and sequestration within *Plasmodium*. Taken together, we have laid the ground work for future studies of *Plasmodium* biology and its role in malaria infection.

# REFERENCES


1. Ashley, E. A., Pyae Phyo, A., and Woodrow, C. J. (2018) Malaria. *The Lancet* **391**, 1608-1621
2. World Health Organization. (2018) Antimalarial drug efficacy and drug resistance.
3. Menard, D., and Dondorp, A. (2017) Antimalarial Drug Resistance: A Threat to Malaria Elimination. a025619
4. Matz, J. M., and Matuschewski, K. (2018) An in silico down-scaling approach uncovers novel constituents of the Plasmodium-containing vacuole. *Scientific Reports* **8**
5. Aurrecoechea, C., Brestelli, J., Brunk, B. P., Dommer, J., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G., Harb, O. S., Heiges, M., Innamorato, F., Iodice, J., Kissinger, J. C., Kraemer, E., Li, W., Miller, J. A., Nayak, V., Pennington, C., Pinney, D. F., Roos, D. S., Ross, C., Stoeckert, C. J., Treatman, C., and Wang, H. (2009) PlasmoDB: a functional genomic database for malaria parasites. **37**, D539-D543
6. Fairlamb, A. H., Gow, N. A. R., Matthews, K. R., and Waters, A. P. (2016) Drug resistance in eukaryotic microorganisms. *Nature Microbiology* **1**, 16092
7. Chiabrando, D., Vinchi, F., Fiorito, V., Mercurio, S., and Tolosano, E. (2014) Heme in pathophysiology: a matter of scavenging, metabolism and trafficking across cell membranes. *Frontiers in Pharmacology* **5**
8. Johnson, D. C., Dean, D. R., Smith, A. D., and Johnson, M. K. (2005) Structure, function, and formation of biological iron-sulfur clusters. *Annual Review of Biochemistry* **74**, 247-281
9. Khow, O., and Suntrarachun, S. (2012) Strategies for production of active eukaryotic proteins in bacterial expression system. *Asian Pacific Journal of Tropical Biomedicine* **2**, 159-162
10. Nikolaeva, D., Illingworth, J. J., Miura, K., Alanine, D. G., Brian, I. J., Li, Y., Fyfe, A. J., Da, D. F., Cohuet, A., Long, C. A., Draper, S. J., and Biswas, S. (2017) Functional characterization and comparison of Plasmodium falciparum proteins as targets of transmission-blocking antibodies. *Molecular & Cellular Proteomics*, mcp.RA117.00003
11. Cobbold, S. A., Santos, J. M., Ochoa, A., Perlman, D. H., and Llinás, M. (2016) Proteome-wide analysis reveals widespread lysine acetylation of major protein complexes in the malaria parasite. **6**, 19722
12. Lam, A. J., St-Pierre, F., Gong, Y., Marshall, J. D., Cranfill, P. J., Baird, M. A., McKeown, M. R., Wiedenmann, J., Davidson, M. W., Schnitzer, M. J., Tsien, R. Y., and Lin, M. Z. (2012) Improving FRET dynamic range with bright green and red fluorescent proteins. **9**, 1005-1012